

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 9

Москва 2023

ОБЩИЙ РАЗДЕЛ

УДК 002:311.311:004.774 Наука 2.0

П.А. Калачихин

Взаимозависимость показателей результативности научных исследований*

Сопоставлены традиционный и сетевой типы показателей на предмет взаимозаменяемости и взаимодополняемости при оценке результатов научной деятельности. Рассмотрены примеры вебметрических индикаторов, а также рейтингов. Дана характеристика сильным и слабым сторонам альтметрик. Обобщены выводы о корреляции различных показателей по материалам исследований других авторов. Предложена модель воронки вебметрических показателей, которая может использоваться как инструмент повышения популярности электронных публикаций.

Ключевые слова: альтметрики, вебметрические показатели, воронкообразная диаграмма, корреляция, Наука 2.0, оценка результатов научной деятельности, цифровые платформы

DOI: 10.36535/0548-0019-2023-09-1

* Работа выполнена в рамках исследования по теме FFFU-2022-0007 Госзадания ВИНТИ РАН.

ВВЕДЕНИЕ

Традиционные показатели измерения результативности научной деятельности, в состав которых входят наукометрические и библиометрические индикаторы, широко распространены и имеют солидную репутацию в академических кругах [1, с. 16]. Сетевые показатели, представленные киберметриками, вебометриками и альтметриками, появились относительно недавно и вызывают настороженность, как и все нововведения.

Ситуация, связанная с уходом в 2022 г. с российского рынка крупных международных компаний, поставляющих библиометрическую информацию, заметно подогревает интерес к сетевым показателям, поскольку они могут рассматриваться как замена традиционным. Однако, существует политический риск потери доступа отечественных пользователей к сетевым метрикам из-за блокировки трафика, в результате возврат к традиционным показателям может обернуться для них проблемой.

Цель настоящего исследования – поиск ответов на вопросы: открывают ли сетевые показатели принципиально новые возможности, дополняют ли они уже известные показатели или противоречат им? Для этого выполнено обобщение полученных другими авторами выводов о наличии или отсутствии корреляции между разными показателями и сопоставлены особенности такой корреляции для различных направленных научных исследований.

Прежде чем переходить к изучению корреляции между различными показателями, следует разобраться в их разновидностях более детально, придерживаясь той хронологической последовательности, в которой они зародились.

РОЛЬ ВЕБОМЕТРИК В ОЦЕНКЕ РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ

Появление дисциплины «киберметрия» (*cybermetry*) связано с бурным развитием компьютерных технологий в 90-х гг. XX в. Приставка «кибер» говорит об отношении к кибернетике, «метрия» – слову «измерение». Не трудно догадаться, что объектом исследования киберметрии является информация, хранящаяся в цифровом формате, хотя в некоторых словарях киберметрия обозначена в качестве синонима к слову «науковедение». Так или иначе, показатели, с помощью которых измеряются количественные характеристики цифровых данных, принято называть *киберметриками* (*cybermetrics*).

Возникновение *вебометрии* (*webometry*) связано с созданием Интернета, который, помимо прочих способов обмена информацией, основан на технологии World Wide Web или кратко – Web¹. Вебометрия появилась на несколько лет позже киберметрии [2] и стала частью киберметрии [3].

Вебометрические показатели, называемые *вебометриками* (*webometrics*), используются для рейтин-

гования сайтов образовательных (реже – научных) организаций [4], а также библиотек [5]. Интерес к вебометрикам связан как со стремлением повышать посещаемость интернет-ресурсов, так и с требованиями законодательства РФ.

Вебометрический рейтинг – это мера популярности сайта на основании сравнительной оценки ключевых показателей деятельности организации в сетевом пространстве. Как правило, рейтинги используются при подведении итогов деятельности организаций, сообществ и отдельных лиц, в частности, в сетевом пространстве [6, с. 88].

Начало вебометрии было положено в 2004 г., когда испанская Лаборатория киберметрии (*Cybermetrics Lab*) Национального исследовательского совета Испании (*Spanish National Research Council, CSIC*) представила Вебометрический рейтинг университетов мира (*Webometrics Ranking of World Universities*), в котором оценка образовательных и научно-исследовательских достижений университетов мира выполнялась на основе анализа представления вузов в сетевом пространстве через сравнение их веб-сайтов [7, с. 328].

Для расчета рейтинга *Webometrics* с помощью основных поисковых систем подсчитывались 4 количественных показателя:

- Размер (*S*) – количество страниц сайта, извлеченных из поисковых систем Google, Yahoo, Live Search и Exalead;
- Видимость (*V*) – общее количество уникальных внешних ссылок в Yahoo Search, Live Search и Exalead;
- Ценные файлы (*R*) – количество выложенных на сайте документов в форматах Adobe Acrobat (.pdf), Adobe PostScript (.ps), Microsoft Word (.doc) и Microsoft PowerPoint (.ppt), подсчитанное поисковой системой Google;
- Scholar (*Sc*) – количество публикаций и цитат, найденных в Google Scholar [8, с. 71].

Формула подсчета итогового значения показателя в рейтинге *Webometrics* выглядит следующим образом:

$$Q_j = 4 \cdot V_j + 2 \cdot S_j + 1 \cdot R_j + 1 \cdot Sc_j, \quad (1)$$

где: Q_j – значение результирующего показателя (рейтинга) для j -го вуза;

V_j – место j -го вуза в ранжировке по количеству уникальных внешних ссылок на страницы официального интернет-сайта вуза;

S_j – место j -го вуза в ранжировке по количеству страниц официального интернет-сайта вуза, покрываемых поисковыми системами;

R_j – место j -го вуза в ранжировке по количеству «ценных» файлов, размещенных на официальном интернет-сайте вуза;

Sc_j – место j -го вуза в ранжировке по количеству страниц и ссылок на официальный интернет-сайт вуза [9, с. 20].

Свертка частных показателей в интегральный может выполняться по различным формулам. У испанских коллег показатели имеют весовые коэффициенты, которые периодически меняются с целью усиления

¹ Web 1.0, Web 2.0 и все последующие поколения Web далее будут обозначаться как «сетевое пространство» (*network space*).

роли размещенных на сайте полнотекстовых файлов и статей, а также их цитирования в Интернете. Другие исследователи для каждого критерия находят порядковый номер организации, а затем суммируют полученные ранги [7, с. 330].

К наиболее популярным вебметрическим индикаторам, которые могут использоваться как независимо, так и в составе рейтинговых и многокомпонентных показателей, относятся следующие.

Просмотр страницы (page view) – процесс, в результате которого пользователь просматривает содержимое страницы сайта. *Глубина просмотра (visit depth)* – среднее количество просмотров страниц сайта, которые пользователь совершает за одно посещение:

$$D_v = \frac{T_p}{T_q}, \quad (2)$$

где: D_v – глубина просмотров;
 T_p – общее количество просмотров страниц за определенный период времени;
 T_q – количество посетителей сайта за определенный период времени [10, с. 28].

Отказ означает ситуацию, когда посетитель зашел на сайт и, просмотрев только одну страницу, не открывает другие страницы в течение одной сессии. *Показатель отказа (bounce rate)* можно вычислить по формуле:

$$R_b = \frac{T_v}{T_e}, \quad (3)$$

где: R_b – показатель отказов;
 T_v – количество посетителей, просмотревших одну страницу (количество отказов);
 T_e – количество просмотров страницы [10, с. 29].

Среднее время пребывания на сайте можно вычислить по формуле:

$$T_s = \frac{T_c}{Q_p}, \quad (4)$$

где: T_s – среднее время пребывания на сайте;
 T_c – количество посетителей, просмотревших только одну страницу (количество отказов);
 Q_p – количество просмотров страницы [10, с. 30].

Минус вебметрических показателей в том, что они могут искажаться различными техническими способами – в процессе вторжения ботов, использования «трюков» анонимизации пользователей и пр. [11, с. 56]. Однако, по мере развития веб-технологий, появляются новые способы как «накрутки» показателей, так и борьбы с ними.

Сетевое общение в эпоху господства Web 1.0 происходило на гостевых книгах, чатах и форумах. Web 2.0 подарил интернет-пользователям возможность общаться в блогах и социальных сетях. Научное сообщество не стало игнорировать эти инновации. В результате появился подход «Наука 2.0»,

ознаменовавший рождение новой концепции научных коммуникаций. В Науке 2.0 ученые могут взаимодействовать более продуктивно, не говоря уже о других преимуществах. Оценка результативности исследований в рамках концепции Наука 2.0 требует специфических показателей, которые будут рассмотрены далее.

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ АЛЬТМЕТРИК В ОЦЕНКЕ РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ

Растущая роль социальных сетей привела к появлению нового показателя исследовательской деятельности – альтернативной метрики или альтметрики. Термин «альтметрика» (altmetrics) – сокращенное от alternative metric – введен американским специалистом Д. Тараборелли (D. Taraborelli) в 2008 г. [12, с. 811].

Альтметрики и вебметрики предназначены для измерения популярности информационных объектов, погруженных в сетевое пространство. В эпоху Web 2.0 вебметрики продолжают развиваться, однако они не предъявляют «социализированных» требований к цифровым платформам, как альтметрики, которые являются порождением Науки 2.0. Под *цифровой платформой* имеется в виду компонент инфраструктуры сетевого пространства, обеспечивающий выполнение транзакций и взаимодействие между пользователями в режиме онлайн.

Альтметрики могут быть привязаны к таким типам цифровых платформ:

- электронные базы данных;
- социальные сети;
- онлайн-сервисы;
- торговые площадки;
- агрегаторы данных.

Среди агрегаторов альтметрической информации в первую очередь следует отметить Altmetric и PlumX Metrics. Некоторые другие платформы, такие как ResearchGate и ScienceOpen, также позволяют собирать данные альтметрик. Отечественные ресурсы Научная электронная библиотека eLibrary и КиберЛенинка способны подсчитывать альтметрики, но уступают по функциональным возможностям зарубежным аналогам.

Библиометрические показатели плохи тем, что не способны различать контекст, в котором цитируются публикации. По своей тональности так называемый «лайк» (like) противоположен «дизлайку» (dislike), а положительный, нейтральный или отрицательный контекст упоминаний в средствах массовой информации (СМИ) еще более существенен, чем при цитировании. Успехи в развитии методов *анализа настроений* (sentiment analysis) уже сейчас позволяют создавать программы, способные определять тональность текста [13, с. 79], однако на данном этапе такой возможности нет. Таким образом, тональность альтметрик должен определять эксперт, и в этом смысле альтметрики приходится сочетать с экспертными оценками.

Специалисты зарубежных электронных библиотек уже активно используют альтметрики в интересах своих пользователей. Внимание же отечественных ученых по отношению к альтметрической информа-

ции представляется однобоким. Причины этого в том, что альтметрические сервисы почти не отражают отечественные ресурсы, российская наука слабо представлена в социальных сетях, обращению к альтметрикам препятствует языковой барьер, и к тому же для ученых неочевидна их польза [14, с. 83].

Альтметрики могли бы стать подспорьем при планировании тематики журналов и при поиске потенциальных авторов публикаций. Они могут помочь научным фондам отслеживать влияние результатов реализации поддержанных проектов. При отборе ученых в общественные или экспертные советы альтметрические показатели могут быть полезны для оценки их научной репутации [15, с. 16].

Тем не менее, в отношении альтметрик высказывается целый ряд критических замечаний.

Обычно утверждается, что альтметрики можно получать очень быстро. Но, наряду с «быстрыми альтметриками» (лайки, закладки, краткие сообщения в блогах), оцениваются и метрики «с задержкой», например, комментарии по поводу документа.

Помимо этого, из-за разнородности групп пользователей социальных сетей на основе альтметрик оценивают, скорее, социальную, прикладную либо образовательную значимость публикации, а не фундаментальную [16, с. 85].

Как библиометрические показатели, так и альтметрики остаются «слабыми» индикаторами, поскольку чреватые недостатками, причем манипулирование – это самый большой из них. Способа полного устранения манипуляций пока не существует. В частности, манипулирование библиометрическими показателями относится к вмешательству человеческого фактора [17]. Использование альтметрических инструментов недобросовестными исследователями позволяет им покупать лайки, репосты, комментарии и так далее [18, с. 38].

Зыбкость и неотлаженность альтметрических показателей в том виде, в каком они сейчас существуют, ставит под вопрос возможность их использования при принятии серьезных управленческих решений, в особенности при распределении финансовых средств на научные исследования [15, с. 14]. Однако в перспективе это заключение может быть пересмотрено.

Как мы видим, использование альтметрик имеет как положительные, так и отрицательные стороны. В отношении альтметрик одни авторы настроены оптимистично, другие занимают более сдержанную, консервативную и осторожную позицию. Прежде чем рассматривать возможность использования альтметрик в управлении наукой на государственном уровне, представляется целесообразным детально изучить корреляцию альтметрик с другими показателями.

АНАЛИЗ КОРРЕЛЯЦИИ МЕЖДУ ПОКАЗАТЕЛЯМИ ОЦЕНКИ РЕЗУЛЬТАТОВ НАУЧНЫХ ИССЛЕДОВАНИЙ

Обзоры исследований по корреляции альтметрик с библиометрическими и прочими типами показателей нередко встречаются в работах других авторов [19-21], поэтому нам представляется уместным рассмотреть эту проблему с разных сторон.

Многие авторы, независимо друг от друга, заявляют о существовании положительной корреляции между альтметриками и традиционными показателями.

Так, анализ наличия альтметрических показателей в статьях о болезни Эбола, опубликованных в журнале «PLOS ONE» в 2010-2015 гг. [22, с. 114-115], показывает сильную положительную связь между цитированием и количеством загрузок текстов публикаций, а также между просмотрами и цитированиями.

В [23, с. 399] статистически установлено, что журналы по политологии с высоким импакт-фактором имеют высокий альтметрический балл на ресурсе Altmetric.com.

Согласно [20, с. 333], корреляции между количеством читателей и цитируемостью особенно сильны в социально-гуманитарных науках, и корреляция между количеством упоминаний в онлайн-документах кажется относительно сильной, в то время как корреляции между цитируемостью и другими альтметрическими индексами очень слабые.

В [24] констатируется, что существует двусторонняя связь между альтметрической активностью и показателями качества научных журналов, так что увеличение одного из них может быть эффективным для увеличения другого.

Между тем, сила корреляции альтметрик с традиционными показателями различается по областям науки.

Так, в [25] отмечается, что самая высокая *альтметрическая плотность* (т.е. среднее число альтметрических показателей на публикацию) приходится на социальные и гуманитарные науки, за которыми следуют медицина, биология и науки о Земле. Математика и информатика, а также естественные и инженерные науки являются областями с наименьшим присутствием и плотностью альтметрик на статью. Для этих областей альтметрики имеют меньшие шансы стать альтернативой или дополнением к цитированию, если в ближайшие годы статистика не изменится. Более того, далеко не все авторы согласны с тем, что альтметрики положительно коррелируют с традиционно используемыми показателями.

В работе [26, с. 11] не удалось установить четкую корреляционную связь между востребованностью научной статьи в пользовательской интернет-среде и в научном сообществе.

В целом для научных публикаций присутствие и плотность альтметрических показателей в социальных сетях низки. Только 15-24% статей демонстрируют альтметрическую активность, и это характерно только для самых последних по времени публикаций [22, с. 105].

Как следствие, многие исследователи негативно отзываются о целесообразности перехода на альтметрики.

В связи со сложившейся тенденцией, а также с уникальными пространственными и временными характеристиками коммуникаций на платформах социальных сетей, альтметрики не могут в полной мере измерить вклад научных исследований, а также заменить цитирование и прочие библиометрические показатели [27, с. 344].

Вопрос о потенциальной взаимодополняемости альтметрик как источника доказательств о других видах воздействия, не отраженных в цитатах, все еще остается открытым и нуждается в дальнейшем изучении. Учитывая растущее присутствие альтметрик в последние годы, предстоит решить, какую роль они могли бы сыграть в оценке исследований, в частности, можно ли рассматривать их как потенциальную замену (или альтернативу) цитированию. Альтметрики могли бы представлять собой интересное дополнение к цитированию, особенно для того, чтобы информировать о других видах воздействия (например, социальном или культурном) и особенно в тех областях, где они имеют более важное значение, а именно в гуманитарных и социальных науках [25].

Помимо этого, разработчики политики в области науки и технологий могут, помимо традиционно используемых показателей оценки исследований, учитывать наличие статей в социальных сетях и показатели альтметрик [24].

В [12, с. 818] метрики использования статьи рекомендуются применять в качестве не альтернативы традиционным библиометрическим показателям цитирования, а как полезное дополнение.

Корреляционный анализ альтметрик в разных случаях может давать различные результаты в зависимости от того, какие метрики и показатели сопоставляются, на основании каких баз данных и платформ они подсчитываются, по каким разделам науки формируются статистические выборки публикаций, в каком режиме доступа находятся эти публикации и т.п. Во множестве случаев найдены положительные корреляции альтметрик и библиометрических показателей, но не более того, поскольку здесь существует много нюансов и в общем виде проблема пока не решена.

В целом научное сообщество считает, что альтметрики не могут стать полноценной заменой традиционным показателям. Далее будет показано, что в отношении вебометрик как другой разновидности сетевых показателей выводы могут быть совершенно другими.

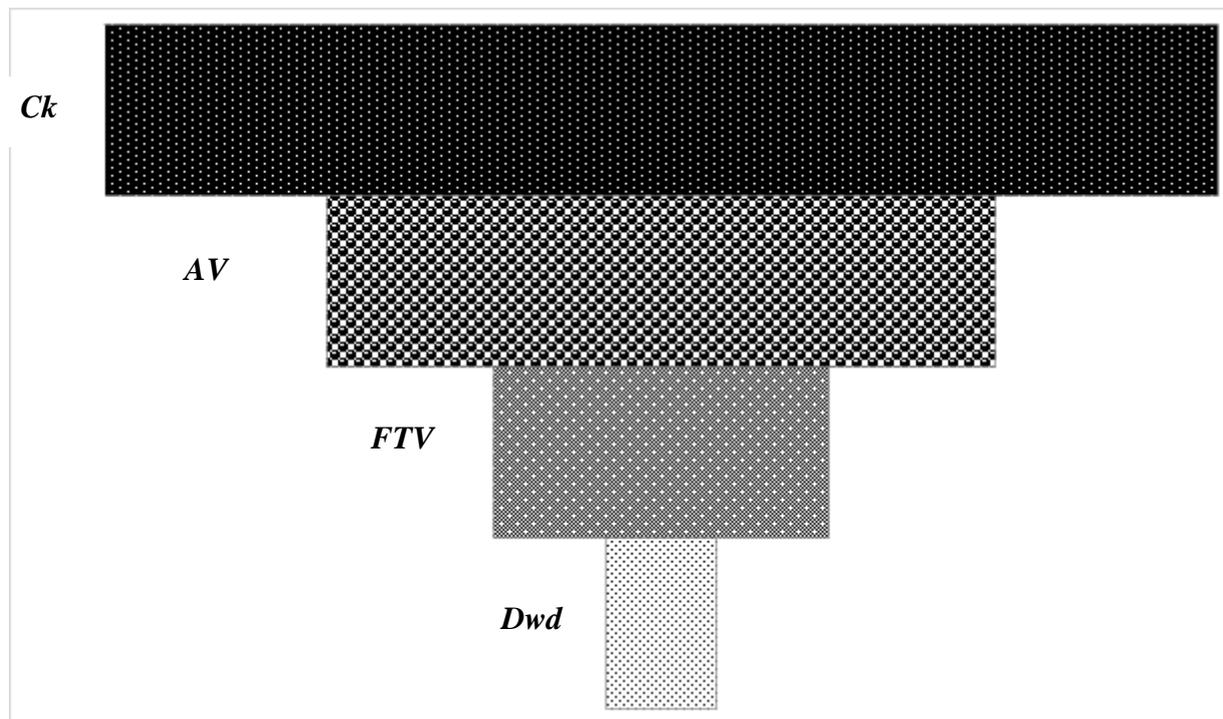
ВИЗУАЛИЗАЦИЯ ВЗАИМОЗАВИСИМОСТИ ПОКАЗАТЕЛЕЙ ИНТЕРЕСА К ЭЛЕКТРОННЫМ ПУБЛИКАЦИЯМ

«Воронкой продаж» (sales funnel) называют статистическую модель, отражающую эффективность любого бизнеса или любого из его уровней в отдельности, например, отдела продаж, почтовых рассылок или всего предприятия [28, с. 207]. Воронка продаж является настолько универсальной моделью, что может использоваться даже для анализа посещаемости сайтов.

Рассмотрим сайт электронной библиотеки или книжного магазина. Воронка продаж для такого сайта показывает, сколько человек зашло на сайт, сколько человек просмотрело описание книги, сколько человек решило купить электронную копию книги. Согласно модели воронки, между этими метриками есть сильная положительная связь: чем больше посетителей сайта, тем больше проданных книг.

Можно предложить альтернативную модель воронки, которую назвать «воронкой вебометрик», напоминающей те воронкообразные диаграммы, которые применяются в управлении продажами или при анализе посещаемости веб-сайтов, однако в ее основе лежат не законы маркетинга, а корреляция между показателями результативности научной деятельности.

На рисунке представлена диаграмма воронки вебометрик электронной публикации.



Воронка вебометрических показателей электронной публикации (где: *Ck* – клики; *AV* – просмотры аннотации; *FTV* – полнотекстовые просмотры; *Dwd* – загрузки [29]).

Ряд значений диаграммы воронки (*funnel chart*) называют «слоем» (*layer*). Так, верхний слой воронки – это «основание» или «голова» (*head*), а нижний – «шея» (*neck*).

В «голове» воронки продаж находятся потенциальные покупатели – «лиды» (*lead*). По мере спуска лиды становятся «горячими», в итоге дело завершается покупкой. Воронка вебометрик никак не связана с продажами. Пожалуй, самой желаемой наградой для автора научной публикации является цитирование его работы другим автором. Но в «шее» воронки вебометрик помещена загрузка файла с копией электронной публикации, поскольку цитирование не относится к вебометрикам, хотя предположительно между ними есть положительная связь. Другими словами, чем чаще читатели просматривают текст или сохраняют файл с публикацией, тем выше вероятность того, что электронная публикация будет ими процитирована.

В воронку вебометрик не должны попадать «сложные» показатели, такие как индикатор глубины просмотра или вебометрический рейтинг. Тем не менее, «простые» метрики воронки могут быть дополнены параметрами, вносящими ограничения. Так, возможна модификация вебометрики без «самопросмотров», когда не учитываются просмотры, сделанные автором или авторами данной публикации. Впрочем, здесь имеет значение пользовательский веб-интерфейс цифровой платформы, который решительным образом определяет состав показателей воронки вебометрик.

Воронка продаж интернет-магазина может быть представлена следующей моделью [30]:

$$R = T \cdot \prod_{i=1}^m conv_i \cdot AOV, \quad (5)$$

где: R – выручка (*revenue*);

T – трафик (*traffic*), т.е. количество посетителей сайта;

$conv_i$ – показатель конверсии i -го этапа продаж, $conv_i \leq 1$;

m – количество этапов продаж; AOV – средний чек (*average order value*).

Применительно к воронке вебометрик, состоящей из m слоев, имеется следующая статистическая зависимость:

$$W_{i+1} = conv_i \cdot W_i, \quad (6)$$

где: W_i – вебометрика i -го слоя воронки;

$conv_i$ – коэффициент конверсии между i -ым и $i + 1$ -м слоями воронки, $conv_i \leq 1$.

Например, W_i может быть количеством AV просмотров аннотации, а W_{i+1} – количеством FTV полнотекстовых просмотров электронной публикации.

Воронку можно считать успешной или эффективной, если она характеризуется высокими показателями конверсии, близкими к единице.

Слабое место модели воронки в том, что вебометрики не всегда накапливаются кумулятивно. Так, согласно базовому сценарию, пользователь должен кликнуть по ссылке, открыв карточку электронной

публикации, изучить аннотацию, затем просмотреть полный текст и, наконец, загрузить файл. Однако эта цепочка действий может не только быть прервана в любой момент, но и произойти по иному сценарию. Например, когда пользователь не стал смотреть аннотацию, а сразу зашел по прямой ссылке и просмотрел полный текст или осуществил загрузку файла к себе на компьютер. При этом нельзя гарантировать, что количество полнотекстовых просмотров или загрузок окажется меньше количества просмотров аннотаций, хотя из общих соображений очевидно, что аннотации просматриваются чаще, чем полные тексты.

Научная публикация – это обнародованный продукт научной деятельности, к которому предъявляются требования уникальности или, по крайней мере, новизны [31]. Подобно тому, как владелец интернет-магазина собирает информацию о конверсии товаров со своей электронной витрины, воронка вебометрик формируется по публикациям в открытом доступе.

По своей популярности у читательской аудитории различаются не только авторские персоны, но и разные проекты одних и тех же авторов. При этом научные проекты служат основаниями для публикаций и указываются в статьях как источники финансирования исследований.

Как правило, цифровые платформы позволяют строить воронки по отдельным публикациям либо по авторским профайлам. Тем не менее, полезно было бы формировать сводные воронки по совокупности публикаций автора или одного из его проектов, сгруппировав данные, размещенные на выбранной платформе.

Задача более высокого порядка – это построение воронки, исходные данные которой komponуются сразу из нескольких платформ. Таким образом, возможность формировать воронки с различными параметрами, в том числе группировками и источниками данных, могла бы стать ценным инструментом для вебометрического анализа.

Конечно, визуализацию в виде горизонтальной гистограммы с накоплением можно заменить на круговую диаграмму, нормируя значения рядов воронки, чтобы они в сумме давали 100%, однако часть информации при этом потеряется. Таким образом, представление вебометрик электронной публикации в виде воронки это не единственно возможный способ визуализации.

Коэффициенты конверсии воронок продаж обычно повышают различными маркетинговыми ходами, например, поисковой оптимизацией сайтов интернет-магазинов, поиском целевых аудиторий, внедрением систем лояльности к клиентам и пр. [32, с. 251-252]. В свою очередь, коэффициенты конверсии воронки вебометрик могут быть оптимизированы доступными для авторов публикаций способами, но при этом без нарушения границ научной этики.

Помимо этого, анализ воронки вебометрик позволяет принимать решения более высокого уровня. Так, воронка вебометрик с группировкой по разным источникам данных позволяет автору определиться, стоит ли ему вообще в дальнейшем размещать свои

публикации в открытом доступе, и какая электронная платформа лучше подойдет.

Таким образом, воронка вебометрик может применяться как аналитический инструмент повышения цитируемости публикаций.

ЗАКЛЮЧЕНИЕ

Возникновение вебометрик связано с развитием и распространением Интернета. Вебометрические показатели активно используются для сбора информации о посещаемости сайтов в коммерческих целях. Однако в научном мире вебометрические показатели применяются, как правило, для рейтингования сайтов образовательных и научных организаций.

Альтметрики появились позже вебометрик, благодаря зарождению социальных сетей. Альтметрики достаточно разнообразны и, как правило, эмоционально окрашены. Несмотря на успехи в области анализа тональности текстов, эффективные инструменты для распознавания тональности альтметрики отсутствуют, что является одним из сдерживающих факторов для внедрения этих индикаторов в оценку результатов научной деятельности. Многие исследователи отмечают, что альтметрики являются полезным инструментом оценки научной деятельности, однако большинство авторов убеждены, что альтметрики не способны заменить цитирования.

Исследования корреляции сетевых показателей с традиционно используемыми выполнялись многократно. Подтверждены положительные корреляции, которые различаются по силе для различных научных направлений. Помимо этого, на наличие и силу корреляции влияет множество других параметров, таких как состав анализируемых показателей и цифровые платформы, с которых снимаются эти метрики.

Выявление хотя и отличающихся по силе, но положительных корреляций традиционно используемых и сетевых показателей позволило нам предложить модель воронки вебометрических показателей. Воронка вебометрик предназначена для анализа популярности электронных научных публикаций. По аналогии с повышением конверсии продаж в бизнесе, воронка вебометрик является средством визуализации интереса к электронным публикациям, которое позволяет принимать решения, способствующие повышению цитируемости работ авторов научных публикаций. Вебометрики результатов поисковых запросов можно использовать и в других задачах, например, для поиска пробелов в знаниях или определения тенденций развития научных направлений.

Таким образом, в то время как переход на показатели альтметрики в управлении научной деятельностью вызывает серьезные сомнения, потенциал комплексного использования простых вебометрических показателей в сочетании с показателями других типов представляется не в полной мере раскрытым.

СПИСОК ЛИТЕРАТУРЫ

1. Гиляревский Р.С. Библиометрия как понимание социальных закономерностей научных коммуникаций // Труды Санкт-Петербургского госу-

дарственного института культуры. – 2022. – Т. 226. – С. 15-23.

2. Мицкевич А. К вопросу о сущности и истоках политической медиаметрии // Научные труды Республиканского института высшей школы. Философско-гуманитарные науки. – 2017. – № 16. – С. 109-121.
3. Ingwersen P. Webometrics – ten years of expansion // International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy (France), May 10-12. – 2006.
4. Гулин К.А., Скородумов П.В. Интернет-портал как средство популяризации деятельности научной организации // Проблемы развития территории. – 2015. – № 5(79). – С. 52-65.
5. Редькина Н.С., Баженов С.Р., Балуткина Н.А., Паршиков Р.М., Рыхторова А.Е., Стукалова А.А., Ударцева О.М., Шевченко Л.Б. Сервис-ориентированные информационные технологии в библиотечных процессах и научной коммуникации // Труды ГПНТБ СО РАН. – 2021. – № 4(12). – С. 37-52. DOI: 10.20913/2618-7575-2021-4-37-52.
6. Ударцева О.М. Анализ популярности библиотечных сайтов с использованием систем сбора статистических данных // Ученые записки (Алтайская государственная академия культуры и искусств). – 2019. – № 3(21). – С. 87-91. DOI: 10.32340/2414-9101-2019-3-87-91.
7. Поляк Ю.Е. О точности вебометрических инструментов. – 2015. – С. 328-331. – URL: <https://www.keldysh.ru/abrau/2015/328.pdf>.
8. Кузнецова А.Д. Обзор вебометрических подходов к изучению деятельности университетов // Интернет и современное общество: Сборник тезисов докладов XX Международной объединенной научной конференции, Санкт-Петербург, 21–23 июня 2017 г. / Университет ИТМО. – Санкт-Петербург: Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2017. – С. 70-71.
9. Борисова И.И. Рейтинг как инструмент повышения конкурентоспособности вузов. – 75 с. – URL: http://www.unn.ru/pages/e-library/methodmaterial/files/9_borisova_2012_migr.pdf.
10. Бажанов Р.С. Основные показатели инструментов веб-аналитики как основа для проведения веб-анализа // Альманах современной науки и образования. – 2014. – № 7(85). – С. 27-30.
11. Канн С.К. Особенности транспорта библиотечной информации в веб-пространстве // Материалы научно-практического семинара «Тенденции развития библиотек в современном обществе: перспективы, возможности, реальность» (г. Новосибирск, 24–25 октября 2018 г.). – Новосибирск: Новосибирский государственный педагогический университет, 2018. – С. 51-62.
12. Маркусова В.А., Миндели Л.Э., Богоров В.Г., Либкинд А.Н. Показатель альтметрики как один из индикаторов научного влияния публикации // Вестник Российской академии наук. –

2018. – Т. 88, № 9. – С. 811-818. DOI: 10.31857/S086958730001694-1.
13. Thelwall M. Big data and social web research methods. – 2018. – 142 p. – URL: <http://www.scit.wlv.ac.uk/~cm1993/papers/IntroductionToWebometricsAndSocialWebAnalysis.pdf>
 14. Чеснялис П.А. Использование альтметрик в справочно-библиографическом обслуживании // Труды ГПНТБ СО РАН. – 2020. – № 1(5). – С. 79-85. DOI: 10.20913/2618-7515-2020-1-79-85.
 15. Юревич М.А., Цапенко И.П. Перспективы применения альтметрики в социогуманитарных науках // Информационное общество. – 2015. – № 4. – С. 9-16.
 16. Бусыгина Т.В. Альтметрия как комплекс новых инструментов для оценки продуктов научной деятельности // Идеи и идеалы. – 2016. – Т. 2, № 2(28). – С. 79-87. DOI: 10.17212/2075-0862-2016-2.2-79-87.
 17. Karanatsiou D., Misirlis N., Vlachopoulou M. Bibliometrics and altmetrics literature review: performance indicators and comparison analysis // Performance Measurement and Metrics. – 2017. – Vol. 18, № 1. – P. 16–27. DOI: 10.1108/pmm-08-2016-0036.
 18. Чеботарева Е.Э. Научные исследования в контексте цифровой экономики // International Journal of Open Information Technologies. – 2017. – Vol. 5, № 11. – P. 36-41.
 19. Bornmann L., Haunschild R. Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data // PLOS ONE. – 2018. – Vol. 13, № 5. DOI: 10.1371/journal.pone.0197133.
 20. Yang S., Zheng M. Performance of citations and altmetrics in the social sciences and humanities // Proceedings of the Association for Information Science and Technology. – 2019. – Vol. 56, № 1. – P. 326-335.
 21. Ouchi A., Saberi M.R., Ansari N.A., Hashempour L., Isfandyari-Moghaddam A. Do altmetrics correlate with citations? A study based on the 1,000 most-cited articles // Information Discovery and Delivery. – 2019. – Vol. 47, № 4. – P. 192–202. DOI: 10.1108/idd-07-2019-0050.
 22. Syamili C., Rekha R.V. Do altmetric correlate with citation?: A study based on PLOS ONE journal // COLLNET Journal of Scientometrics and Information Management. – 2017. – Vol. 11, № 1. – P. 103-117. DOI: 10.1080/09737766.2016.1260815.
 23. Gul S., Shueb S., Shah R., Shah T.A. Altmetrics for the journals of politics: correlating altmetrics with journal metrics // 2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS). – IEEE. – 2018. – P. 397-401.
 24. Sedighi M. Altmetrics analysis of selected articles in the field of social sciences // Global Knowledge, Memory and Communication. – 2022. – №. ahead-of-print. DOI: 10.1108/GKMC-07-2021-0124.
 25. Costas R., Zahedi Z., Wouters P. Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective // Journal of the Association for Information Science and Technology. – 2015. – Vol. 66, № 10. – P. 2003-2019. DOI: 10.1002/asi.23309.
 26. Артамонова А.С., Третьякова О.В. К вопросу об альтернативных способах оценки научных результатов // Социальное пространство. – 2020. – Т. 6, № 5. – С. 1-14. DOI: 10.15838/sa.2020.5.27.7.
 27. Zhao R., Wang X., Liu Z., Qi Y., Zhang Z., Chang R. Research on the impact evaluation of academic journals based on altmetrics and citation indicators // Proceedings of the Association for Information Science and Technology. – 2019. – Vol. 56, № 1. – P. 336-345. DOI: 10.1002/pr2.00027.
 28. Соломатин М.С., Сайбель Н.Ю. Использование воронки продаж в бизнесе // Сб. статей Международной научно-практической конференции «Проблемы формирования единого научного пространства» (г. Волгоград, 05 мая 2017 г.). В 4-х частях. Часть 1. – Волгоград: Общество с ограниченной ответственностью "Аэтерна", 2017. – С. 206-208.
 29. Periodic table of scientometric indicators. – URL: <https://mylibrarianship.wordpress.com/2022/05/26/periodic-table-of-scientometric-indicators>
 30. Татаровский Ю.А. Моделирование в анализе продаж: эволюция подходов // Сборник материалов XVIII Международной научно-практической конференции «Наука, образование, общество: тенденции и перспективы развития» (г. Чебоксары, 11 мая 2020 г.). – Чебоксары: Общество с ограниченной ответственностью "Центр научного сотрудничества "Интерактив плюс", 2020. – С. 207-210. DOI: 10.21661/r-530869.
 31. Нестеров А.В. О научных произведениях и продуктах. – Москва: НИУ ВШЭ, препринт, ноябрь 2014. – 14 с. – URL: <http://a-podkidyshev.ru/MinON/nauka/Proizvedeniya--nauki--01.pdf>
 32. Парфенова Е.Н., Казлитина Е.В. Повышение эффективности цифровой воронки продаж как современного механизма построения результативной маркетинговой стратегии // Сборник материалов 5-й Всероссийской научно-практической конференции «Российские регионы как центры развития в современном социокультурном пространстве» (г. Курск, 25 октября 2019 г.). – Курск: Юго-Западный государственный университет, 2019. – С. 250-253.

Материал поступил в редакцию 06.06.23.

Сведения об авторе

КАЛАЧИХИН Павел Андреевич – кандидат экономических наук, ведущий научный сотрудник ВИНТИ РАН, Москва
e-mail: pakalachikhin@viniti.ru

В.Н. Комарица

Анализ ключевых слов в научных статьях

Рассматриваются методы извлечения ключевых слов и реферирования научного текста; показано, что автоматизированные методы реферирования обеспечивают наиболее точную передачу содержания текста и не требуют дополнительного использования лексических средств.

Описывается подготовка корпусов текстов для анализа. Ключевые слова, полученные статистическим методом, сопоставляются с ключевыми словами, выделенными авторами.

*Основным критерием для определения ключевых слов стала частотность их появления в тексте, которая вычислялась с помощью семантико-статистического анализа текста *advego.com*.*

Установлено, что все ключевые слова имеют устойчивые дефиниции и на 96% соответствуют терминологии отрасли; ключевые слова, выделенные авторами, имеют до 30% несовпадений с ключевыми словами, полученными статистическим методом.

Ключевые слова: *ключевые слова, термины, корпуса текста, смысл текста, семантическая структура текста*

DOI: 10.36535/0548-0019-2023-09-2

ВВЕДЕНИЕ

Определяющее значение для систематизации, классификации и оптимизации поиска информации имеют ключевые слова [1, 2]. Различают как отдельные ключевые слова, так и их совокупности. Набор ключевых слов позволяет получать высокоуровневое и компактное описание содержания научной публикации [3–7].

Ключевые слова характеризуются тем, что:

- являются наиболее употребительными (частотными) словами (терминами);
- представлены значимой лексикой, достаточно обобщены по своей семантике (средней степени абстракции), стилистически нейтральны, не оценочны;
- связаны друг с другом сетью семантических связей, пересечением значений;
- более половины тематического компонента текста составляют ключевые слова, а минимальный набор ключевых слов приближается к инварианту его содержания при их логическом упорядочивании;
- набор из 5–15 или 8–10 ключевых слов соответствует объёму оперативной памяти человека;
- при частом повторении ключевых слов в тексте поисковые системы могут расценивать ресурс как спам и не показывать его в строке результатов поиска;

- набор ключевых слов определяют контексты слов, обладающих максимальной предсказуемостью [2].

Ключевые слова являются обозначенной понятийной текстовой единицей, которая как самостоятельно, так и в словосочетаниях образует как самостоятельные, так и группы терминов.

Постоянные изменения, происходящие в научном дискурсе, создают потребность в выделении терминов из научных текстов, в их систематизации и включении в специализированные научно-технические словари [8–14].

Существующие методы и алгоритмы извлечения ключевых слов из научных текстов разделяют на семантические и статистические. В работах [5, 6; 15–18] такие методы рассматриваются с условием адаптации к каждой конкретной ситуации. Семантический подход основан на установлении логико-семантической связи понятий в заданной предметной области. Статистический подход рассматривает использование специальных формул для выявления частотности употребления слов в тексте [3].

Автоматическое (статистическое) извлечение ключевых слов представляется перспективным направлением исследований в области анализа текста, обработки естественного языка и поиска информации.

Ключевые слова представляют текстовые документы в компактном виде, что позволяет выполнять их автоматическое индексирование, суммирование и классификацию [19]. В тексте ключевые слова образуют сеть логико-семантических связей, вершины графа которой соответствуют объектам предметной области, а дуги задают отношения между ними. Объектами могут быть: понятия, события, свойства, процессы [1, 20]. Алгоритмы и методы автоматизированного подбора ключевых слов, основанные на вычислении метрик, сортировке слов по важности для документа или коллекции документов позволяют определять группу ключевых слов, наиболее полно характеризующую анализируемый текст [21–24]. Автоматическое реферирование основано на сжатии заданного текста путем выбора предложений с ключевыми словами, которые содержат наиболее важную информацию [25, 26].

Для извлечения ключевых слов из текста алгоритм GenEx, основанный на машинном обучении, предложен П.Д. Терни [27].

Традиционно авторы выделяют ключевые слова на основе индивидуальных предпочтений и имеющегося практического опыта работы в предметной области.

Одна из задач настоящего исследования – сопоставление ключевых слов, подобранных автором, с ключевыми словами, выделенными с применением статистических методов.

По результатам изучения методик извлечения ключевых слов и реферирования научных текстов было установлено, что использование автоматизированных методов обеспечивает наилучшее сжатие со-

держания и максимально точное описание смысла реферируемого научного текста с минимальным применением лексических средств.

ЗАДАЧИ И ПРАКТИЧЕСКАЯ ЗНАЧИМОСТЬ ИССЛЕДОВАНИЯ

Цель настоящего исследования – определить состав, особенности и отличие ключевых слов, выделенных из текста авторами, от ключевых слов, полученных автоматизированным путем. Для этого необходимо:

- подготовить корпус текстов на базе научных статей по тематике журнала;
- выявить ключевые слова, определить частотность их употребления в содержательно-смысловом блоке основного корпуса текста (тематические рубрики);
- визуализировать полученные результаты;
- определить значимость терминов в тексте.

Схема сопоставления ключевых слов, полученных автоматизированным и семантическим способами, представлена на рис. 1.

В настоящей работе были использованы:

- 1) теоретико-аналитический метод, включающий обзор и анализ научной литературы;
- 2) метод сплошной выборки: сбор корпуса текстов по определенной теме без установки дополнительных параметров отбора;
- 3) метод сопоставительного и структурного анализа: распределение слов в порядке убывания частоты.

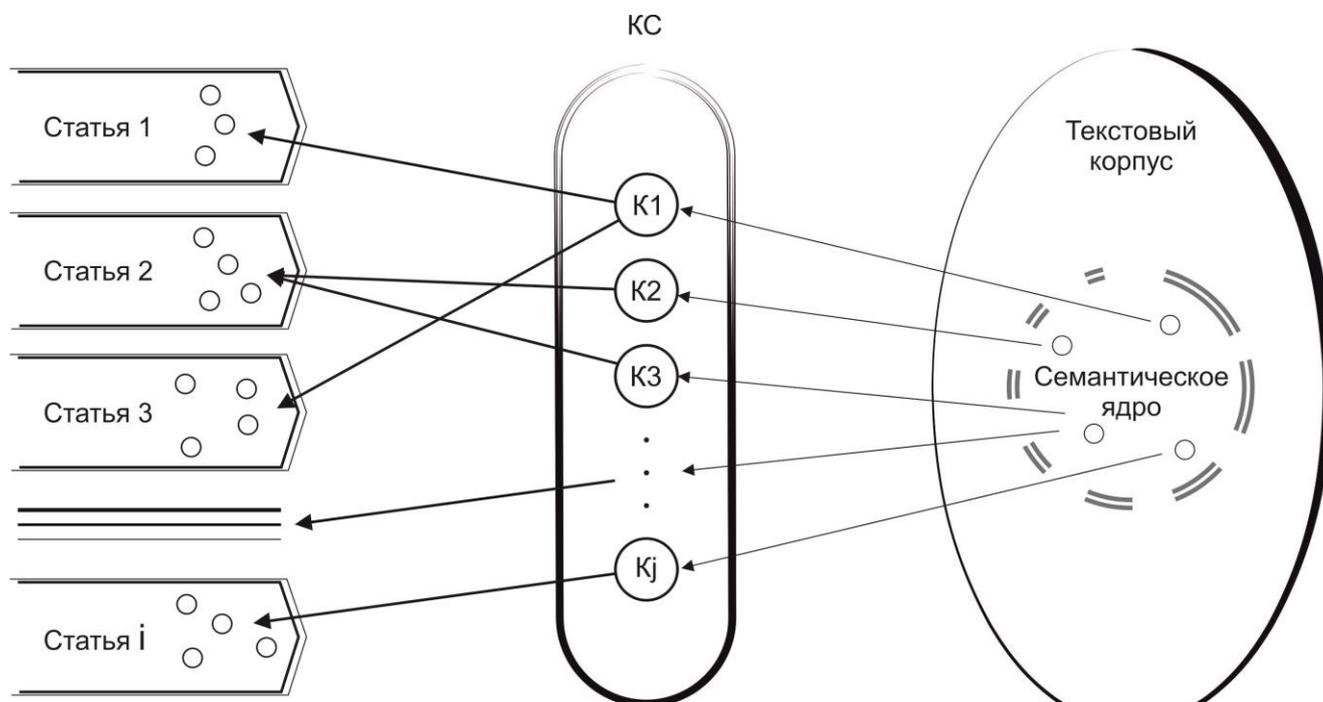


Рис. 1. Схема сопоставления ключевых слов, полученных автоматизированным и семантическим способами

Материалом для нашего исследования стали статьи журнала «Трубопроводный транспорт нефти и нефтепродуктов» за 2020–2022 гг. На первом этапе было сформировано 19 корпусов, в которые были включены тексты без предварительной обработки методом сплошной выборки:

- основной корпус текста содержит все статьи за 2020–2022 гг. – 916 275 словоупотреблений (табл. 1);
- корпус текста, содержащий авторские ключевые слова всех статей, – 1 318 словоупотреблений;
- 17 корпусов текста, составленных с учетом распределения статей по тематическим рубрикам журнала (табл. 2).

Отобранные 19 корпусов текста были рассчитаны с помощью аналитического сервиса advego.com. В результате выявлены наиболее употребляемые слова и рассчитаны показатели их частотности – отношение количества употреблений отдельного слова к общему количеству слов в тексте. В табл. 3 и 4 пред-

ставлены выделенные расчётным путем и отсортированные в порядке убывания показателя частотности ключевые слова. На рис. 2 графически показаны распределения частот и словоупотреблений в основном корпусе текста.

Анализ данных табл. 3 и 4 показывает, что ключевые слова, имеющие высокий уровень распространенности, являются однокомпонентными и в целом характеризуют рассматриваемую предметную область.

В результате статистической обработки основного корпуса текста было получено 1 318 ключевых слов, что составляет 0,14% от общего количества словоупотреблений в основном корпусе текста. Проверка выделенных ключевых слов показала, что 96% слов соответствует отраслевой терминологии – это свидетельствует об эффективности использования статистического метода для извлечения ключевых слов из текста.

Таблица 1

Параметры основного корпуса текста журнала (статьи 2020–2022 гг.)

Год / выпуск журнала		Общее число словоупотреблений	Число	
			слов в семантическом ядре	авторских ключевых слов
2020	1	52 493	12 806	70
	2	52 110	12 506	69
	3	51 673	12 402	62
	4	49 320	11 837	81
	5	50 585	12 140	75
	6	54 150	12 996	65
	∑	304 144	74 687	422
2021	1	51 826	12 438	77
	2	56 366	13 528	83
	3	50 690	12 166	77
	4	49 471	11 873	79
	5	52 185	12 524	103
	6	51 623	12 390	77
	∑	309 285	74 919	496
2022	1	56 900	13 656	73
	2	50 474	12 114	53
	3	52 366	12 568	71
	4	45 015	11 039	63
	5	53 450	12 828	69
	6	44 641	11 238	77
	∑	302 846	73 443	406
Всего:		916 275	223 049	1 324

Таблица 2

Матрица распределения статей по разделам (R) журнала

Раздел	2020					2021					2022					Итого				
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	∑	%
R(1)	2	3	2	2	1	2	2	3		2	1	1	2	2	2	3		2	32	17
R(2)	2	3	1	2	2	2	6	2	5	3	6	4		2	1	2	4	3	50	25
R(3)	1		1							1	1		1		1		1	1	8	4
R(4)			1	1			1				2		1			1	1		8	4
R(5)	1	1	1				1		1	1		1	1		3	1	2		14	8
R(6)								1	1				1						3	2

Раздел	2020					2021					2022						Итог			
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	∑	%
R(7)		1		2	2		1	3	1				1		2		1		14	7
R(8)		1		1					1								2		5	3
R(9)								1	2	1	1	1					1		7	3
R(10)	1			1	2		1	2		2	1		2	1		1		14	7	
R(11)	1	1		1		2				1								6	3	
R(12)	2		1		2	3					1	2	1	1		1		14	7	
R(13)	1	1	3	2	1	1			2	1	1			1			1	15	8	
R(14)													1					1	1	
R(15)														1				1	1	
R(16)					1													1	1	
R(17)							1	1										2	1	

Таблица 3

Частотный список распределения ключевых слов по тематическим разделам журнала

Раздел	Авторские	Расчетные
R(1)	трубопровод, напряжение, разрушение, состояние, изгиб	трубопровод, напряжение, дефект, изгиб, давление
R(2)	нефть, грунт, трубопровод, резервуар, модель	нефть, грунт, трубопровод, температура расчет
R(3)	сварной, контроль, трубопровод, вертикальный, неразрушающий	сварной, соединение, шов, дефект, зона
R(4)	трубопровод, ремонт, дефект капитальный, надежность	трубопровод, ремонт, участок, метод, работа
R(5)	защита, покрытие, катодная коррозия, электрохимическая	защита, покрытие, испытание, коррозия грунт
R(6)	обеспечение, программное автоматизированная, информационный, контур	система, процесс, анализ, безопасность давление
R(7)	нефть, насос, режим, гидравлический, оптимизация	нефть, насос, работа, режим мощность
R(8)	сертификация, система, диагностика, стандартизации, магистральный	материал, вкладыш, высокий грунт, давление
R(9)	оценка, качество, соответствие арматура, сертификация	оценка, качество, соответствие требование, нефть
R(10)	нефть, качество, измерение, свойство, качество	нефть, качество, вязкость, измерение, метод
R(11)	авария, безопасность, ликвидация, опасный, нефть	авария, безопасность, ликвидация, объект, подача
R(12)	система, управление, сеть, анализ, коэффициент	система, управление, модель, оценка сеть
R(13)	разлив, ликвидация, нефтепродукт, испарение, деградация	нефть, концентрация, нефтепродукт, очистка, топливо
R(14)	автоматизированные, управление, информационный, угрозы, кибератака	асутп, угроза, информационный, сеть безопасность
R(15)	системы, моделирование, расчет проектирование, цифровой	система, проектирование, объект проектирование, труба
R(16)	резервуар, железобетонный, научно-технический, отрасль Главтранснефть	резервуар, нефть, нефтепровод емкость, крыша
R(17)	монополия, характеристики, регулирование, экономика отрасль	государственный, естественный регулирование, монополия, деятельность

Таблица 4

Частотный список распределения ключевых слов в основном корпусе текста

Авторские	Расчетные
нефть, трубопровод, система, резервуар магистральный	нефть, трубопровод, система, грунт, покрытие

ВИЗУАЛИЗАЦИЯ РЕЗУЛЬТАТОВ ЧАСТОТНОГО АНАЛИЗА – ПОСТРОЕНИЕ ОБЛАКА КЛЮЧЕВЫХ СЛОВ

Для наглядного представления распределения частотности ключевых слов в корпусе текста мы использовали элемент визуального дизайна – органическую форму в виде облака. Криволинейные формы облака задаются текстом различного размера, плотности цвета и направленности. Размер шрифта написания слова в облаке рассчитывается из условия частоты использования слова в тексте по формуле:

$$S_i = \frac{f_{max}(t_i - t_{min})}{t_{max} - t_{min}}, t_i > t_{min}; S_i = 1,$$

где: S_i – размер шрифта, f_{max} – максимальный размер шрифта, t_i – вес слова (термина),
 t_{min} – минимальный вес слова (термина),
 t_{max} – максимальный вес слова (термина).

Результаты подсчета параметров размера шрифта ключевых слов основного корпуса текста по этой формуле мы оформили в виде облака ключевых слов (рис. 3).

По получившемуся облаку ключевых слов (см. рис. 3) можно видеть, что наиболее значимые в тексте ключевые слова – нефть, трубопровод и система. Применение этих ключевых слов наиболее точно задает принадлежность текста к предметной области – трубопроводный транспорт нефти.

ВЫВОДЫ

Сопоставление ключевых слов, выделенных авторами из текста, с ключевыми словами, полученными с применением статистических методов, показало:

1) ключевые слова, выделенные авторами и полученные статистическим методом, имеют устойчивые дефиниции и на 96% соответствуют отраслевой терминологии;

2) ключевые слова, выделенные авторами, имеют до 30% несовпадения с полученными статистическим методом;

3) использование ключевых слов обеспечивает соответствие реферата содержанию статьи.

Результаты авторского и статистического методов выделения ключевых слов из текста статьи показывают, что ключевые слова: нефть, трубопровод и система – имеют максимальные показатели частоты употребления в статьях журнала «Трубопроводный транспорт нефти и нефтепродуктов». Установлено, что включение в общий объем статьи 0,14% терминов характеризует текст как специализированный. Перспективу дальнейшего исследования мы видим в более подробном анализе интеграции авторского и статистического методов выделения ключевых слов из текста.

СПИСОК ЛИТЕРАТУРЫ

1. Москвитина Т.Н. Ключевые слова и их функции в научном тексте // Вестник Челябинского государственного педагогического университета. – 2009. – № 11-1. – С. 270-283.

2. Роднянский В.Л. О роли ключевых слов в понимании текста // Психолингвистические проблемы семантики и понимания текста. – Калинин, 1986. – С. 106-113.
3. Степанова Д.В., Жданович М.А. Особенности выделения ключевых слов из научных технических текстов предметной области «информационные технологии» // Вестник Минского государственного лингвистического университета. Серия 1: Филология. – 2020. – № 6(109). – С. 98-105.
4. Jinqing Yang, Yi Bu, Wei Lu, Yong Huang, Jiming Hu, Shengzhi Huang, Li Zhang. Identifying keyword sleeping beauties: A perspective on the knowledge diffusion process // Journal of Informetrics. – 2022. – Vol. 16, Issue 1. DOI: <https://doi.org/10.1016/j.joi.2021.101239>.
5. Jinho Choi, Sangyoon Yi, Kun Chang Lee. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution // Information & Management. – 2011. – Vol. 48, Issue 8. – P. 371-381. DOI: <https://doi.org/10.1016/j.im.2011.09.004>.
6. Jordan Pascual Espada, Jaime Solis Martinez, Irene Cid Rico, Luis Emilio Velasco Sanchez. Extracting keywords of educational texts using a novel mechanism based on linguistic approaches and evolutive graphs // Expert Systems with Applications. – 2023. – Vol. 213, Part A. – P. 118842. DOI: <https://doi.org/10.1016/j.eswa.2022.118842>.
7. Шехтман Н.А. Ключевые слова в научном тексте // Научно-техническая информация. Сер. 2. – 2003. – № 10. – С. 35-39.
8. Zhenzhen Xu, Junsheng Zhang. Extracting Keywords from Texts based on Word Frequency and Association Features // Procedia Computer Science. – 2021. – Vol. 187. – P. 77-82. DOI: doi.org/10.1016/j.procs.2021.04.035.
9. Иконникова В.А., Цверкун Ю.Б. Лингвокультурологические исследования англоязычных отраслевых терминологий и современные технологии в лингвистике. – Москва: МГИМО, 2023. – 186 с.
10. Багана Ж., Таранова Е.Н. Терминообразование в языке науки (Серия: Научная мысль. Лингвистика). – Москва: ИНФРА-М, 2012. – 142 с.
11. Екимов М.А. Особенности подбора ключевых слов в русскоязычных научных статьях (на материале журналов по истории из списка ВАК) // Вестник магистратуры. – 2020. – № 1-5(100). – С. 65-71.
12. Митин Н.А., Орлов Ю.Н. Статистический индикатор тематики научных текстов. Проектирование будущего // Проблемы цифровой реальности. – 2018. – № 1(1). – С. 74-80.
13. Комарица В.Н. Информационный и терминологический анализ текстов авторефератов диссертаций (на примере предметной области – трубопроводный транспорт углеводородов) // Научно-техническая информация. Сер. 1. – 2020. – № 1. – С. 29-33.
14. Сощенко А.Е., Комарица В.Н. Анализ зависимости между числом публикаций и количе-

- ством цитирования статей в научной периодике трубопроводного транспорта углеводородов // Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. – 2015. – № 3(19). – С. 108-115.
15. Башмакова А.Ю. Выявление ключевых слов тематического поля «образование/education». Филологические науки // Вопросы теории и практики. – 2021. – Т. 14, № 12. – С. 3986-3990.
 16. Zara Nasar, Syed Waqar Jaffry, Muhammad Kamran Malik. Textual keyword extraction and summarization: State-of-the-art // Information Processing & Management. – 2019. – Vol. 56, Issue 6. – P. 102088. DOI: doi.org/10.1016/j.ipm.2019.102088.
 17. Zeng P., Tan Q., Yan Y., Xie Q., Xu J., Cao W. Automatic keyword extraction using word embedding and clustering // International Conference on Computer Systems, Electronics and Control (ICCSEC). – 2017. – P. 1402–1408. DOI: 10.1109/ICCSEC.2017.8447033.
 18. Ying Yang, Yaru Sun, Dawei Yang, Guang Mei, Zhihong Wang. Empower Keywords Generation for Short Texts with Graph-to-Sequence Learning // Procedia Computer Science. – 2022. – Vol. 207. – P. 100-109. DOI: doi.org/10.1016/j.procs.2022.09.042.
 19. Алексеенко А.А., Глазкова А.В. Разработка приложения для подбора ключевых слов к научным текстам // Материалы Всероссийской конференции молодых ученых «Математическое и информационное моделирование». – Тюмень, 2022. – С. 250-256.
 20. Aytug Onan, Serdar Korukoglu, Hasan Bulut. Ensemble of keyword extraction methods and classifiers in text classification // Expert Systems with Applications. – 2016. – Vol. 5. – P. 232-247. DOI: doi.org/10.1016/j.eswa.2016.03.045.
 21. Arjun Duvvuru, Srinivasan Radhakrishnan, Deepali More, Sagar Kamarthi, Sivarit Sultornsane. Analyzing Structural & Temporal Characteristics of Keyword System in Academic Research Articles // Procedia Computer Science. – 2013. – Vol. 20. – P. 439-445. DOI: doi.org/10.1016/j.procs.2013.09.300.
 22. Кретов А.А. Маркеры и ключевые слова в научных текстах // Мир лингвистики и коммуникации: электронный научный журнал. – 2012. – № 27. – С. 1-13.
 23. Москвитина Т.Н. Методы выделения ключевых слов при реферировании научного текста // Вестник Томского государственного педагогического университета. – 2018. – № 8(197). – С. 45-50.
 24. Дубинина Е.Ю. Выделение ключевых слов текста научной статьи в процессе создания автоматического реферата // Вестник Воронежского государственного университета. Серия: Филология. Журналистика. – 2020. – № 1. – С. 26-28.
 25. Камшилова О.Н. Малые формы научного текста: ключевые слова и аннотация (информационный аспект) // Известия Российского государственного педагогического университета им. А.И. Герцена. – 2013. – № 156. – С. 106-117.
 26. Батура Т.В., Бакиева А.М. Система авто-реферирования научно-технических текстов // Научное программное обеспечение. Труды семинара. – Новосибирск, 2019. – С. 47-52.
 27. Turney P.D. Learning to extract keyphrases from text // CoRR cs.LG/0212013. – 2002. – URL: <https://arxiv.org/abs/cs/0212013> (дата доступа: 22.05.2023 г.).

Материал поступил в редакцию 08.06.23.

Сведения об авторе

КОМАРИЦА Валентин Николаевич – кандидат технических наук, заместитель начальника отдела издательских проектов и медиакоммуникаций, ООО «Научно-исследовательский институт трубопроводного транспорта» (ООО «НИИ Транснефть»), Москва e-mail: KomaritsaVN@niitnn.transneft.ru

Инструменты поддержки исследовательских процессов открытой науки

Представлены результаты анализа инструментов поддержки научных исследований с целью построения системы таких открытых инструментов на онлайн-платформе «Библиотека для открытой науки». Показано, что использование платформ и инструментов на различных этапах научного процесса позволит следовать принципам открытости и доступности, повысить результативность научной работы, сократить время на поиск необходимой информации.

Ключевые слова: открытая наука, инструменты для поддержки исследовательских процессов, принципы открытой науки, научный процесс

DOI: 10.36535/0548-0019-2023-09-3

ВВЕДЕНИЕ

Открытая наука имеет ряд принципов и инструментов, направленных на то, чтобы научные исследования были как можно более прозрачны, доступны, достоверны и воспроизводимы. Существует множество онлайн-инструментов для поддержки исследовательских процессов. Использование подобных инструментов позволяет ученым повышать эффективность своей работы, сокращать время на поиск необходимой информации, делать исследовательские рабочие процессы более открытыми и достоверными. Сегодня в научных исследованиях широко используются онлайн-платформы и сервисы.

Цель настоящего исследования – изучение онлайн-инструментов для поддержки исследовательских процессов открытой науки. В дальнейшем мы полагаем построить систему открытых инструментов для поддержки научных исследований в ГПНТБ СО РАН на онлайн-платформе «Библиотека для открытой науки».

ОБЗОР ЛИТЕРАТУРЫ ПО ПРОБЛЕМЕ

Движение за внедрение открытой науки смещает способы проведения научных исследований в сторону более открытых, прозрачных и совместных практик [1, 2]. Открытая наука включает набор действий, принципов и инструментов, направленных на то, чтобы сделать научные исследования доступными, повысить их эффективность. Использование учеными различных платформ для совместной работы меняет характер исследовательского процесса, объединяет географически «разбросанных» специалистов для сотрудничества, а также для обмена объектами, идеями и опытом [3, 4]. Открытые хранилища данных, платформы рецензирования, серверы препринтов и сервисы для предварительной регистрации протоколов исследований помогают решать вопросы кризиса репликации, достоверности, установления

приоритета открытий и идей [5-10]. Но не все специалисты знакомы с этими возможностями и платформами [11, 12]. Многие из них не знают о новых технологиях и стандартах в области открытых данных. Простого размещения файлов в общедоступном репозитории недостаточно: и данные, и их коды должны быть описаны, чтобы можно было понять, как они организованы и как запускать код [13].

В профессиональной литературе отмечается, что часто исследователи тратят много времени на сбор данных и их структурирование вместо того, чтобы использовать это время на сам процесс исследования. Если руководствоваться принципами FAIR¹, то эта проблема станет менее значимой, использование различных платформ и инструментов позволит следовать принципам открытости эффективнее [14–17]. Так, R. Schonfeld утверждает, что ученые могут повысить эффективность своей деятельности, используя платформы и системы, поддерживающие их исследовательский рабочий процесс [18].

Для оценки владения принципами FAIR можно использовать инструмент FAIR-Aware – онлайн-инструмент, который помогает исследователям выявлять свои знания принципов FAIR и понимать, как сделать набор научных данных более соответствующим этим принципам, прежде чем загружать его в хранилище. FAIR-Aware состоит из 10 вопросов, к каждому из которых предлагается дополнительная информация и даются практические рекомендации для понимания пользователями его принципов [19].

В движение за открытую науку достаточно активно включились библиотеки, в основном в отношении публикаций в открытом доступе. Многие зарубежные библиотеки, особенно университетские, предостав-

¹ FAIR – Findability (находимость), Accessibility (доступность), Interoperability (интероперабельность) и Reusability (повторное использование).

ляют свою инфраструктуру (репозитории) для обмена и хранения данных, поддерживают управление исследовательскими процессами [20, 21]. Некоторые академические библиотеки разрабатывают программы для предоставления набора инструментов и услуг с целью поддержки открытых исследований на протяжении всего их жизненного цикла [2]. В результате анализа сайтов российских вузовских библиотек было выявлено, что в нашей стране, к сожалению, таких систем поддержки практически нет, есть только их разрозненные элементы [22].

Изучение литературы по проблеме показало, что, несмотря на интеграцию цифровых инструментов в практику открытой науки, исследования таких инструментов, поддерживающих открытую науку, немногочисленны. В этом отношении интересен опрос, проведенный библиотекой Утрехтского университета в Нидерландах, об использовании традиционных и онлайн инструментов в научной работе. Его авторы отмечают, что результаты этого опроса могут помочь исследователям в организации своей работы [23].

Наше общение с учеными НИУ СО РАН и проведенное анкетирование показали, что не все исследователи знакомы с технологиями размещения открытых данных, больше половины не знают о принципах FAIR или только слышали о них и нуждаются в помощи в управлении данными и продвижении результатов своих научных работ [24, 25].

МЕТОДИКА И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Сначала мы идентифицировали онлайн-инструменты из целого ряда различных источников. Изучали профессиональную литературу, в которой описывались различные инструменты и сервисы для поиска и сбора данных, обеспечения целевой поддержки управления исследованиями, а также долгосрочного архивирования и публикации данных; результаты опросов по использованию инструментов для поддержки исследовательских рабочих процессов, вебинары для ученых, рекомендации ЮНЕСКО [10, 12, 23, 26–29].

Поле анализа было расширено за счет веб-поиска, просмотра сайтов, предназначенных в помощь ученым, и содержащих различные ресурсы и сервисы для исследователей, где выбирались открытые инструменты [30–32]. Был использован и собственный опыт применения инструментов для поддержки исследовательских процессов.

Инструменты оценивались с позиций открытости, бесплатности и доступности в условиях ограничений (рестрикции). Инструмент должен быть активен на момент создания системы и иметь веб-сайт с описанием функций и условий использования. Из многих инструментов разработанных коммерческими организациями, мы отбирали те, у которых есть бесплатная версия.

Исследовательский процесс станет более эффективным и открытым, если ученые будут применять практики открытой науки и использовать различные инструменты и сервисы.

На этапе планирования исследования ученый может:

- обмениваться своими идеями и подходами на раннем этапе, чтобы получать отзывы научного сообщества;

- просматривать репозитории данных, чтобы выделять информацию, которую можно повторно использовать в ходе исследования;
- проверять гипотезу, предварительно регистрировать исследование;
- оценивать найденные публикации по интересующей теме на основе экспертных оценок;
- создавать совместные проекты.

Для того чтобы находить результаты обсуждения проблемы, ее экспертные оценки, открытые репозитории данных, а также партнеров по исследованиям, на этом этапе можно использовать, например, такие инструменты и платформы:

- Open Science Framework (OSF) – проект программного обеспечения с открытым исходным кодом, который способствует открытому сотрудничеству в научных исследованиях. В качестве инструмента совместной работы OSF помогает исследовательским группам работать над проектами в частном порядке или делает весь проект общедоступным. Есть возможность предварительной регистрации исследований (<https://osf.io/dashboard>);

- AsPredicted – платформа, которая позволяет одним исследователям предварительно регистрировать свои исследования, а другим – читать и оценивать эти предварительные регистрации (<https://aspredicted.org/>);
- Реестр репозитория исследований Re3data (<https://www.re3data.org/>).

На этапе поиска информации ученый может:

- находить публикации/данные/коды;
- получать к ним доступ;
- совместно просматривать публикации;
- открыто их обсуждать и комментировать.

На этом этапе применяются онлайн-сервисы, позволяющие: выявлять необходимую информацию и устанавливать к ней доступ легальными способами; получать оповещения/рекомендации по интересующей тематике; выбирать инструменты, ускоряющие процесс усвоения информации, и программы для управления библиографической информацией.

На активном этапе исследования ученый может:

- обмениваться методами и ранними выводами с помощью препринтов, получать обратную связь от коллег и выявлять ошибки или проблемы до публикации;
- депонировать данные, необходимые для проверки выводов, и программное обеспечение для их анализа или визуализации;

- использовать общую научно-исследовательскую инфраструктуру для работы по общим программам изучения сложных проблем;

- выполнять совместные научные разработки;
- привлекать научных волонтеров.

На этом этапе используются препринты, репозитории, сервисы для проведения экспериментов и опросов, а также лабораторные блокноты, системы для набора текста и подготовки статей, инструменты для визуализации и анализа данных, например:

- R Discovery – инструмент на базе искусственного интеллекта, упрощающий поиск литературы. Его алгоритмы просматривают статьи из Microsoft Academic, PubMed, PubMed Central и

Crossref, чтобы рекомендовать лучшие из них для изучения (<https://discovery.researcher.life/>);

- **BASE** – одна из самых объемных поисковых систем в мире, особенно для академических веб-ресурсов. Бесплатно предоставляет полные тексты около 60% проиндексированных документов открытого доступа (<https://www.base-search.net/>);

- **Unpaywall** – браузерное расширение – выявляет бесплатные копии научных статей, оригиналы которых находятся под paywall (<https://unpaywall.org/>);

- **OpenRefine** – инструмент для работы с «сырыми» данными, их очистки и преобразования из одного формата в другой (<https://openrefine.org/>);

- **Tatool** – пакет программного обеспечения с открытым исходным кодом для создания и проведения компьютерных экспериментов (<http://www.tatool-web.com/#/>).

На завершающем этапе исследования ученый может:

- публиковать в журналах открытого доступа и/или размещать свои статьи в репозитории открытого доступа;

- использовать открытое экспертное рецензирование;

- создавать свой профиль исследователя;

- общаться в научных соцсетях;

- распространять результаты исследования за пределами научной аудитории;

- определять влияние исследователей на результаты научной работы.

На этом этапе применяются инструменты выбора журналов для публикации в открытом доступе, облегчающие процесс публикации, сервисы постпубликационного рецензирования, оценки исследований, научные социальные сети, например:

- **Preprints.org** – бесплатный сервер препринтов, субсидируемый издательством MDPI (Multidisciplinary Digital Publishing Institute), это междисциплинарная платформа для препринтов, которая делает научные рукописи из всех областей исследований доступными для немедленного цитирования и постоянного доступа (<https://www.preprints.org/>);

- **Dissemin** – веб-платформа, собирающая академические публикации и анализирующая их доступность в Интернете, а также раскрывающая политику издателей. Исследователи используют Dissemin, чтобы обеспечить свободный доступ к своим публикациям (<https://dissemin.in/>);

- **Hypothesis** – платформа для обсуждения, комментирования, чтения в социальных сетях, организации исследований и составления личных заметок на веб-страницах, в файлах PDF и EPUB (<https://web.hypothes.is/>);

- **ImpactStory** – веб-инструмент с открытым исходным кодом, предоставляющий альтернативные показатели, помогает исследователям выявлять влияние результатов своих исследований, включая журнальные статьи, сообщения в блогах, наборы данных и программное обеспечение (<https://profiles.impactstory.org/>).

Все используемые нами инструменты оценивались с позиции открытости и доступности. Например, с помощью инструмента Altmetric можно оценивать публикации как свои, так и других ученых. Для оценки своих исследований или публикаций органи-

зация должна получить лицензию для размещения значка Altmetric на веб-сайте или блоге. Но сегодня это невозможно, так как при обращении в Altmetric приходит отказ на предоставление лицензий российским организациям. Однако этот инструмент был включен в список отобранных, так как им можно пользоваться при просмотре других публикаций, оценить отдельные статьи на тех платформах, которые работают с Altmetric, например, Dimensions.

Окончательный список открытых инструментов для поддержки научных исследований содержит 137 позиций. Полный список выявленных и отобранных инструментов доступен в репозитории Zenodo [33].

ЗАКЛЮЧЕНИЕ

Выполненный нами анализ позволил выявить группу открытых инструментов для сопровождения рабочих исследовательских процессов, что позволит построить на платформе «Библиотека для открытой науки» в ГПНТБ СО РАН систему открытых инструментов для поддержки научных исследований.

Использование различных инструментов с соблюдением практик открытой науки показывает, что это позволит ученым повысить эффективность своей научной работы, сократить время на поиск необходимых публикаций и данных, сделать исследовательские рабочие процессы более открытыми и достоверными.

СПИСОК ЛИТЕРАТУРЫ

1. Mckiernan E.C., Bourne P.E., Brown C.T. et al. How open science helps researchers succeed // eLife. – 2016. – № 5. DOI: <https://doi.org/10.7554/eLife.16800>.
2. Wang H., Gainey M., Campbell P. et al. Implementation and assessment of an end-to-end Open Science & Data Collaborations program [version 2; peer review: 2 approved] // F1000Research. – 2022. – № 11. – Art. 501. DOI: <https://doi.org/10.12688/f1000research.110355.2>.
3. Ponte D., Mierzejewska B.I., Klein S. The transformation of the academic publishing market: multiple perspectives on innovation // Electron Markets. – 2017. – № 27. – P. 97–100. DOI: <https://doi.org/10.1007/s12525-017-0250-9>.
4. The Open Science Training Handbook. – URL: https://open-science-training-handbook.github.io/Open-Science-Training-Handbook_EN/ (дата обращения 02.02.2023).
5. Heller L., The R., Bartling S. Dynamic Publication Formats and Collaborative Authoring // Bartling S., Friesike S. (eds) Opening Science. – Cham: Springer, 2014. DOI: https://doi.org/10.1007/978-3-319-00026-8_13.
6. Rahal R.-M., Havemann J. Science in Crisis. Is Open Science the Solution? // MetaArXiv. – 2019. DOI: [10.31222/osf.io/3hb6g](https://doi.org/10.31222/osf.io/3hb6g).
7. Heck T. et al. Open Science Practices in Higher Education: Discussion of Survey Results from Research and Teaching Staff in Germany // Education for Information. – 2020. – Vol. 36, № 3. – P. 301-323. DOI: [10.3233/EFI-190272](https://doi.org/10.3233/EFI-190272).
8. Malički M., Jerončić A., ter Riet G. et al. Preprint Servers' Policies, Submission Requirements,

- and Transparency in Reporting and Research Integrity Recommendations // JAMA. – 2020. – Vol. 324, № 18. – P. 1901–1903. DOI:10.1001/jama.2020.17195.
9. Potterbusch M., Lotrecchiano G. Shifting paradigms in information flow: An open science framework (OSF) for knowledge sharing teams // Informing Science: the International Journal of an Emerging Transdiscipline. – 2018. – № 21. – P. 179–199. DOI: <https://doi.org/10.28945/4031>.
 10. ArXives of Earth science // Nature Geoscience. – 2018. – Vol. 11, № 149. DOI: <https://doi.org/10.1038/s41561-018-0083-y>.
 11. Ide K., Nakayama J. Researchers support preprints and open access publishing, but with reservations: A questionnaire survey of MBSJ members // Genes to Cells. – 2023. – Vol. 28, № 5. – P. 333–337. DOI: <https://doi.org/10.1111/gtc.13015>.
 12. Bezuidenhout L., Havemann J. The varying openness of digital open science tools [version 2; peer review: 1 approved, 1 approved with reservations] // F1000Research. – 2021. – № 9. – Art. 1292. DOI: <https://doi.org/10.12688/f1000research.26615.2>.
 13. Hicks D. J. Open science, the replication crisis, and environmental public health // Accountability in Research. – 2023. – Vol. 30, № 1. – P. 34–62. DOI: 10.1080/08989621.2021.1962713.
 14. Chen X., Jagerhorn M. Implementing FAIR Workflows along the research lifecycle // Procedia Computer Science. – 2022. – № 211. – P. 83–92. DOI: <https://doi.org/10.1016/j.procs.2022.10.179>.
 15. Wilkinson M., Dumontier M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship // Sci Data. – 2016. – № 3. – Art. 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>.
 16. Baždarić K., Vrkić I., Arh E., Mavrinac M., Gligora Marković M., Bilić-Zulle L. et al. Attitudes and practices of open data, preprinting, and peer-review—A cross sectional study on Croatian scientists // PLoS ONE. – 2021. – Vol. 16, № 6. – Art. e0244529. DOI: <https://doi.org/10.1371/journal.pone.0244529>.
 17. Müller H., Lopes-Dias C., Holub P., Plass M., Jungwirth E., Reihls R., Torke P.R., Malatras A., Berger A., Coombs H., Dillner J., Merino-Martinez R. BIBBOX, a FAIR toolbox and App Store for Life Science Research // New biotechnology. – 2023. – Vol. 77. – P. 12–19. DOI: 10.1016/j.nbt.2023.06.001.
 18. Schonfeld R.C. What is Researcher Workflow? // Ithaka S+R. – URL: <https://sr.ithaka.org/blog/what-is-researcher-workflow/> (дата доступа 27.04.2023).
 19. FAIR Aware: Your first step towards your FAIR data (set). – URL: <https://fairaware.dans.knaw.nl/> (дата обращения 19.06.2023).
 20. Редькина Н. С. Библиотека в условиях информационной экосистемы открытой науки // Научно-техническая информация. Сер. 1. – 2021. – № 10. – С. 9–18.
 21. Шевченко Л. Б. Комплексная поддержка открытой науки: обзор практик // Научно-техническая информация. Сер. 1. – 2022. – № 12. – С. 28–32. DOI: 10.36535/0548-0019-2022-12-5.
 22. Шевченко Л. Б. Информационная поддержка научных исследований в российских вузах в условиях открытой науки // Информационные ресурсы России. – 2022. – № 2 (186). – С. 107–116. DOI: 10.52815/0204-3653_2022_02186_107.
 23. Kramer B., Bosman J. Innovations in scholarly communication – global survey on research tool usage // F1000Res. – 2016. – № 5. – Art. 692. DOI: 10.12688/f1000research.8414.1.
 24. Шевченко Л. Б. Открытая наука: учёные – «за», библиотеки? // Научные и технические библиотеки. – 2023. – № 2. – С. 113–131. DOI: <https://doi.org/10.33186/1027-3689-2023-2-113-131>.
 25. Шевченко Л. Б. Отношение исследователей и библиотекарей к открытой науке // Наука, технологии и информация в библиотеках (LIBWAY-2023): Международная научно-практическая конференция, посвященная 105-летию ГПНТБ СО РАН (Новосибирск, 20–22 марта 2023 г.). – Новосибирск, 2023 – URL: <https://www.youtube.com/watch?v=VIVXWwoSpR8&t=7135s>
 26. Рекомендация по открытой науке Юнеско. – URL: [https://unesdoc.unesco.org/ark:/48223/pf0000379949_rus_\(дата обращения 02.02.2023\)](https://unesdoc.unesco.org/ark:/48223/pf0000379949_rus_(дата обращения 02.02.2023)).
 27. Felden J., Möller L., Schindler U. et al. PANGAEA – Data Publisher for Earth & Environmental Science // Sci Data. – 2023. – № 10. – Art. 347. DOI: <https://doi.org/10.1038/s41597-023-02269-x>.
 28. Martin E.L., Barrote V.R., Cawood P.A. A resource for automated search and collation of geochemical datasets from journal supplements // Sci Data. – 2022. – № 9. – Art. 724. DOI: <https://doi.org/10.1038/s41597-022-01730-7>.
 29. Kramer B., Bosman J. Open Science – What’s in it for me? Practices and tools for your workflow : workshop for OIS Center der Ludwig-Boltzmann Gesellschaft // Open Access Network Austria, Vienna, Austria, 20 September 2017. – URL: <http://tinyurl.com/vienna-openscience> (дата обращения 05.04.2023).
 30. Open Access Scholarly Publishing Association. – URL: <https://oaspa.org/> (дата обращения 12.06.2023).
 31. Scolary. – URL: <https://scolary.com/about> (дата обращения 11.05.2023).
 32. OPERAS Tools Research and Development White Paper, July 2021 // OPERAS. – URL: <https://operas-eu.org/special-interest-group-living-book/operas-tools-research-and-development-white-paper-june-2021/#Tools-Definitions-and-criteria> (дата обращения 12.04.2023).
 33. Shevchenko L. Open tools [Data set] // Zenodo. – 2023. – URL: <https://doi.org/10.5281/zenodo.8053708>

Материал поступил в редакцию 20.06.23

Сведения об авторе

ШЕВЧЕНКО Людмила Борисовна – кандидат педагогических наук, старший научный сотрудник Отдела научных исследований открытой науки Государственной публичной научно-технической библиотеки Сибирского отделения Российской академии наук, г. Новосибирск
e-mail: shevchenkol@spsl.nsc.ru

ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК [002:658]:004.451.5

Н.А. Мазов, В.Н. Гуреев

Ведение базы данных публикаций организации с использованием библиографических ресурсов открытого доступа*

Обращение к внешним библиографическим системам – неизбежный этап в работе научных библиотек и информационных служб при создании собственных ресурсов. С одной стороны, информация из внешних источников может широко использоваться при ведении институциональной базы данных / репозитория – на этапах поиска информации о публикациях организации, настройки оповещений о появлении новых записей и их экспорта в нужных форматах для сокращения времени ввода. С другой стороны, наиболее полные данные из внутренних систем могут использоваться для корректировки сведений во внешних базах данных, повышая точность публикационных профилей организаций и увеличивая представленность библиографической информации. С развитием открытой науки ожидаемым становится использование бесплатных библиографических систем, что особенно актуально с учетом дороговизны и иных трудностей с доступом к коммерческим продуктам. На массиве публикаций одного из институтов РАН показаны возможности работы с открытыми библиографическими ресурсами по наполнению институциональной базы данных. Проведено сравнение прежде задействованных коммерческих систем Web of Science и Scopus с открытыми ресурсами РИНЦ, Dimensions и Lens с позиций охвата этими базами данных научного контента организации, а также их применимости к библиотечным технологическим процессам.

Ключевые слова: институциональная база данных, репозиторий, публикационная активность, научная коммуникация, РИНЦ, Dimensions, Lens, Scopus, Web of Science

DOI: 10.36535/0548-0019-2023-09-4

ВВЕДЕНИЕ

Институциональные базы данных и репозитории научных публикаций являются эффективным инструментом научной коммуникации, повышают видимость ведущихся в учреждении исследований, позволяют проводить библиометрический анализ разного уровня сложности, предоставляют единый доступ к интеллектуальной продукции организации и обеспечивают сохранность массива накопленных знаний. Обзоры российского опыта по ведению институциональных баз данных и репозиториях, к сожалению, показывают их ограниченное распространение в научно-образовательных организациях, а также недостаточные функциональные возможности многих из них, в том числе

трудности с навигацией и поиском информации [1, 2]. Тем не менее, все больше учреждений демонстрируют эффективно организованную работу по ведению баз данных, включая БЕН РАН [3], Институт катализа СО РАН [4], Институт вычислительного моделирования СО РАН [5], Межведомственный суперкомпьютерный центр РАН [6], ГПНТБ СО РАН [7, 8] и некоторые другие.

Большинство подходов к эффективному построению и наполнению институциональных баз данных и репозиториях предполагает активное обращение к внешним библиографическим системам, позволяющее решать сразу несколько задач:

1) библиотеки в автоматическом режиме и на периодической основе получают актуальные сведения о новых публикациях организации (без обращения к авторам или другим службам организации);

* Исследование выполнено по проектам ГПНТБ СО РАН (122040600059-7) и ИНГТ СО РАН (FWZZ-2022-0028).

2) появляется возможность оперативно импортировать записи во внутреннюю систему организации, устраняя потребность в ручном вводе данных и экономя ценное время библиографов;

3) библиотекари могут вносить необходимые корректировки в сами внешние системы и, таким образом, поддерживать публикационные профили организации и ее авторов в актуальном состоянии.

При ведении внутренних баз данных многие организации как в России [2, 5, 9], так и за рубежом [10–13] на определенных этапах работы чаще всего обращаются к международным библиографическим системам Web of Science компании Clarivate и Scopus издательства Elsevier, что продиктовано их наибольшей авторитетностью и богатым опытом в обработке метаданных. В российских организациях до недавнего времени дополнительно решались задачи подготовки библиометрических показателей по этим системам с учетом требований Министерства высшего образования и науки (обеспечивавшего доступ до 2022 г.), Высшей аттестационной комиссии, грантовых фондов. Необходимые отчетные библиометрические показатели из этих систем в том или ином виде встраивались во внутренние базы данных, определяя их структуру и наполнение. При активном содействии российских представительств Clarivate и Elsevier успешно решались и задачи по массовой корректировке публикационных профилей российских научных и образовательных учреждений в базах Web of Science и Scopus.

Несмотря на широкую распространенность этих двух систем, в последнее десятилетие с развитием открытой науки у них появились серьезные конкуренты. Так, внушительные массивы метаданных научных публикаций предоставляют базы данных Lens (открыта в 2013 г.), Semantic Scholar (2015 г.), Dimensions (2018 г.), OpenAlex (2022 г.) и некоторые другие. Во многих из них доступен не только стандартный набор метаданных, но и актуальная в последние годы информация об аффилиациях авторов, финансировании исследований, различных идентификаторах (публикаций, организаций, авторов, источников и пр.), списках пристатейной литературы и прочих востребованных атрибутах научных публикаций, что приближает их по качеству информации к коммерческим базам данных. Получили развитие знаковые проекты I4OA – Initiative for Open Abstracts (<https://i4oa.org>), I4OC – Initiative for Open Citations (<https://i4oc.org>) и OpenCitations (<https://opencitations.net>) [14], позволяющие системам в рамках открытых лицензий беспрепятственно заимствовать друг у друга метаописания публикаций. Кроме преимуществ бесплатного доступа и сопоставимых функциональных возможностей, открытые системы нередко демонстрируют и существенно больший охват научных источников в сравнении с Web of Science и Scopus [15].

Таким образом, тенденция развития открытого доступа к метаданным публикаций в совокупности с разного рода проблемами с доступом к коммерческим системам, включая тривиальную дороговизну, актуализирует использование бесплатных информационных ресурсов, в том числе при ведении институциональных баз данных и репозиториев.

В настоящей статье на примере институциональной базы данных публикаций Института нефтегазовой геологии и геофизики (ИНГГ) СО РАН описан опыт технологического перехода с внешних коммерческих на открытые библиографические ресурсы, используемые в технологических процессах сбора и наполнения внутренней системы. Рассмотрены также задействованные в работе базы данных открытого доступа RINCC, Dimensions и Lens, проведено их сравнение с ранее используемыми системами Web of Science и Scopus с позиций охвата публикаций ИНГГ СО РАН и функциональной совместимости с технологическими процессами ведения институциональной базы данных. Описанный опыт может быть полезен научным библиотекам, ответственным за ведение баз данных публикаций сотрудников обслуживаемых научных организаций.

ВЫБОР ОТКРЫТЫХ БИБЛИОГРАФИЧЕСКИХ СИСТЕМ ДЛЯ ВЕДЕНИЯ ИНСТИТУЦИОНАЛЬНОЙ БАЗЫ ДАННЫХ

Обоснованный выбор библиографической системы, обращение к которой будет встроено в технологическую цепочку ведения институциональной базы данных, может существенно предопределить качество информации во внутренней системе. Основными требованиями к внешним системам, апробированными нашим собственным многолетним опытом и опытом наших коллег (см. обзорную часть статьи [2]), стали следующие:

1) реализация расширенного поиска, создание глобальных и сводных запросов, в том числе по авторским идентификаторам (ID). Эта функция необходима как для настройки оповещений по широкому запросу, так и для решения прикладных библиометрических задач;

2) наличие авторских профилей, желательно с присвоением внутренних идентификаторов. Как правило, запрос по названию организации не дает полной информации из-за большого числа разночтений, многих способов перевода и транслитерации названия на английский язык и, кроме того, не охватывает публикации сотрудников без указания аффилиации. Запрос по авторам дает намного более полные результаты, в том числе по неаффилированным с организацией публикациям, которые часто нужны для заполнения заявок на гранты, при аттестации сотрудников, оценке членов диссертационных советов;

3) создание системы оповещений, которая позволяет в автоматическом режиме и на периодической основе отслеживать появление новых публикаций сотрудников организации и, при необходимости, получать сведения о цитировании ранее опубликованных работ;

4) поддержка полноты метаданных: с учетом неуклонно расширяющегося в последние годы набора метаданных научных публикаций, а также необходимости их массовой обработки, полнота метаданных может существенно сократить время подготовки библиографических записей. Набор метаданных год от года меняется, преимущественно в сторону увеличения, и уже сейчас к традиционным выходным данным публикации добавились такие обязательные по-

ля, как аффилиация, число участвовавших в исследовании организаций, наличие зарубежных соавторов, сведения о финансировании исследования, коды тематических рубрик и различные идентификаторы публикации (DOI, ISBN, ISSN, EDN и др.);

5) обеспечение полноты выгрузки метаданных, а также возможности выбора и настройки форматов их экспорта. Зачастую набор доступных для выгрузки полей не совпадает с тем, что отображается в веб-интерфейсе системы, поэтому данный параметр мы оценивали отдельно. Эта функция позволяет существенно экономить время, затрачиваемое библиотекарями сотрудниками на подготовку библиографических записей в институциональной базе данных.

На основе недавнего обзорного исследования, посвященного открытым библиографическим ресурсам [15], из 12 библиографических систем для дальнейшего тестирования нами были выбраны Dimensions и Lens. Хотя в обеих базах данных предлагаются дополнительные платные (и весьма полезные) функции, в нашей работе рассматривались общедоступные бесплатные модули. В сравнение была включена система РИНЦ как единственная российская национальная библиографическая база данных, хотя в ней до сих пор не реализована система оповещений, а использование авторских профилей для формирования списка публикаций организации предполагает наличие подписки на Science Index. Добавлены также традиционно применяемые до конца 2022 г. Web of Science и Scopus. Далее мы представим краткие описания каждого из информационных ресурсов с позиций их участия в работе по ведению институциональной базы данных.

Российский индекс научного цитирования, РИНЦ (<https://elibrary.ru>) – база данных Научной электронной библиотеки, образованная в 2005 г., – является основным региональным источником библиографических метаданных о российских и русскоязычных публикациях, а также публикациях некоторых стран СНГ. Задача ведения институциональной базы данных с использованием РИНЦ сопряжена с рядом проблем, до сих пор ожидающих своего решения. Во-первых, для пользования расширенными функциональными возможностями, включая выгрузку данных и редактирование профилей организаций и авторов, необходима подписка на модуль Science Index, что не позволяет отнести РИНЦ к безусловно бесплатным системам. Во-вторых, даже в платном варианте до сих пор не реализована установка оповещений о новых публикациях, а также настройка формата выгрузки данных. В-третьих, формирование сводных и гибких запросов и хранение запросной истории нуждаются в значительном совершенствовании, чтобы соответствовать современным информационным системам. Поэтому работа с РИНЦ, к сожалению, требует затратного по времени регулярного визуального мониторинга публикаций, а также существенной доработки метаданных. Открытым остается и вопрос об индексации в РИНЦ публикаций российских авторов в зарубежных системах, которые прежде импортировались из ныне недоступной в России базы данных Scopus.

Dimensions (<https://www.dimensions.ai>) – база данных британской компании Digital Science, ведущая историю с 2018 г., – уникальна нацеленностью на отслеживание полного цикла научной работы от ее замысла на этапе заявок на финансирование до реализации в виде конечных продуктов – патентов, клинических исследований или постановлений [16, 17]. Согласно литературным данным, Dimensions считается наиболее близким по своим возможностям аналогом Web of Science и Scopus [18–20] и активно используется для ведения репозитория [10]. Являясь системой коммерческого распространения, Dimensions предоставляет бесплатно значительную часть своих метаданных, которых вполне хватает для работы как обычным исследователям, так и научным библиотекам.

Lens (<https://www.lens.org>) – австралийская библиографическая база данных компании Cambia, запущенная в 2013 г., – характеризуется предоставлением доступа к большим объемам патентной информации, однако в равной мере она охватывает и значительные пласты научных публикаций, предоставляя широкие возможности их анализа. Также особенностью Lens является разработка мета-идентификатора записей, объединяющего в себе все прочие идентификаторы научного объекта – DOI, ORCID, идентификаторы публикаций в других системах и пр. [21, 22]. Многочисленные функциональные возможности Lens наиболее приближены к таковым в Web of Science и Scopus.

Scopus (<https://www.scopus.com>) – база данных издательства Elsevier – была образована в 2004 г. Основываясь на экспертном подходе в отборе контента для индексирования, разработчики Scopus основные усилия приложили к равномерному и широкому охвату источников, так что в системе в большем объеме, чем в Web of Science, представлены метаданные публикаций на различных национальных языках, индексируется большее число журналов и материалов конференций. Кроме того, в Scopus представлен собственный набор библиометрических показателей. В плане удобства работы с метаданными система Scopus не уступает Web of Science, а ее систему автоматического формирования авторских профилей можно считать одной из лучших [23].

Web of Science (<https://www.webofscience.com>) – американская база данных компании Clarivate – это наиболее авторитетная на сегодняшний день библиографическая система, основной характеристикой которой является строгий экспертный отбор научного контента для индексирования. Образованная в 1964 г. Юджином Гарфилдом, эта система до сих пор задает библиометрические стандарты и представляется наиболее используемой в оценке научных результатов, рейтинговании журналов и университетов. Между тем развитие многих современных сервисов в Web of Science, равно как и расширение спектра индексируемых источников во многом возникли благодаря жесткой конкуренции на рынке библиографических баз данных. По всем параметрам, необходимым научным библиотекам, равно как и по удобству работы с профилями организаций и авторов, а также их корректировке в самой системе, Web of Science представляет собой образец наиболее высокого качества.

К недостаткам двух последних систем следует отнести коммерческий характер их распространения, невозможность подписаться на работу в системе обычным пользователям, а также остающиеся пробелы в индексировании научного контента. Так, обе системы критикуются за смещенность в сторону англоязычной литературы [24] и естественнонаучной тематики [25, 26], за слабый охват типов документов кроме журнальных статей, а в последние годы – и за недостаточный охват документов в целом. Несмотря на платный характер распространения, обе базы данных предлагают ограниченный набор функций для бесплатной работы. В частности, авторы имеют возможность просмотра данных о своих публикациях и библиометрических показателях, реализована возможность внесения корректировок в авторские профили. Кроме того, в 2021 г. компанией Clarivate представлена мобильная версия My Research Assistant (MyRA), дающая возможность проводить информационный поиск по публикациям из Web of Science Core Collection (CC) последних пяти лет. Поскольку набор отображаемых полей ограничен, а установка оповещений и выгрузка в бесплатных модулях не реализованы, промышленное применение этих систем в научных библиотеках невозможно. Реализуемый в отдельных библиотеках планомерный просмотр авторских профилей, на наш взгляд, неоправданно трудозатратен и к тому же не дает гарантированной полноты информации. В частности, в Web of Science поиск проводится лишь по авторам, которые провели верификацию своих записей (а это крайне ограниченное число исследователей), в Scopus же введены строгие лимиты на объем

выдаваемой информации как по авторам, так и по их публикациям.

В табл. 1 представлены основные характеристики рассматриваемых систем с точки зрения их применимости к ведению институциональной базы данных и к возможностям вносить корректировки в сами внешние системы.

Из данных табл. 1 видно, что наибольшие функциональные возможности предоставляются в распространяемых по подписке системах Web of Science и Scopus. РИНЦ в значительной степени может использоваться для мониторинга публикаций российских авторов, хотя широта набора доступных функций во многом определяется тем, подписана организация на платный модуль Science Index или нет. Dimensions и Lens предоставляют в бесплатном доступе практически сопоставимые с Web of Science и Scopus функциональные возможности и потенциально могут стать полноценной заменой коммерческим системам. При этом охват источников в них намного шире, чем в Web of Science и Scopus.

Несмотря на более широкий охват источников в бесплатных системах, а также на сопоставимые функциональные возможности, важным было практически оценить их применимость к ведению институциональной базы данных конкретной организации в сравнении с ранее используемой технологией обращения к коммерческим системам. Для этого все пять ресурсов сравнивались по двум параметрам: отражению авторов и охвату публикаций отдельной организации на примере ИНГТ СО РАН.

Таблица 1

Наполнение и характеристики международных политематических библиографических баз данных, определяющие эффективность их использования в ведении институциональных баз данных / репозиториях, а также поддержке публикационных профилей организаций (данные на апрель 2023 г.)

База данных	Характеристика								
	Глобальный / сводный запрос	Наличие авторских профилей с уникальными ID	Оповещения	Полнота		Возможность создания / правки профиля автора	Число		
				мета-данных	выгрузки		индексируемых журналов, тыс.	проиндексированных научных публикаций, млн	проиндексированных ссылок
РИНЦ	-	+	-	+	+	(при подписке на Science Index)	5,6	41,3	620 млн
Dimensions	+(GET-запрос)	+	+	+	+	+	137	135	1,7 млрд
Lens	+	-	+	+	+	+	51	255	1,9 млрд
Scopus	+	+	+	+	+	+	28	84	1,8 млрд
Web of Science CC	+	+	+	+	+	+	21,5	86	1,9 млрд

ПОСТРОЕНИЕ ЗАПРОСОВ НА ПОИСК НОВЫХ ПУБЛИКАЦИЙ, ОЦЕНКА ОХВАТА АВТОРОВ И ВОЗМОЖНОСТИ КОРРЕКТИРОВКИ АВТОРСКИХ ПРОФИЛЕЙ

Публикации ИНГГ СО РАН организованы в виде базы данных «Труды сотрудников ИНГГ СО РАН» (<http://ibc.ipgg.sbras.ru>) [27], которую сопровождает дополнительный модуль «Авторские идентификационные профили» с указанием сведений обо всех авторских идентификаторах сотрудников Института в различных библиографических системах [28]. База данных содержит полные сведения о публикационной активности научных сотрудников Института со времени его основания – 1957 г. За более чем полувековую историю ИНГГ СО РАН претерпел несколько крупных структурных изменений, включая разделение на две исследовательские организации – Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН и Институт геологии и минералогии им. В.С. Соболева СО РАН. Кроме публикаций головной организации мы учитывали публикации трех иногородних филиалов – Западно-Сибирского, Томского и Ямало-Ненецкого.

Наполнение базы данных Института в последние два десятилетия во многом строилось на автоматизированном сборе сведений о новых публикациях по оповещениям в Web of Science и Scopus [2, 29], а также на визуальном просмотре публикаций в РИНЦ из-за уже упомянутого отсутствия в РИНЦ системы оповещений. Преимущества автоматизированного сбора библиографической информации из внешних источников подтверждаются литературными данными [11] и нашим собственным опытом, указывающим на крайне низкую долю сведений, поступающих от самих сотрудников.

Запросные строки формировались из авторских идентификаторов в Scopus и фамилий авторов с учетом разных вариантов их транслитерации в Web of Science. К запросам добавлялись идентификаторы самой организации, а также адресные запросы со множественными вариантами написания названий организации [30]. Следует отметить, что в 2022 г. в Web of Science были введены уникальные внутренние авторские идентификаторы, по которым теоретически стало возможным строить запрос аналогично запросу в Scopus, т. е. не по фамилиям, а по уникальным номерам. Однако в процессе тестирования выяснилось, что идентификаторы авторов в Web of Science носят динамический характер и, кроме того, при объединении дублирующихся авторских профилей освободившиеся идентификаторы могут назначаться другим авторам. Автообновление профилей в Web of Science долгое время отсутствовало, в отличие от Scopus. Оно было анонсировано лишь в апреле 2023 г. [31]. Все это делает авторские идентификаторы Web of Science непригодными для построения запросной строки. В то же время в 2022 г. новый импульс в развитии получил другой авторский идентификатор в Web of Science – ResearcherID (RID). Он был в автоматическом режиме присвоен всем авторам, имеющим хотя бы одну публикацию в этой системе (за ис-

ключением тех, у кого уже были номера ResearcherID). Кроме того, из системы были удалены ResearcherID авторов, которые получили их прежде, но не имели публикаций в Web of Science. Таким образом, ResearcherID стало возможным использовать в запросах аналогично идентификаторам Scopus AuthorID.

Подход на основе запросов по авторам позволяет не только обнаруживать публикации организации, которые по техническим причинам оказались не связанными с ее профилем, но и неаффилированные с организацией публикации сотрудников, необходимые для аттестации исследователей, подачи заявок на гранты и пр. Запрос в Web of Science включал верифицированное в системе название организации и множество вариантов его написания для учета возможно не связанных с верифицированным профилем публикаций, а также варианты написания имен и фамилий всех активных сотрудников организации (как уже было отмечено, с 2022 г. при наличии доступа к системе стало возможным использовать в запросной строке и RID авторов). Запрос в Scopus выглядел аналогично с той лишь разницей, что вместо вариантов написания имен / фамилий авторов использовались уникальные идентификаторы каждого из сотрудников в системе.

В РИНЦ работа проводилась с использованием интерфейса платного модуля Science Index, который предполагает привязку авторов к организации, правку их профилей, а также отражение структуры организации с внесением фамилий авторов в соответствующие публикационные профили лабораторий и подразделений.

Запросы в Web of Science и Scopus носили гибкий характер и корректировались по мере изменений штата сотрудников или структуры организации. Далее на основе оповещений на почту сотрудника библиотеки приходили сообщения о появлении новых публикаций, которые можно было выгрузить из Web of Science и Scopus в заданных форматах, а затем преобразовать нужным образом и ввести в институциональную базу данных. Можно было своевременно вносить коррективы в профили самих авторов (например, при появлении публикаций однофамильцев в их профилях, формировании дублей и др.) и организации в целом для поддержания актуальных сведений о ее публикационной активности.

С приостановкой для российских организаций доступа в 2022 г. к полной версии Web of Science прекратилось и поступление оповещений о новых публикациях. В базе данных Scopus, к полной версии которой в России также был приостановлен доступ в 2023 г., оповещения продолжают поступать, но отсутствует возможность корректировки запросов, равно как и выгрузки информации. Для сохранения технологического процесса сбора библиографической информации и ее добавления в институциональную базу данных схема запросов, настройки оповещения, выгрузки данных и их импорта в институциональную базу была перенесена на системы Dimensions и Lens как наиболее близкие и бесплатные аналоги Web of Science и Scopus.

Доля авторов ИНГГ СО РАН, имеющих публикационные профили / публикации в РИНЦ, Dimensions, Lens, Scopus и Web of Science (RID), %

БД Труды сотрудников ИНГГ СО РАН	РИНЦ	Dimensions	Lens	Scopus	Web of Science (RID)
100	80,3	59,2	58,8	46,6	38,1

Dimensions, как видно из табл. 1, присваивает внутренние идентификаторы каждому автору, поэтому для построения запроса была взята схема, аналогичная схеме в Scopus, т. е. были использованы уникальные идентификаторы авторов – сотрудников ИНГГ СО РАН – в Dimensions. Подобные сложные запросы в Dimensions составляются непосредственно в адресной строке в виде GET-запроса. Идентификатор и название организации в бесплатной версии Dimensions недоступны, поэтому запрос был ограничен исключительно фамилиями авторов, чего по нашему предыдущему опыту вполне достаточно для учета подавляющей части публикаций организации.

В Lens, в отличие от Dimensions, отсутствуют внутренние идентификаторы авторов. Вместо этого система заимствует информацию об авторах напрямую из ORCID [32], а также дает возможность поиска по фамилии авторов подобно Web of Science. Кроме того, в отличие от бесплатной версии Dimensions, в Lens возможен поиск по организации – как по международному идентификатору организаций ROR (**R**esearch **O**rganization **R**egistry – реестр научных организаций), так и по вариантам написания. Нами был выбран способ формирования сводного запроса по идентификатору организации и варианту написания фамилий и имен авторов – сотрудников Института. Запрос был реализован в окне расширенного поиска.

Поскольку во всех рассматриваемых системах ядро запросов составляют фамилии авторов с указанием их идентификаторов либо разночтений в написании, для оценки полноты получаемой в каждой из систем информации важным представляется оценить долю авторов, имеющих публикации в той или иной системе. В табл. 2 представлен охват авторов ИНГГ СО РАН в РИНЦ, Dimensions, Lens, Scopus и Web of Science. Как уже было отмечено, в Lens отсутствуют автоматически формируемые авторские профили с назначением уникальных идентификаторов, вместо которых используются идентификаторы ORCID. Однако при регистрации авторов в системе и указании на собственные публикации в Lens система назначает уникальный идентификатор, что можно сопоставить с идентификатором SPIN-код в РИНЦ, который можно получить после заполнения соответствующей анкеты.

Как видно из данных табл. 2, наибольшее количество авторов ИНГГ СО РАН зафиксировано в базе данных РИНЦ, где высокая представленность выверенных авторских профилей связана, прежде всего, с национальным характером системы, но и не в меньшей степени с ежегодно проводимым редактированием авторских профилей сотрудниками библиотеки

Института. За РИНЦ следуют базы данных Dimensions, Lens, Scopus и Web of Science, что в целом хорошо согласуется с долей публикаций Института в этих системах, как будет показано далее. Не охваченное ни одной системой небольшое число авторов представлено преимущественно инженерным составом Института, спорадически появляющимся в качестве соавторов в отдельных публикациях, а также молодыми научными сотрудниками с публикациями по итогам локальных научных конференций, не индексируемые ни одной из внешних библиографических систем.

Возможности корректировки авторских профилей имеются во всех пяти системах. В РИНЦ для авторов, равно как и для сотрудников библиотек предусмотрен широкий набор инструментов для уточнения библиографических сведений о публикациях в рамках подписки Science Index. В Dimensions и авторы, и сотрудники библиотек могут по обратной связи отправлять запросы на корректировку библиографических данных в техподдержку, которая оперативно реагирует и вносит соответствующие изменения. В Lens для авторов предусмотрен функциональный инструментальный по ведению собственного профиля, в то время как возможности сторонних лиц по корректировке авторских профилей ограничены. В Web of Science и Scopus, даже при отсутствии подписки авторам предоставляется соответствующий инструментальный для устранения дублирующихся профилей, добавления индексируемых в этих системах, но не привязанных к профилям публикаций, а также для удаления публикаций однофамильцев из своих профилей. В Scopus за авторов корректировать сведения могут и работники библиотек или лица, администрирующие авторские профили сотрудников организаций. В Web of Science и Scopus работники библиотек могут обращаться с запросами по корректировке профилей сотрудников напрямую в техподдержку баз данных с помощью инструментов обратной связи и по почте.

ТЕСТИРОВАНИЕ ФОРМАТОВ / ПОЛНОТЫ ВЫГРУЗКИ И ОЦЕНКА ОХВАТА КОНТЕНТА

После настройки оповещений по описанным выше запросам по авторам, которые реализованы во всех системах, кроме РИНЦ (где привязка автора к организации, тем не менее, автоматически означает учет его публикаций в показателях), важным было протестировать доступные форматы выгрузки данных. Гибкая настройка экспортируемых полей и наиболее широкий набор форматов для экспорта в настоящее время реализованы в Scopus и Web of Science. Все системы дают возможность выгрузки табличных данных (в форматах CSV или Tab-Delimited). Везде,

кроме РИНЦ, реализована выгрузка в формате RIS для программ управления ссылками. В табл. 3 представлена общая информация по количеству и настройке экспортируемых полей.

Формат выгрузки, а также объем и полнота экспортируемых записей имеют большое значение для последующей обработки, конвертации и импортирования метаданных во внутреннюю систему организации, что существенно сокращает время ввода данных. Не менее важен и набор полей, доступных для выгрузки, который не всегда коррелирует с набором метаданных, отражаемых в библиографическом описании публикации на веб-странице в базе данных. Набор полей, необходимых для институциональной базы данных, во многих случаях индивидуален, а кроме того зависит от периодически меняющихся требований как внутри организации, так и на уровне контролирующих инстанций, формулирующих требования и набор показателей по отчетности. Поэтому мы не сравниваем конкретный набор экспортируемых полей в различных системах (см. при необходимости недавнее подробное исследование [33]), одна-

ко в целом констатируем их достаточное число в Dimensions и Lens для решения большинства текущих задач, стоящих перед российскими научными организациями. Определенным недостатком Dimensions и РИНЦ является отсутствие возможности выбора полей для выгрузки, отчего увеличиваются время и объем выгрузки.

Важным шагом для обоснования выбора внешних библиографических систем открытого доступа в технологическом процессе ведения институциональной базы данных стало выявление доли проиндексированных во внешних системах публикаций Института. Полученные по описанным выше запросам списки публикаций из Dimensions и Lens в автоматическом режиме были соотнесены по идентификатору DOI с имеющимися в институциональной базе данных публикациями. Идентификаторы публикаций в Web of Science, Scopus и РИНЦ планомерно проставлялись во внутренней базе данных по мере их ввода. Для анализа был рассмотрен массив публикаций ИНГГ СО РАН за 10-летний период с 2013 по 2022 г. Данные представлены на рис. 1 и в табл. 4.

Таблица 3

Варианты экспорта данных из РИНЦ, Dimensions, Lens, Scopus и Web of Science

Базы данных	Выбор форматов выгрузки	Выбор полей для выгрузки	Число полей для выгрузки	Максимальное число выгружаемых записей
РИНЦ	–	–	99 полей	неограниченно в рамках подписки на Science Index
Dimensions	+	–	16 полей (бесплатная версия) 53 поля (платная версия)	500 (бесплатная версия) 5000 (платная версия)
Lens	+	+	32 поля	1000
Scopus	+	+	5 категорий, 30 полей	20000
Web of Science	+	+	4 категории, 29 полей	1000

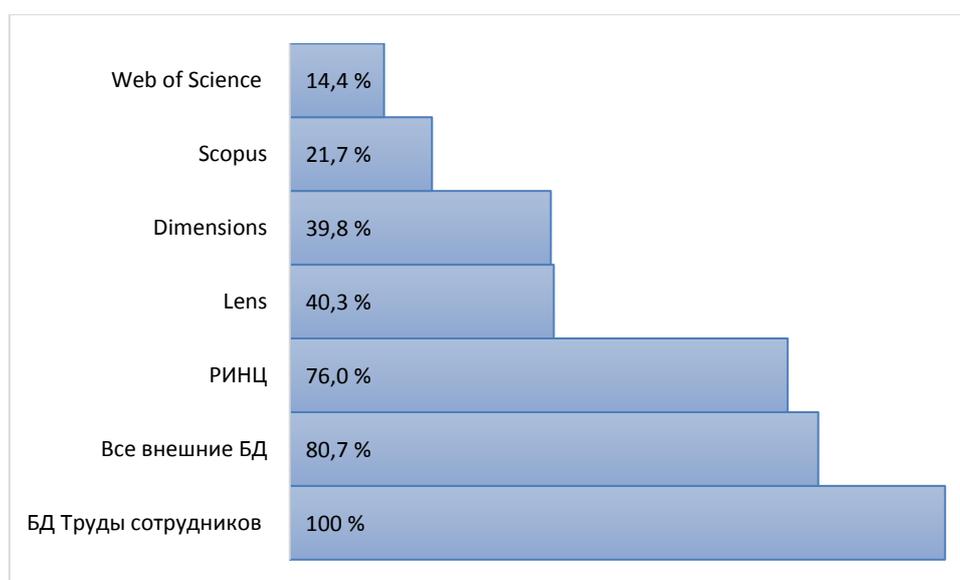


Рис. 1. Долевой охват публикаций ИНГГ СО РАН в РИНЦ, Dimensions, Lens, Scopus и Web of Science за 2013–2022 гг.

Пересечения (в процентах) в индексации публикаций ИНГГ СО РАН в РИНЦ, Dimensions, Lens, Scopus и Web of Science за 2013–2022 гг. Абсолютные значения можно получить напрямую из системы (<http://ibc.ipgg.sbras.ru>) либо по запросу авторам.
Сортировка в порядке убывания объема индексируемых данных

База данных	Труды сотрудников	РИНЦ	Lens	Dimensions	Scopus	Web of Science
Труды сотрудников	100	76,0	40,3	39,8	21,7	14,4
РИНЦ			35,9	35,8	19,0	12,6
Lens				39,7	20,1	13,8
Dimensions					20,1	13,8
Scopus						12,9

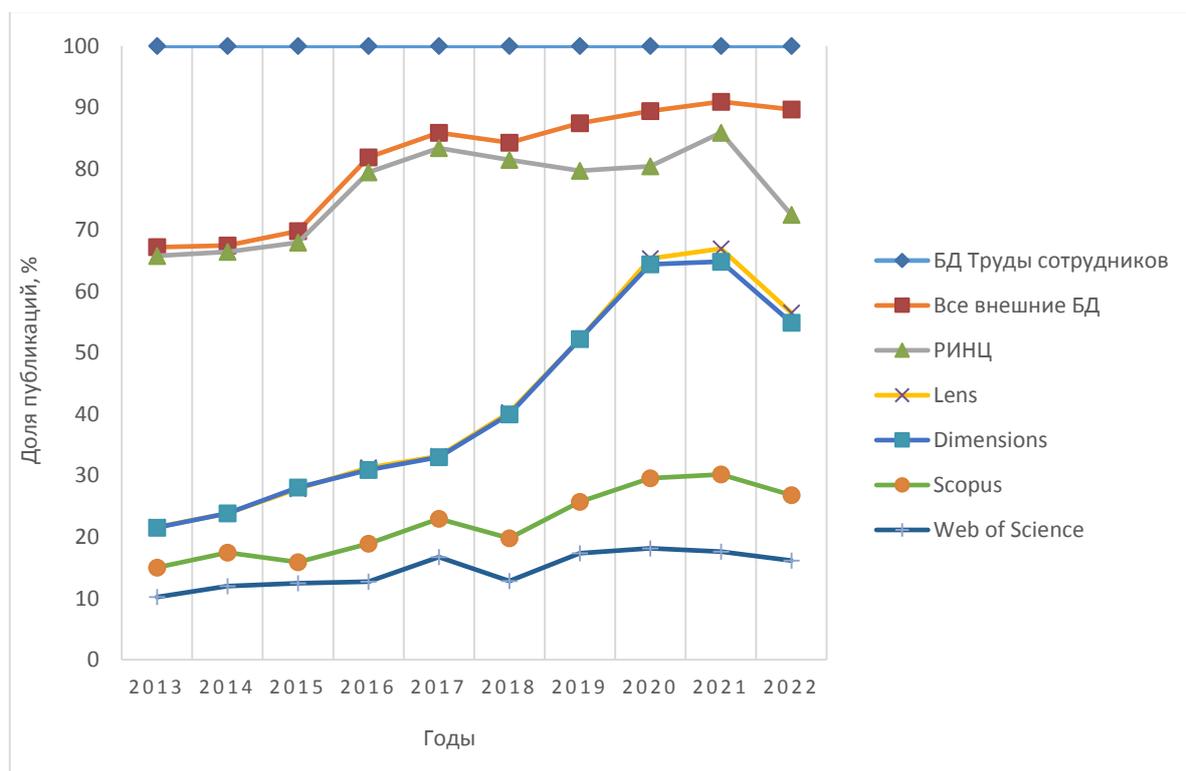


Рис. 2. Индексация публикаций ИНГГ СО РАН в РИНЦ, Dimensions, Lens, Scopus и Web of Science, а также хотя бы в одной из внешних баз данных по годам за 2013–2022 гг., в порядке убывания, %.

Данные рис. 1 и табл. 4 демонстрируют ожидаемо наибольший охват публикаций отечественной системой РИНЦ. Фактически одинаковое число публикаций проиндексировано в бесплатных системах Dimensions и Lens, что объяснимо идентичной основой формирования обеих систем – данными Crossref [16, 21]. За ними следуют Scopus и Web of Science. На все системы приходится более 80 % публикаций, что позволяет говорить как о высокой степени автоматизации в процессе сбора библиографических сведений о публикациях сотрудников Института, так и о достойном качестве публикуемых работ, в большинстве входящих в ту или иную библиографическую систему. На оставшуюся долю нигде не индексируемых работ приходится прежде всего публикации по результатам интеллектуальной деятельности, материалы локальных кон-

ференций, учебные пособия и пр., которые поступают в библиотеку Института от самих авторов.

При оценке пригодности внешних библиографических ресурсов для ведения институциональной базы данных много информации могут дать сведения о равномерности распределения индексации публикаций организации в ретроспективе. На рис. 2 показана индексация публикаций ИНГГ СО РАН в анализируемых системах по годам за рассматриваемый 10-летний период.

Данные рис. 2 позволяют сделать несколько наблюдений:

1) Dimensions и Lens регистрируют фактически равное число публикаций, что, однако, связано с общим источником данных в Crossref. Индексация публикаций Института в этих системах существенно превосходит таковую в Scopus и Web of Science и со временем приближается к уровню индексации в РИНЦ;

2) прослеживается неуклонный рост доли индексируемых публикаций ИНГГ СО РАН практически во всех внешних системах. При этом в РИНЦ, Web of Science и Scopus рост достаточно плавный, что связано с учетом библиотекой Института всех публикаций в этих базах данных, в том числе без идентификатора DOI. Стремительный же рост доли индексируемых публикаций института в Dimensions и Lens в последние годы преимущественно может объясняться распространением в российском публикационном сегменте идентификаторов DOI, поскольку и Dimensions, и Lens (как и ряд других библиографических систем) заимствуют существенную долю своего контента из системы регистрации DOI Crossref. В последние же годы нами отмечена положительная тенденция российских журналов присваивать DOI статьям на ретроспективу, что, безусловно, существенно повышает представленность российских научных достижений в международном научном сообществе;

3) со временем существенно повышается доля публикаций ИНГГ СО РАН, проиндексированных хотя бы в одной из внешних библиографических систем. Так, если в целом за 10 лет не нашли отражения ни в одной из внешних баз данных 19,3 % публикаций Института, то в 2022 г. доля таких публикаций составляла лишь 10,3 %, тогда как в 2013 г. доля таких работ достигала 32,8 %. Здесь могут играть роль как общие тренды на присвоение DOI, делающих видимыми публикации, так и повышение качества проводимых в Институте исследований;

4) РИНЦ индексирует наибольшую долю публикаций, включая практически все статьи, индекси-

руемые в остальных системах. Однако индексация часто проходит с большим запозданием (см. следующий пункт), а отсутствие системы оповещений и удобной выгрузки делает работу с этой системой крайне неудобной;

5) спад индексации публикаций в 2022 г. объясняется преимущественно задержкой в отображении данных во внешних системах, причем этот спад более ярко выражен в системе РИНЦ.

Наши наблюдения позволяют сделать вывод, что российские публикации значительно быстрее проходят индексацию в зарубежных системах, в том числе открытого доступа, чем они попадают в отечественную базу данных. Для отечественных журналов разница в сроках индексации может достигать месяца и более, а для переводных версий российских журналов и публикаций российских авторов в зарубежных журналах разрыв уже составляет от нескольких месяцев до 1,5 лет, что неприемлемо для ведения институциональных баз данных и репозиторий.

Важным в обосновании выбора внешних библиографических ресурсов представляется и распределение индексации по типам документов. Поскольку ИНГГ СО РАН функционирует в сфере естественнонаучных дисциплин, основными типами документов являются статьи и материалы конференций. Монографии, главы в монографиях, учебные пособия, патентные документы, авторские свидетельства, диссертации, препринты и пр. занимают незначительную долю в общем публикационном потоке Института. На рис. 3 показано распределение по трем указанным типам документов.

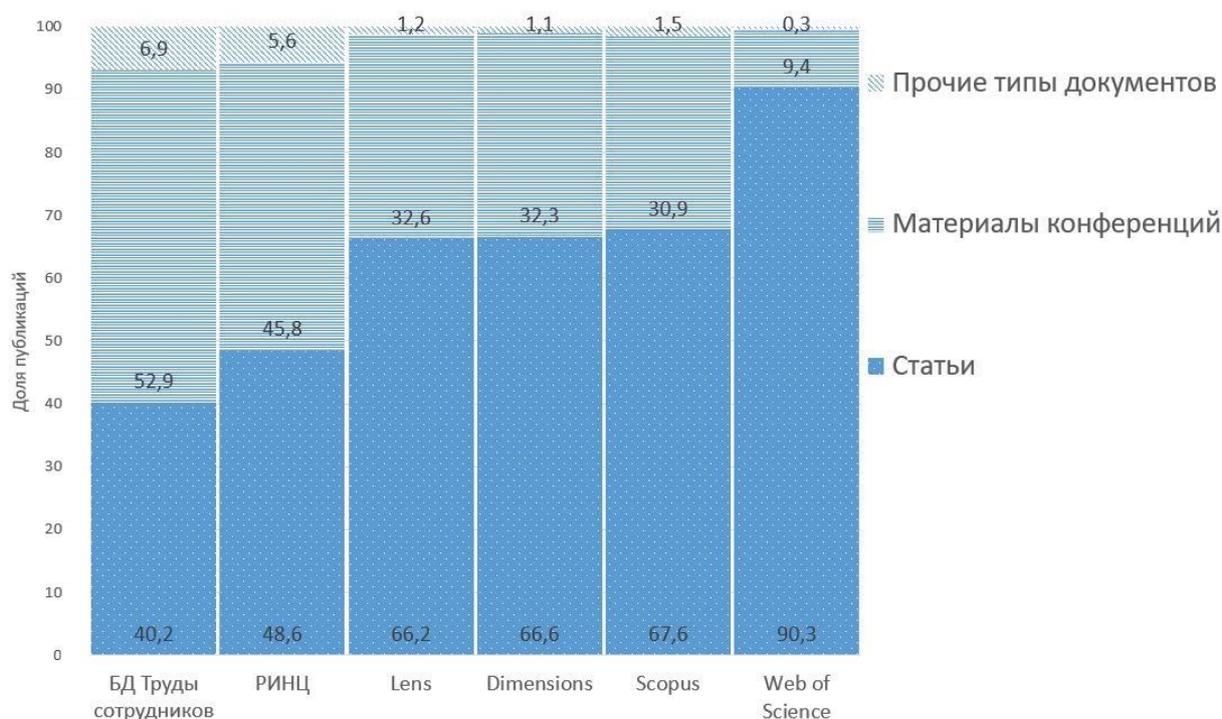


Рис. 3. Распределение индексации публикаций ИНГГ СО РАН в РИНЦ, Dimensions, Lens, Scopus и Web of Science по типам документов за 2013–2022 гг. Показано долевое отношение индексируемых типов документов к общему числу индексируемых документов в каждой из пяти баз данных, %

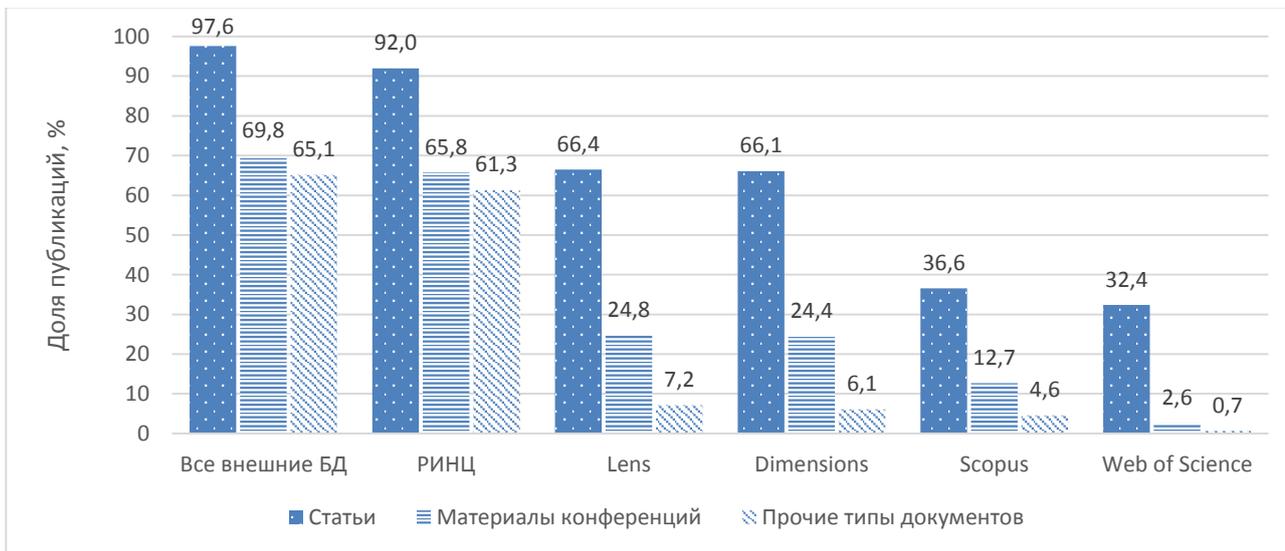


Рис. 4. Распределение индексации публикаций ИНГГ СО РАН в РИНЦ, Dimensions, Lens, Scopus и Web of Science по типам документов за 2013–2022 гг. Показано долевое отношение индексируемых типов документов к общему числу публикаций Института, %

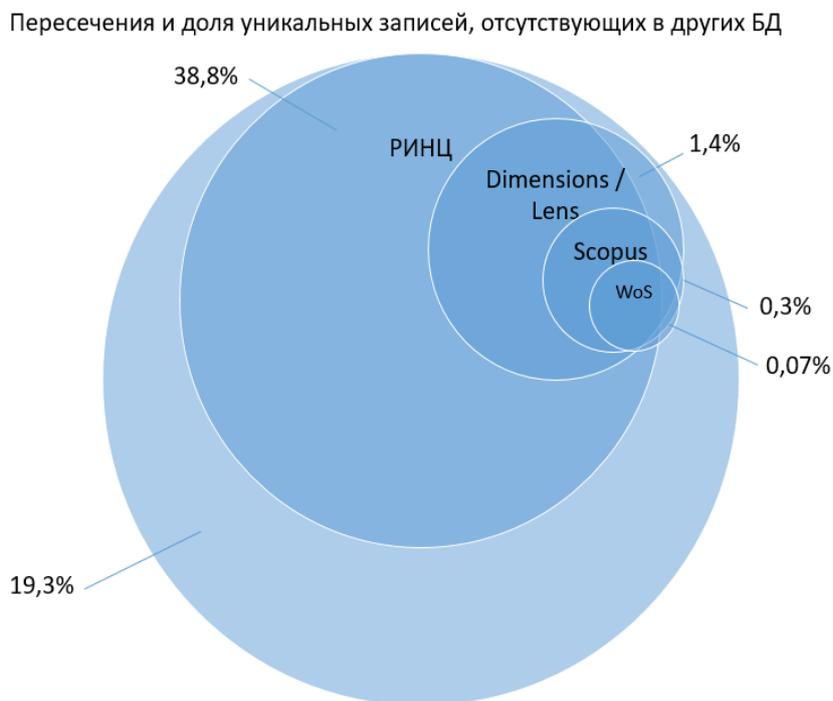


Рис. 5. Степень пересечения контента между РИНЦ, Dimensions, Lens, Scopus и Web of Science на выборке публикаций ИНГГ СО РАН за 2013–2022 гг. С учетом практически идентичной индексации записей в Dimensions и Lens для наглядности они показаны вместе

Как видно из графика, представленного на рис. 3, наиболее сбалансированно материалы конференций и статьи индексируются в РИНЦ. В Dimensions, Lens и Scopus отмечена более высокая доля статей по сравнению с материалами конференций; тем не менее, доля последних представляется достаточно высокой, что может быть связано с проводимой в России работой по продвижению материалов конференций в международные базы данных и присвоением каждой из публикаций в сборниках идентификаторов DOI [34, 35]. Web of

Science из публикаций сотрудников ИНГГ СО РАН индексирует преимущественно статьи, притом что материалы конференций на самом деле составляют большую часть публикуемого контента, что видно в первой колонке рис. 3. Что касается отражения прочих типов публикаций, то они одинаково слабо представлены во всех международных системах.

На рис. 4 представлена доля индексируемых типов публикаций Института во внешних системах по отношению к их общему количеству.

Распределение индексации по типам документов по отношению к общему числу публикаций ИНГГ СО РАН снова демонстрирует, во-первых, наиболее широкий охват публикаций Института в РИНЦ, а во-вторых, наилучшую индексацию научных статей, которая по всем внешним ресурсам достигает почти 100 %. В меньшей мере внешние системы способны поиску сведений о материалах конференций, где не индексируется треть публикаций. Совсем слабо во внешних базах данных представлены прочие, более редкие в естественных науках, типы документов. При этом и по материалам конференций, и по прочим типам документов наиболее высокие значения демонстрирует РИНЦ. В то же время, с учетом уже обозначенных проблем с задержками в индексации контента в РИНЦ, а также с трудностями в своевременном получении сведений об их включении в базу данных и сложностью выгрузки, обращение к Dimensions и Lens может дать хорошие результаты, превосходящие по полноте и не уступающие по качеству результатам работы с коммерческими ресурсами Web of Science и Scopus. Особенно актуальным обращение к Dimensions и Lens становится в последние годы, когда доля индексируемого ими контента приближается к РИНЦ (рис. 2).

На рис. 5 приведены результаты изучения степени пересечения индексируемого контента в различных базах данных.

Как видно из данных рис. 5, пересечения между всеми внешними базами данных достаточно высоки. В РИНЦ наиболее высока доля уникального контен-

та, отсутствующего в других системах. Dimensions и Lens в совокупности индексируют 1,4 % публикаций, которых нет ни в одной из других систем, тогда как на долю Web of Science и Scopus приходится мизерная общая доля таких публикаций – менее одного процента. При этом следует отметить, что на рис. 5 показаны данные за весь рассматриваемый 10-летний период, в который доля нигде не индексируемых статей последовательно сокращалась с 32,8 % в 2013 г. до 10,3 % в 2022 г., как было показано на рис. 2. Следует иметь в виду, что основным типом публикаций, по которым организации ежегодно отчитываются перед различными инстанциями, являются статьи в научных журналах. Поэтому даже при преобладании материалов конференций (см. рис. 3) наиболее важно своевременное отражение в институциональных базах данных именно журнальных статей. На рис. 6 показано распределение индексации только с учетом статей в научных журналах.

Данные рис. 6 наглядно показывают крайне важную роль внешних библиографических систем в агрегации сведений о самом значимом типе публикаций – статьях в научных журналах. При этом базы данных Dimensions и Lens, каждая из которых охватывает 2/3 статей ИНГГ СО РАН (а по 2022 г. – более 80 %), очевидно являются необходимым источником библиографической информации, при том что обе системы практически полностью включают сведения, индексируемые в коммерческих базах данных Web of Science и Scopus, что согласуется с литературными данными [36].

Пересечения и доля журнальных статей, отсутствующих в других БД

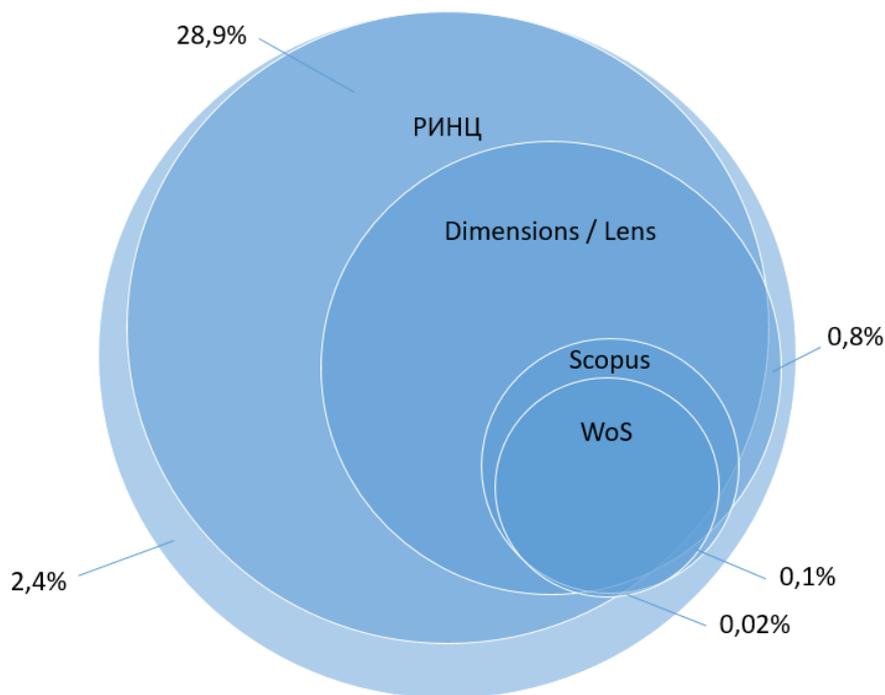


Рис. 6. Степень пересечения контента между РИНЦ, Dimensions, Lens, Scopus и Web of Science на выборке журнальных статей ИНГГ СО РАН за 2013–2022 гг. С учетом практически идентичной индексированности записей в Dimensions и Lens для наглядности они показаны вместе

ЗАКЛЮЧЕНИЕ

Результаты нашего исследования показали, что системы Dimensions и Lens могут рассматриваться как полноценная замена коммерческим базам данных Web of Science и Scopus в качестве источников библиографической информации при ведении институциональных баз данных и репозиториях. Их безусловное преимущество – это широкий охват научного контента, в том числе российских публикаций, оригинальных версий российских журналов и материалов конференций, статей российских авторов в зарубежных журналах. По функциональным возможностям эти две системы хотя и уступают коммерческим системам, тем не менее вполне пригодны для решения задач, стоящих перед научными библиотеками при ведении баз данных публикаций сотрудников научных организаций.

В сравнении с РИНЦ базы данных Dimensions и Lens охватывают меньшее число источников, однако со временем разрыв сокращается, а индексация контента зарубежными системами происходит значительно быстрее. Безусловные же функциональные преимущества зарубежных систем перед РИНЦ делают их необходимым дополнением к отечественной системе при ведении институциональных баз данных и репозиториях.

Выбор библиотекарей и сотрудников информационных служб между Dimensions и Lens, на наш взгляд, может быть обусловлен личными предпочтениями в работе с той или иной системой, поскольку, как показало наше исследование, и контентное наполнение, и функциональные возможности у них сопоставимы.

В качестве ограничений данной работы можно назвать следующие. В Lens нами учитывались только публикации с DOI, притом что в базе данных есть много публикаций ИНГГ СО РАН и без этого идентификатора, поскольку кроме Crossref в Lens много записей из системы OpenAlex, во многом полагающейся на web-индексирование всей сети Интернет. Поскольку обработка таких записей затратна по времени и требует визуального просмотра, эта работа на момент написания статьи еще не была проведена. В Dimensions публикаций Института без DOI обнаружено не было. Таким образом, Lens может иметь преимущество перед Dimensions, что не было показано в настоящей статье. Приведенные нами результаты оценки применимости внешних библиографических ресурсов к ведению институциональных баз данных и (или) репозиториях проводились на примере института естественно-научного профиля, поэтому результаты могут быть полезны преимущественно в организациях естественно-научного цикла.

СПИСОК ЛИТЕРАТУРЫ

1. Стукалова А.А. Функциональные возможности репозитория вузов – участников программы «Приоритет-2030» // Труды ГПНТБ СО РАН. – 2022. – № 2. – С. 36–47.
2. Мазов Н.А., Гуреев В.Н. Базы данных публикаций научной организации как основа информационных исследований // Научно-техническая информация. Сер. 1. – 2022. – № 5. – С. 8–18; Mazov N.A., Gureyev V.N. Publication Databases of Research Organizations as a Tool for Information

- Studies // Scientific and Technical Information Processing. – 2022. – Vol. 49(2). – P. 108–118.
3. Захарова С.С., Гуреева Ю.А. Научные публикации: от картотеки трудов до библиографических профилей // Библиосфера. – 2017. – № 2. – С. 85–89.
4. Альперин Б.Л., Ведягин А.А., Зибарева И.В. SciAct – информационно-аналитическая система Института катализа СО РАН для мониторинга и стимулирования научной деятельности // Труды ГПНТБ СО РАН. – 2015. – Т. 9. – С. 95–102.
5. Ковязина Е.В. Электронный архив научных публикаций: этапы развития // Научные и технические библиотеки. – 2014. – № 2. – С. 19–26.
6. Власова С.А. Автоматизированная система поддержки базы данных научных трудов сотрудников академических учреждений // Информационные ресурсы России. – 2020. – № 5. – С. 29–31.
7. Бусыгина Т.В., Балуткина Н.А., Лаврик О.Л., Мандригина Л.А., Елепов Б.С. Библиографическая база данных трудов сотрудников учреждений СО РАН по нанотехнологиям как инструмент для проведения наукометрических исследований // Информационный бюллетень РБА. – 2013. – Т. 66. – С. 192–200.
8. Баженов С.Р., Балуткина Н.А., Стукалова А.А. Концепция новой информационно-поисковой системы ГПНТБ СО РАН на основе ИРБИС64+ // Научные и технические библиотеки. – 2023. – № 3. – С. 80–101.
9. Баженов С.Р., Данилин М.В., Rogoznikova O.A. Интеграция базы данных публикаций организации с индексами научного цитирования: реализация средствами САБ ИРБИС64 // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: Труды 22-й Международной конференции «Крым-2015» (6–14 июня 2015 г., Судак). – Москва: Изд-во ГПНТБ России, 2015. – С. 1–4.
10. Lappalainen Y., Narayanan N. Harvesting publication data to the institutional repository from Scopus, Web of Science, Dimensions and Unpaywall using a custom R Script // The Journal of Academic Librarianship. – 2023. – Vol. 49(1). – Art. no. 102653.
11. Zhang H., Boock M., Wirth A.A. It Takes More than a Mandate: Factors that Contribute to Increased Rates of Article Deposit to an Institutional Repository // Journal of Librarianship and Scholarly Communication. – 2015. – Vol. 3(1). – Art. no. eP1208.
12. Bull J., Schultz T.A. Harvesting the Academic Landscape: Streamlining the Ingestion of Professional Scholarship Metadata into the Institutional Repository // Journal of Librarianship and Scholarly Communication. – 2018. – Vol. 6(1). – Art. no. eP2201.
13. Li Y. Harvesting and Repurposing Metadata from Web of Science to an Institutional Repository Using Web Services // D-Lib Magazine. – 2016. – Vol. 22(3/4). DOI: 10.1045/march2016-li.
14. Peroni S., Shotton D. OpenCitations, an infrastructure organization for open scholarship // Quantitative Science Studies. – 2020. – Vol. 1(1). – P. 428–444.
15. Гуреев В.Н., Мазов Н.А. Возрастание роли открытых библиографических данных в условиях

- ограничения доступа к коммерческим информационным системам // Управление наукой: теория и практика. – 2023. – Т. 5(2). – С. 49–76.
16. Herzog C., Hook D., Konkiel S. Dimensions: Bringing down barriers between scientometricians and data // *Quantitative Science Studies*. – 2020. – Vol. 1(1). – P. 387–395.
17. Hook D.W., Porter S.J., Herzog C. Dimensions: Building Context for Search and Evaluation // *Frontiers in Research Metrics and Analytics*. – 2018. – Vol. 3. – Art. no. 23.
18. Orduna-Malea E., Delgado-Lopez-Cozar E. Dimensions: re-discovering the ecosystem of scientific information // *Profesional De La Informacion*. – 2018. – Vol. 27(2). – P. 420–431.
19. Thelwall M. Dimensions: A competitor to Scopus and the Web of Science? // *Journal of Informetrics*. – 2018. – Vol. 12(2). – P. 430–435.
20. Visser M., van Eck N.J., Waltman L. Large-scale comparison of bibliographic data sources: Web of Science, Scopus, Dimensions, and Crossref // *Proceedings of the 17th International Conference on Scientometrics & Informetrics (ISSI2019) (2–5 September 2019, Rome, Italy)*. Vol. 2. – Rome: Edizioni Efesto, 2019. – P. 2358–2369.
21. Jefferson O.A., Koellhofer D., Warren B., Jefferson R. The Lens MetaRecord and LensID: An open identifier system for aggregated metadata and versioning of knowledge artefacts. – 2019. – URL: <https://doi.org/10.31229/osf.io/t56yh> (дата обращения 13.06.2023).
22. Penfold R. Using the Lens database for staff publications // *Journal of the Medical Library Association*. – 2020. – Vol. 108(2). – P. 341–344.
23. *Research Metrics Guidebook*. – Elsevier, 2018. – URL: <https://www.elsevier.com/research-intelligence/resource-library/research-metrics-guidebook> (дата обращения 13.06.2023).
24. Vera-Baceta M.A., Thelwall M., Kousha K. Web of Science and Scopus language coverage // *Scientometrics*. – 2019. – Vol. 121(3). – P. 1803–1813.
25. Mongeon P., Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis // *Scientometrics*. – 2016. – Vol. 106(1). – P. 213–228.
26. Martín-Martín A., Orduna-Malea E., Delgado López-Cózar E. Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison // *Scientometrics*. – 2018. – Vol. 116(3). – P. 2175–2188.
27. Мазов Н.А., Гуреев В.Н. IPGGTR Труды сотрудников ИНГГ СО РАН (реферативно-полнотекстовая библиография): Свидетельство о государственной регистрации программы для ЭВМ // Свидетельство о программе 2020621025; RU; № 2020620872, заявл. 10.06.2020, опубл. 19.06.2020, ИНГГ СО РАН.
28. Мазов Н.А., Гуреев В.Н. IPGGAU Авторские идентификационные профили: Свидетельство о государственной регистрации программы для ЭВМ // Свидетельство о программе 2020621128; RU; № 2020620879, заявл. 10.06.2020, опубл. 02.07.2020, ИНГГ СО РАН.
29. Мазов Н.А., Гуреев В.Н. Библиографическая база данных трудов сотрудников организации: цели, функции, сфера использования в наукометрии // *Вестник Дальневосточной государственной научной библиотеки*. – 2016. – № 2. – С. 84–87.
30. Гуреев В.Н., Мазов Н.А. Редактирование профиля организаций в Scopus и РИНЦ: сравнение возможностей // *Научно-техническая информация. Сер. 1*. – 2016. – № 3. – С. 10–22; Gureev V.N., Mazov N.A. Editing organization profiles in SCOPUS and the RSCI: Facilities comparison // *Scientific and Technical Information Processing*. – 2016. – Vol. 43(1). – P. 66–77.
31. *Web of Science Release Notes, April 13 2023: Automatic updates to claimed profiles...* – Clarivate, 2023. – URL: <https://clarivate.com/webofsciencegroup/release-notes/wos/web-of-science-release-notes-april-13-2023-2/> (дата обращения: 13.06.2023).
32. Naak L.L., Fenner M., Paglione L., Pentz E., Ratner H. ORCID: a system to uniquely identify researchers // *Learned Publishing*. – 2012. – Vol. 25(4). – P. 259–264.
33. Лутай А.В., Любушко Е.Э. Сравнение качества метаданных в БД CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection. – РФФИ, 2023. – URL: https://podpiska.rfbr.ru/storage/reports2021/2022_meta_quality.html (дата обращения: 13.06.2023).
34. Стерлигов И.А. Российский конференционный взрыв: масштабы, причины, дальнейшие действия // *Управление наукой: теория и практика*. – 2021. – Т. 3(2). – С. 222–251.
35. Косяков Д.В. Анализ аномального роста количества российских публикаций в сборниках трудов конференций, проиндексированных в БД Scopus // *Научно-техническая информация. Сер. 1*. – 2023. – № 4. – С. 13–24.
36. Singh V.K., Singh P., Karmakar M., Leta J., Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis // *Scientometrics*. – 2021. – Vol. 126(6). – P. 5113–5142.

Материал поступил в редакцию 13.06.23.

Сведения об авторах

МАЗОВ Николай Алексеевич – кандидат технических наук, ведущий научный сотрудник лаборатории информационно-системного анализа Государственной публичной научно-технической библиотеки СО РАН; Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН, г. Новосибирск
e-mail: MazovNA@ipgg.sbras.ru

ГУРЕЕВ Вадим Николаевич – кандидат педагогических наук, ведущий научный сотрудник лаборатории информационно-системного анализа Государственной публичной научно-технической библиотеки СО РАН; Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН, г. Новосибирск
e-mail: GureevVN@ipgg.sbras.ru