



DOI: 10.19181/smtp.2023.5.2.4

EDN: CXJUNG

ВОЗРАСТАНИЕ РОЛИ ОТКРЫТЫХ БИБЛИОГРАФИЧЕСКИХ ДАННЫХ В УСЛОВИЯХ ОГРАНИЧЕНИЯ ДОСТУПА К КОММЕРЧЕСКИМ ИНФОРМАЦИОННЫМ СИСТЕМАМ



**Гуреев
Вадим Николаевич^{1,2}**

¹ Государственная публичная научно-техническая библиотека
Сибирского отделения РАН, Новосибирск, Россия

² Новосибирский государственный технический университет,
Новосибирск, Россия



**Мазов
Николай Алексеевич^{1,3}**

¹ Государственная публичная научно-техническая библиотека
Сибирского отделения РАН, Новосибирск, Россия

³ Институт нефтегазовой геологии и геофизики им. А. А. Трофимука
Сибирского отделения РАН, Новосибирск, Россия

Для цитирования: Гуреев В. Н. Возрастание роли открытых библиографических данных в условиях ограничения доступа к коммерческим информационным системам / В. Н. Гуреев, Н. А. Мазов // Управление наукой: теория и практика. 2023. Т. 5, № 2. С. 49–76. DOI 10.19181/smtp.2023.5.2.4. EDN CXJUNG.

АННОТАЦИЯ

В статье представлен обзор функциональных возможностей и наполнения открытых библиографических баз данных, которые могут быть востребованы в отсутствие доступа к платным коммерческим информационным системам. Базы данных рассматриваются прежде всего с позиций их пользы для исследователей и сотрудников научных библиотек, перед которыми стоят задачи поиска научной и патентной литературы, библиометрической оценки отдельных персоналий и организаций, продвижения публикаций в международное информационное пространство, поиска коллабораторов или решения библиометрических задач. Акцент сделан на политематических ресурсах с широким охватом международных источников.

На основе анализа литературы, посвящённой сравнительному изучению различных библиографических ресурсов, а также исходя из собственного опыта работы с рассмотренными системами авторами сделан вывод о принципиальной возможности решения фактически всех информационно-поисковых и библиометрических задач с помощью современных открытых баз данных и их инструментария. Отмечается, что массивы метаданных сами по себе становятся базовой и не уникальной характеристикой библиографических систем, в то время как на первое место в борьбе за внимание пользователя выходят аналитические функциональные возможности.

КЛЮЧЕВЫЕ СЛОВА:

библиографические базы данных, информационный поиск, открытый доступ, AMiner, Dimensions, Exaly, Google Scholar, Lens, OpenAlex, Scilit, scite, Semantic Scholar, РИНЦ

БЛАГОДАРНОСТИ:

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

ВВЕДЕНИЕ

В последнее десятилетие информационное сопровождение исследований и оценка научной деятельности в России во многом базировались на коммерческих международных системах, прежде всего Web of Science и Scopus. Обе системы активно использовались для проведения информационного поиска, библиометрических исследований, библиотечного комплектования и решения ряда других задач. Выбор этих систем во многом определялся строгостью отбора источников для индексации [1; 2] и политематическим характером формирования баз данных. Заметную роль в распространении и популяризации этих систем сыграли усилия региональных представительств владельцев Web of Science – компании Clarivate и Scopus – издательства Elsevier.

Широкий доступ к обеим системам во всех российских научных и образовательных организациях, с одной стороны, был удобен, поскольку использовались унифицированные подходы и требования к работе с библиографической информацией и библиометрическими показателями. С другой стороны, системы не идеально подходили под нужды оценки именно российской науки, как и многих других национальных систем, что неоднократно освещалось в литературе. Высказывалась критика за смещённость в сторону англоязычной литературы [3] и естественно-научной тематики [4; 5], за слабый охват типов документов кроме журнальных статей, а в последние годы – и за недостаточный охват документов, который порой приводит к неточной оценке научных объектов и соответственно неверным выводам и управленческим решениям [6]. Кроме того, основанная на показателях этих баз данных система оценки научного труда, принятая в Министерстве высшего образования и науки, грантовых фондах или Высшей аттестационной комиссии привела к ситуации, в которой многие другие не менее достойные библиографические ресурсы остались и остаются неизвестными большому кругу пользователей.

Основные задачи, которые прежде решались преимущественно с помощью Web of Science и Scopus и при отсутствии полного доступа к ним требуют перехода на новые информационные ресурсы, схематично представлены в табл. 1.

Таблица 1

Круг задач, решаемых с помощью международных библиографических баз данных, и заинтересованные в их результатах стороны

Контролирующие инстанции	Руководство организаций	Научные журналы	Научные сотрудники	Научные библиотеки
оценка организаций; оценка диссоветов; оценка заявок на научные проекты и их результатов.	приём на работу; карьерное продвижение; премирование, ПРНД; аттестация сотрудников.	поиск рецензентов; определение конкурентов; верификация ссылок при рецензировании;	поиск литературы; поиск коллег для совместных работ; определение конкурентов; поиск журналов для рукописей;	информационные справки; ведение внутренних библиографических баз данных и репозиториев; корректировка сведений о публикациях организации во внешних системах;
		определение рейтинга.	оценка собственной продуктивности.	оптимизация комплектования.
Синий цвет – круг задач, решаемых с помощью библиометрии; зелёный цвет – информационно-поисковые задачи.				

Обратим внимание, что причины отсутствия доступа к коммерческим библиографическим системам могут быть разными, включая недостаточное финансирование, небольшой размер организаций или их ненаучный профиль [7; 8], притом что заинтересованность в информационной поддержке в каждом подобном случае может быть высокой. Поэтому текущая ситуация в России с отсутствием доступа к проприетарным библиографическим продуктам не является уникальной. Также следует отметить, что в России прекращение доступа оказывает наибольший эффект на сферу управления наукой, в которой при многолетнем использовании устоявшегося набора показателей при дальнейшей невозможности их получения возникает необходимость реформирования всей системы оценки [9].

Что касается большинства прочих задач, приведённых в табл. 1, которые являются намного более насущными для развития науки и технологий, то в текущей парадигме открытого доступа к информации они вполне могут решаться с помощью ряда других инструментов. Например, базовые возможности тематического поиска, систему авторских профилей и основные библиометрические показатели авторов предоставляют практически все библиографические базы данных открытого доступа – Google Scholar, Semantic Scholar, AMiner. При этом некоторые из них, например, Lens, позволяют проводить полноценный поиск по пристатейным спискам литературы, что

делает их значимыми конкурентами коммерческих баз данных в качестве источников и инструментов проведения библиометрического анализа.

Оставляя за рамками работы проблемы оценки отечественной науки, которые являются предметом отдельного исследования, в настоящей статье авторы представляют обзор возможностей открытых библиографических систем (а также открытых модулей платных систем) для решения поисковых и ряда библиометрических задач, стоящих перед исследователями и научными библиотеками.

Рассмотрены возможности бесплатных модулей Web of Science, Scopus и Dimensions, а также открытые системы Google Scholar, AMiner, Lens, Scilit, Semantic Scholar и Exaly. Описаны функциональные особенности и платной базы данных scite, которая, однако, доступна по подписке отдельным пользователям (как и организациям в целом) и распространяется за умеренную плату. Приведены характеристики новой системы OpenAlex, которая пока не представила веб-интерфейс и поэтому в основном интересна узкому кругу специалистов по наукометрии. Информация бралась преимущественно из открытых материалов с описаниями баз данных, в том числе из научных статей за авторством сотрудников соответствующих систем; некоторые моменты были уточнены в личной переписке авторов с разработчиками баз данных. Каждая из систем была протестирована на личных публикационных профилях авторов статьи, а также на 10-летнем массиве базы данных «Труды сотрудников ИНГГ СО РАН» [10]. Результаты этого практического исследования приняты к опубликованию в журнале «Научно-техническая информация. Серия 1: Организация и методика информационной работы» и будут доступны в одном из выпусков в 2023 г.

ОТКРЫТЫЕ БИБЛИОГРАФИЧЕСКИЕ ДАННЫЕ

Инициатива открытого доступа к научным знаниям, тесно связанная с развитием информационных технологий, изначально была направлена на обеспечение свободного доступа к полным текстам научных публикаций. В последние годы инициатива охватывает всё более широкий список объектов. Так, многие журналы уже в обязательном порядке требуют от авторов выставления в общий доступ исходных данных, которые всё чаще используются отдельно и независимо от публикаций. Кроме повышения прозрачности и воспроизводимости результатов, дополнительными выгодами становятся снижение рисков в утрате этих данных, а также их повторное использование другими исследователями, что в целом повышает эффективность научных исследований.

Движение за открытый доступ не обошло стороной метаданные и библиографические системы [7]. Для обеспечения авторитетности и широкого охвата разных дисциплин владельцы баз данных используют различные подходы и наработки. Так, в Clarivate для отбора контента работает большой штат специалистов, открыты представительства во многих странах мира, способ-

ствующие поиску и продвижению национальных журналов в Web of Science, налажены связи с крупными издательствами. В реферативных службах, например, в Chemical Abstracts и РЖ ВИНТИ, имеется штат предметных специалистов по мониторингу и отбору научной информации в тематические базы данных. Издательские дома соответственно во многом используют свой издательский опыт и возможности для ведения баз данных, как это реализовано в Scopus (Elsevier) или Scilit (MDPI). Собственные методы по наполнению баз данных, связанные со значительными технологическими возможностями, прямым доступом к веб-индексации и программными наработками по поиску информации, применяются корпорациями Google и Microsoft в их системах Google Academia и Microsoft Academic (массив данных закрытого в 2021 г. проекта Microsoft Academic в настоящее время используется в системе OpenAlex).

В последние два десятилетия появился ещё один подход к организации баз данных, связанный с открытыми лицензиями на распространение метаданных. В рамках этого подхода одни системы используют возможности и преимущества других, заимствуя их метаданные и дополняя собственными. В последние несколько лет это привело к появлению сразу нескольких новых информационных продуктов. На рис. 1 представлен список основных крупных политематических библиографических систем, образованных в XXI в.

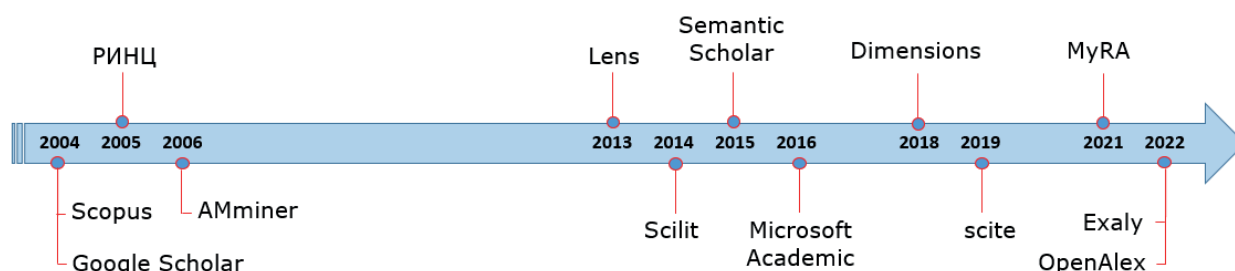


Рис. 1. Основные политематические библиографические базы данных, появившиеся в XXI в.

Важную роль в развитии новых баз данных сыграла система регистрации DOI Crossref, в которой представлены метаданные публикаций большого количества научных издателей, а также установлены взаимосвязи между публикациями на основе анализа пристатейных списков литературы [11]. Не являясь полноценной библиографической базой данных в общепринятом смысле, Crossref тем не менее выполняет важнейшую роль поставщика метаданных в прочие системы, при этом набор метаданных, распространяемых в открытом доступе, неизменно становится всё более разнообразным. Одновременно с этим расширяется набор типов индексируемых документов, куда уже включены отдельные части публикаций, наборы данных, рецензии, отчёты, стандарты и пр. Не меньшее влияние на развитие открытого доступа к метаданным оказал ныне закрытый проект Microsoft Academic, в рамках которого с 2016 по 2021 гг. были собраны десятки миллионов библиографических записей о научных публикациях [6; 12]. Активно заимствуются

открытые метаданные из тематических систем, таких как PubMed. Важную роль в развитии баз данных сыграл открытый проект уникальных авторских профилей ORCID [13], из которого многие системы собирают сведения о публикациях, грантах, местах работы и прочей информации, внесённой авторами в ORCID.

Образованию и развитию новых библиографических систем способствует не только доступ к основной библиографической информации, но и ко множеству дополнительных полей, которые также были охвачены инициативами открытого доступа. Так, к основной информации, включающей авторов научных публикаций, заглавия, источники опубликования и выходные сведения, в последние годы добавляются аннотации, списки литературы, информация об аффилиациях и финансировании и др. Например, аннотации переходят в открытый доступ в рамках инициативы I4OA – Initiative for Open Abstracts (<https://i4oa.org>) при их размещении самими издателями в Crossref. Значимым стал недавний проект I4OC – Initiative for Open Citations (<https://i4oc.org>), направленный на обеспечение всеобщего доступа к пристатейным спискам литературы научных публикаций с целью сделать анализ цитирования более доступным. Несмотря на противодействие данной инициативе некоторых издательств, в 2022 г. почти для половины всех статей из системы Crossref (60 из 134 млн) списки литературы были открытыми, тогда как прежде они присутствовали почти исключительно в платных библиографических указателях [14]. Схожие цели преследует и другой проект – OpenCitations (<https://opencitations.net>) [15].

Стоит упомянуть и переход в открытый доступ многих журнальных баз данных. В частности, полностью открыты рассчитываемые на массиве Scopus журнальные показатели в базе данных Scimago Journal Rank [16], а существенная часть информации из платной системы Journal Citation Reports доступна в бесплатной базе данных Master Journal List [17].

В табл. 2 приведены общие характеристики основных международных мультидисциплинарных библиографических баз данных открытого доступа, появившихся в XXI в. В сравнительных целях также приводятся сведения по системам платного доступа Web of Science, Scopus и Dimensions. Включена в обзор и система scite, которая, как отмечено выше, доступна для подписки отдельными исследователями за небольшую плату. В таблицу также включены данные по национальной системе РИНЦ, которая не ставит целью охватить всю международную литературу, но в то же время является важным информационным ресурсом регионального уровня, охватывая публикации, периодические издания, организации и авторов не только из России, но и из некоторых стран СНГ. В анализ не включены национальные системы других регионов, такие как южноамериканская Scielo, Korea Citation Index и некоторые другие по причине их низкой востребованности в России. Кроме того, не рассматриваются и социальные сети, основой которых являются библиографические метаданные, такие как ResearchGate или Academia, ввиду того что цели их создания, логика представления данных и инструментарий отличаются от основных целей библиографических баз данных.

Таблица 2

Общие характеристики международных политематических библиографических баз данных.
Ранжирование в порядке их образования (данные на апрель 2023 г.)

База данных	Владелец (страна)	Сайт	Тип данных	Число записей	Число цитирований	Число авторов	Активные индексируемые журналы	Основные источники наполнения	Открытый доступ	Русскоязычный интерфейс	Ссылки
Web of Science ¹	Clarivate (США)	https://www.webofscience.com/	журнальные статьи (≈86 млн в Web of Science CC), патенты (109 млн), наборы данных (14,5 млн)	196 млн	1,9 млрд	н/д	34 тыс., из них 21,5 тыс. в WoS CC	издатели, отбираемые экспертами Clarivate	-	+	[18]
Scopus	Elsevier (Нидерланды)	https://www.scopus.com/	преимущественно журнальные статьи	84 млн	1,8 млрд	17,6 млн	28 тыс.	издатели, отбираемые экспертами Elsevier	-	+	[19]
Google Scholar ²	Google (США)	https://scholar.google.com/	научные статьи, монографии, материалы конференций, диссертации	389 млн	н/д	н/д	н/д	издатели, репозитории, веб-сканирование	+	+	[20; 21]
РИНЦ	Научная электронная библиотека (Россия)	https://elibrary.ru/	научные статьи, монографии, материалы конференций, диссертации, патенты	41,3 млн	620 млн	1 млн	75,5 тыс., из них 5,6 тыс. в РИНЦ	издатели, прочие базы данных	+	+	[22; 23]
AMiner	Tsinghua University (Китай)	https://www.aminer.org/	журнальные статьи, материалы конференций	316 млн	1,1 млрд	130 млн	н/д	веб-сканирование	+	-	[24–26]
Lens	Cambia (Австралия)	https://www.lens.org/	научные публикации (255 млн), патенты (146 млн)	401 млн	1,9 млрд ²	н/д	51 тыс.	Mircosoft Academic, Crossref, PubMed, ORCID, OpenAlex	+/-	+	(машинный перевод) [27; 28]
Scilit	MDPI (Швейцария)	https://app.scilit.net/	научные статьи	155 млн	1,2 млрд	16 млн	114 тыс.	издатели, Crossref, PubMed, репозитории, ORCID	+	-	[29]
Semantic Scholar	Allen Institute for Artificial Intelligence (США)	https://www.semanticscholar.org/	н/д	211 млн	2,4 млрд	80 млн	н/д	Crossref, издатели, PubMed, репозитории, веб-сканирование	+/-	-	[30–32]

Продолжение табл. 2											
Microsoft Academic ³	Microsoft (США)	https://www.microsoft.com/en-us/research/project/academic/publications/	все типы публикаций	225 млн	2 млрд	н/д	н/д	веб-сканирование	+	–	[6; 12]
Dimensions	Digital Science (Великобритания)	https://www.dimensions.ai/	научные публикации (135 млн), патенты (151 млн), наборы данных (12 млн), гранты (6 млн), книги (1,6 млн)	305 млн	1,7 млрд	313 млн	137 тыс.	Crossref, PubMed, ORCID	+/-	–	[7; 33]
scite	scite Inc. (США)	https://scite.ai/	научные публикации	92 млн	1,2 млрд	260 млн	н/д	Crossref, PubMed,	–	–	[34]
Web of Science My Research Assistant (MyRA) ⁴	Clarivate (США)	–	преимущественно журнальные публикации	≈15 млн	н/д	–	21,5 тыс.	издатели	+	–	
OpenAlex	OurResearch (Канада)	https://openalex.org/	статьи, книги, наборы данных, диссертации	240 млн	н/д	213 млн	118 тыс.	Microsoft Academic, Crossref, PubMed, репозитории, ORCID, веб-сканирование	+	–	[35; 36]
Exaly ⁵	Exaly (н/д)	https://exaly.com/	журнальные публикации	143 млн	н/д	4,9 млн	80 тыс.	н/д	+	–	[37]

Примечания:
¹ Все базы данных Web of Science.
² Данные Google Scholar на 2018 г.
³ Данные Microsoft Academic на ноябрь 2019 г.
⁴ В мобильном приложении Web of Science My Research Assistant доступны данные Web of Science за последние 5 лет. Расчёт числа публикаций проведён авторами.
⁵ Данные в Exaly заканчиваются 2021 годом. На запрос авторов об индексировании статей за последующие годы разработчики не ответили.

Из данных табл. 2 видно, что наполнение большинства недавно образованных систем существенно выше, чем у Web of Science и Scopus. Это позволяет рассматривать их как полноценную замену при решении информационно-поисковых задач. Проиндексированные же пристатейные списки литературы позволяют использовать соответствующие системы и в различных библиометрических целях. При этом многие базы данных демонстрируют прозрачные алгоритмы формирования источниковой основы, что повышает степень достоверности библиометрических исследований, проводимых на массивах этих систем.

При выборе баз данных для поисковой работы и библиометрических исследований не менее важными являются следующие критерии.

1. Полнота метаданных: с учётом неуклонно расширяющегося в последние годы набора метаданных научных публикаций, а также необходимости их массовой обработки, полнота метаданных играет важную роль в выборе

библиографической системы. К набору нужных сведений и для эффективного поиска научной литературы, и для проведения библиометрического анализа, кроме обязательных заглавия, строки авторов, источника и выходных данных, можно отнести также:

- аннотации – необходимы для более «тонкого» проведения информационного поиска;
- ключевые слова – нужны для проведения информационного поиска и предметной классификации;
- идентификаторы публикаций (DOI, внутренние идентификаторы в системах) – востребованы при отсылке к полным библиографическим описаниям и текстам, а также при создании глобальных запросов и библиометрических расчётов;
- аффилиации авторов – важны для расчёта доли участия авторов / коллективов / организации в исследованиях, для определения авторских коллабораций и др.;
- финансирование исследований – используется для отчётных целей по грантам, госзаданиям и пр.;
- коды тематических рубрик – необходимы для создания запросов по дисциплинарным направлениям.

2. Широкий набор фильтров, непосредственно связанный с полнотой индексируемых метаданных. Для эффективного решения многих поисковых задач крайне важно наличие возможности поиска по каждому индексируемому полю. Принцип использования единой поисковой строки с последующим ранжированием результатов по скрытым от пользователя алгоритмам, без предоставления возможности самостоятельной фильтрации результатов, зачастую не даёт нужных результатов при решении поисковых задач.

3. Система оповещений, которая позволяет исследователям при правильной настройке отслеживать появление новых работ по тематическим запросам, а научным библиотекам и прочим службам – в автоматическом режиме и на периодической основе получать сведения об индексировании новых публикаций сотрудников организации и цитировании ранее опубликованных работ.

4. Полнота выгрузки метаданных, а также возможность выбора и настройки форматов их выгрузки. Данная опция нужна для проведения библиометрических исследований, в том числе с использованием специальных программ, для автоматизации процессов ведения внутренней базы данных организации, сокращения времени на ручную обработку записей.

5. Реализация расширенного поиска, создания глобальных и сводных запросов. Эта функция необходима как для настройки оповещений по широкому запросу, так и для решения библиометрических задач.

6. Система авторских профилей, призванная упростить поиск информации о публикациях исследователя, повысить степень представленности его работ международному научному сообществу, а также сделать более точной библиометрическую оценку его публикационной активности. Из универсальных идентификаторов на текущий момент наибольшее распространение получил ORCID [13], который активно используется в электронных редакциях

журналов, профилях рецензентов, указывается в научных публикациях, а также внедряется в библиографические базы данных. Между тем для эффективной работы научных библиотек и прочих служб при выполнении разных информационно-поисковых задач более полезными оказываются внутренние идентификаторы авторов. Их наличие существенно повышает точность поиска нужных сведений в определённой системе, несмотря на все ограничения, связанные с дублями, неточной атрибуцией публикаций однофамильцев и т. д. [38]. Так, становится возможным объединение авторских профилей в единый запрос для тематического поиска, например, по лаборатории или факультету, установка оповещений о новых публикациях организации или её подразделения и пр. [39; 40].

7. Реализация поиска по пристатейной литературе также представляется важной, поскольку списки цитирований являются краеугольным камнем проведения библиометрического анализа. Кроме того, авторы могут отыскивать несвязанные с их профилями цитирования на их публикации и привязывать к соответствующим публикациям.

8. Наличие подробного тематического рубрикатора, значимого как при проведении информационного поиска, так и для решения библиометрических задач.

В табл. 3 представлена степень соответствия анализируемых систем вышеприведенным критериям.

Таблица 3

Характеристики международных политематических библиографических баз данных, определяющие эффективность их использования в решении информационно-поисковых и библиометрических задач

База данных	Веб-интерфейс / API	Полнота метаданных	Широкий набор фильтров	Оповещения	Полнота выгрузки	Глобальный / сводный запрос	Возможность создания / корректировки профиля автора	Наличие авторских профилей с уникальными ID ¹	Поиск по пристатейной литературе	Рубрикатор, число рубрик
Web of Science ²	+ / +	+	+	+	+	+	+	+	+	Категории WoS, 21 – верхний уровень; 254 – нижний уровень; 2500 тематик
Scopus ²	+ / +	+	+	+	+	+	+	+	+	ASJC, 4 – верхний уровень; 30 – средний уровень; 330 – нижний уровень
Google Scholar	+ / –	–	–	+	–	–	+	+	–	н/д
РИНЦ	+ / +	+	+	–	– ³	–	+	+	+	ГРНТИ, 100 – верхний уровень; >800 – средний уровень; >7000 – нижний уровень
AMiner	+ / +	–	–	+	–	–	+	–	–	8,7 млн концепций, выделенных на основе машинного обучения
Lens	+ / +	+	+	+	+	+	+	–	+	см. OpenAlex
Scilit	+ / +	–	+	+	–	–	–	–	–	н/д
scite	+ / +	–	+	+	+	–	+	+	–	см. OpenAlex
Semantic Scholar	+ / +	–	–	+	–	–	+	+	–	24

<i>Продолжение табл. 3</i>										
Microsoft Academic	+ / +	<i>неприменимо</i>								<i>см. OpenAlex</i>
Dimensions	+ / +	+	+	+	+	+ ⁴	+	+	+	ANZSRC, 236; несколько млн концепций, выделенных на основе машинного обучения
Web of Science My Research Assistant (MyRA)	+ / -	-	+	+	-	-	-	-	-	Категории WoS, 21 – верхний уровень; 254 – нижний уровень
OpenAlex	- / +	+	+	-	+	+	-	-	+	19 – верхний уровень; 65 тыс. тематик на основе машинного обучения
Exaly	+ / -	-	+	-	-	-	-	-	-	8 – верхний уровень, 93 – средний уровень, 272 – нижний уровень; 69 тыс. тематик на основе автоматической обработки естественного языка
Примечания:										
¹ В Exaly, Lens и Scilit алгоритм назначения внутренних авторских идентификаторов непрозрачный. Он присваивается при регистрации и редактировании списка своих публикаций. В отдельных случаях идентификаторы могут быть присвоены авторам, не прошедшим регистрацию, тогда как у других авторов с неменьшим числом публикаций идентификаторы отсутствуют.										
² Только в платной версии.										
³ Выгрузка в РИНЦ реализована только при подписке на Science Index для организаций и охватывает только публикации организации.										
⁴ Сводный запрос в Dimensions формируется в адресной строке.										

Данные табл. 3 подтверждают, что Web of Science и Scopus даже при возникновении существенной конкуренции остаются наиболее функциональными системами, хотя и уступают по широте индексации большинству из новых информационных ресурсов. В то же время видно, что фактически все системы открытого доступа могут эффективно использоваться для информационного поиска, а такие базы данных, как Dimensions в её бесплатном варианте и Lens, могут применяться и как полноценная альтернатива коммерческим продуктам. Кроме того, каждая из этих систем имеет свою характерную особенность, которая позволяет наиболее эффективно использовать её для решения тех или иных задач.

ОСОБЕННОСТИ БИБЛИОГРАФИЧЕСКИХ БАЗ ДАННЫХ

Доступные по подписке системы Web of Science и Scopus многократно и подробно исследованы, поэтому мы опускаем их описание. За подробным обзором отсылаем читателей к недавней работе [41].

Несмотря на то что в основе многих систем лежат одни и те же данные из Crossref, Microsoft Academic или тематических баз данных, в каждом случае ядро заимствованных метаданных обогащается дополнительной библиографической информацией, предлагаются оригинальные, в том числе альтернативные, метрики оценки научных публикаций или проводится глубокий семантический анализ полных текстов. Это определяет и общую содержательную направленность, а также круг пользователей той или иной системы.

Google Scholar, или Google Академия нацелена на максимально возможный сбор информации, что, однако, оборотной стороной имеет слабую защищённость от спама и неverified публикаций [42]. Система де-

дубликации записей, на наш взгляд, неэффективна, а интерфейс для поиска и анализа библиографической информации относительно скудный.

Российский индекс научного цитирования (РИНЦ) является основным региональным источником библиографических метаданных о российских и русскоязычных публикациях, а также публикациях некоторых стран СНГ. База данных вполне могла бы использоваться для оценки российской науки как система, наиболее полно индексирующая отечественные публикации. Для решения же поисковых и библиометрических задач, стоящих перед исследователями и специалистами по научной информации, РИНЦ во многом не подходит и не может полноценно заместить международные системы. Прежде всего это связано с тем, что РИНЦ не индексирует на регулярной основе зарубежные источники, поэтому невозможно проводить качественный информационный поиск, учитывающий международные потоки публикаций. Кроме того, как видно из табл. 3, в РИНЦ до сих пор не реализована возможность установки оповещений, нет возможности проводить комплексные запросы, а выгрузка метаданных реализована лишь в одном формате и доступна только организациям, подписанным на надстройку Science Index, что существенно препятствует использованию системы для библиометрических исследований.

AMiner нацелена на сетевой анализ научных объектов и сконцентрирована прежде всего на исследователях, их влиянии на общество и визуализации этого влияния, рекомендациях по построению коллабораций, анализу взаимосвязей между авторами, предоставлении возможности к изучению развития научного сообщества [25; 26].

Lens как библиографическая база данных характерна предоставлением доступа к большим объёмам патентной информации и широкими возможностями их анализа, а также внедрении мета-идентификатора записей, объединяющего в себе все прочие идентификаторы научного объекта – DOI, ORCID, идентификаторы публикации в других системах и пр. [8; 28]. Многочисленные функциональные возможности Lens наиболее приближены к таковым в Web of Science и Scopus. Дополнительные расширенные возможности предоставляются по отдельной подписке, доступной как отдельным пользователям, так и организациям. В то же время подписка преимущественно требуется при коммерческом использовании данных или для решения внутренних задач организаций. Для решения поисковых и библиометрических задач в научных целях подписка не требуется, а набор открытых функциональных возможностей сопоставим с таковым в Web of Science или Scopus.

Scilit издательства MDPI позиционирует себя как систему, которая включает всю опубликованную научную литературу, собранную на единой платформе, с быстрой индексацией, удобную широкому кругу пользователей, преследующих цели повышения видимости научных результатов, поиска научной информации, библиометрической оценки и пр. [29].

scite указывает на возможность решения множества задач для широкого круга пользователей. Основной особенностью системы является применение машинного обучения в решении давней задачи по выявлению коннотации цитирования [34]. На 25-миллионной полнотекстовой коллекции выделены три типа цитирования: поддерживающие определённые результаты, опровер-

гающие их или просто упоминающие ту или иную работу. Иными словами, делается попытка разграничить значительные, отрицательные и поверхностные цитирования (см. также [43]). По мнению разработчиков, такой подход позволяет исследователям быстрее ориентироваться в больших потоках научной информации путём отсеивания публикаций с отрицательными цитированиями, научным журналам – дополнительно оценивать природу ссылок на публикуемые материалы; организациям – лучше понимать контекст своих публикаций и пр. Кроме того, система позволяет решить задачу с цитированиями ретрагированных статей или ошибочных ссылок (см., например, [44]): в ней проводится проверка точности цитирований и актуальный статус публикаций с выдачей предупреждения авторам об отозванных из печати статьях или ошибках в метаданных. По-видимому, данный подход будет особенно востребован в связи с распространением публикаций, написанных при участии нейронных сетей, где важной проблемой являются вымышленные и в действительности несуществующие цитируемые источники.

Semantic Scholar основные усилия концентрирует на облегчении информационного поиска в условиях растущих объёмов научной информации, в том числе для поиска рецензентов, для чего в анализе собранных полнотекстовых коллекций применяется искусственный интеллект [31; 32]. Его результаты используются как в ранжировании результатов поиска, так и в библиометрических задачах оценки публикаций. В частности, предложены новые подходы к определению веса цитирований на основе их расположения в том или ином разделе статьи [45]. Нужно отметить, что по такому же пути пошли разработчики Web of Science, предложив графический инструмент анализа цитирований «Расширенная пристатейная библиография» (Enriched Cited References) [46], и система scite, где также учитывается место цитирований в структуре цитирующей публикации [34]. Хотя к точности подобных алгоритмов у исследователей пока остаются вопросы [47], развитие инструментов анализа цитирования уверенно идёт в указанном направлении.

Microsoft Academic основную цель видела в охвате как можно более широкого круга научных документов, особенно тех, которые распространяются в обход традиционной журнальной модели, используют тип открытого рецензирования и редко используют DOI, отчего не индексируются большинством баз данных [6; 12]. Кроме того, в Microsoft Academic в основном было решено полагаться на результаты машинного обучения в обработке метаданных, а не на усилия большого числа экспертов, как это реализовано в коммерческих базах данных. Такой подход, по мнению разработчиков, должен был существенно повысить как точность метаданных, так и объективность результатов различных видов оценки. Проект прекратил существование в 2021 г., а накопленные данные перешли в новый проект OpenAlex.

Dimensions уникальна нацеленностью на отслеживание полного цикла научной работы – от её замысла на этапе заявок на финансирование до реализации в виде конечных продуктов (патентов, клинических исследований или директив) [7; 48]. Наряду с Lens, Dimensions может рассматриваться как наиболее близкий, но во многом открытый аналог Web of Science и Scopus, что подтверждают многочисленные исследования (см. ниже). Часть функций, как и в Lens, доступна по подписке, однако открытой части системы

вполне достаточно для решения как информационно-поисковых, так и библиометрических задач.

OpenAlex как продолжатель **Microsoft Academic** имеет схожие с этой системой цели и особенности [35; 36].

Exaly предлагает пользователю уникальные данные лингвистического анализа полных текстов на основе алгоритмов обработки естественного языка. Например, пользователю представлены наиболее часто употребляемые в публикациях слова и части речи, что даёт дополнительную ценную информацию для глубокого контент-анализа публикационной активности исследователя.

Таким образом, в начале XXI в. 40-летняя история почти безальтернативного доминирования **Web of Science (SCI)** в области библиографической информации подошла к концу не только из-за появления конкурирующей системы **Scopus**, но и в связи с ускоряющимся переходом различных метаданных в открытый доступ. Такая ситуация сделала возможным появление целого ряда новых библиографических систем в разных странах, многие из которых были образованы в последнее десятилетие. При этом открытый доступ и высокая конкуренция затронули сами **Web of Science** и **Scopus**, которые с каждым годом выкладывают в общий доступ всё больше индексируемых данных о публикациях.

На начало 2023 г. в обеих системах в общедоступном режиме был возможен просмотр основной библиометрической информации по авторам. При этом если до недавнего времени в **Web of Science** обновление авторских профилей и получение идентификаторов **ResearcherID** проводилось лишь при непосредственном участии авторов, то в 2022 г. **ResearcherID** были инициативно присвоены **Clarivate** всем авторам, имеющим хотя бы одну публикацию в **Web of Science**, а с апреля 2023 г. **Clarivate** наконец-то внедрила систему автообновления библиометрической информации по профилям авторов [49]. В дополнение к основной библиометрической информации по авторам, в открытой веб-версии баз данных **Web of Science Core Collection** доступны для просмотра полные списки публикаций авторов, а также списки цитирующих их публикаций.

Кроме того, заслуживает безусловного внимания мобильная версия **Web of Science My Research Assistant (MyRA)**, в которой возможен полнофункциональный поиск всех публикаций основной коллекции **Web of Science** за последние 5 лет, использование развёрнутой системы фильтров поиска, установка оповещений, создание подборок и экспорт записей в формате **RIS** для использования в программах управления ссылками. Этого более чем достаточно для того, чтобы отслеживать новейшие разработки в своих областях знаний для большинства исследователей.

Беглый взгляд на инициативы открытого доступа и внедряемые в библиографические системы функции позволяет предположить, что ценность и цена накопления данных как таковых в последние годы значительно снизились. В жёсткой конкурентной борьбе за внимание пользователя на первое место выходят удобство использования системы, глубина предметизации индексируемых объектов для облегчения и уточнения поиска, а также уникальные

подходы к анализу больших массивов научной информации с применением новейших технологий искусственного интеллекта, автоматической обработки естественного языка, визуализации результатов анализа и пр. Многие инициативы направлены на решение задачи по более глубокому пониманию природы цитирований, их целей и коннотации, что издавна привлекало [50–52] и продолжает привлекать [43; 53–55] внимание исследователей. Если прежние подходы к определению веса цитирований определялись по источнику в целом (например, ссылка из Nature в показателе SJR считается более значимой, чем ссылка из какого-либо нового журнала с низким рейтингом [56]), то в новых подходах вес зависит непосредственно от контекста ссылки. Безусловно, такая ситуация способствует развитию информационных библиографических систем и несёт значительную выгоду всем заинтересованным пользователям.

СРАВНЕНИЕ БИБЛИОГРАФИЧЕСКИХ СИСТЕМ В НАУЧНОЙ ЛИТЕРАТУРЕ

В международной литературе на регулярной основе появляются публикации, в которых проводится сравнение библиографических систем с различных точек зрения. Чаще всего сравнения проводят с Web of Science и Scopus. Исчерпывающий обзор со сравнительным анализом наполнения, фильтров и метрик в этих двух системах приводится в работе [41]. Особенности индексации различных типов документов в этих системах освещены в статье [57]. Далее перечислены недавние работы (последних пяти лет) со сравнительным анализом прочих библиографических систем.

Сравнение Google Scholar с Web of Science и Scopus проводится с позиций как охвата литературы этими базами данных, так и объёмов учитываемых в них цитирований научной литературы [5; 58; 59]. Уровень цитирований определённой группы исследователей в области медицины в этих системах исследуется в работе [60]. Ещё в одном исследовании этих же трёх систем на выборке журналов по управлению производством делается вывод о надёжности информации о цитировании в Google Scholar, так что эту систему можно использовать в областях, слабо охваченных Scopus и Web of Science [61].

Для выявления степени применимости Crossref, Lens, OpenAlex, Semantic Scholar, Scopus и Web of Science в библиометрических целях недавно проведён детальный анализ индексируемых этими системами метаданных [62]. Объём записей сразу для 12 библиографических систем оценивался в работе [21], где сделан вывод о преимуществе Google Scholar перед всеми современными системами в охвате научной литературы.

В качестве наиболее серьёзного конкурента Web of Science и Scopus чаще всего упоминают Dimensions – продукт компании Digital Science, принадлежащий издательскому дому Holtzbrinck Publishing Group, в который также входит издательство Springer Nature. Dimensions позиционируется как система с большим охватом источников, в значительной части бесплатная для

использования, но в то же время достаточно надёжная в обеспечении качества индексируемого контента, в отличие, например, от Google Scholar. Так, при сравнении Web of Science, Scopus и Dimensions было определено, что в Dimensions содержится на 82,22 и 48,17% журналов больше, чем в Web of Science и Scopus соответственно, тогда как доля отсутствующих в Dimensions журналов из двух других баз данных минимальна и не превышает 1–3% [63]. В этом же исследовании в качестве серьёзного преимущества Dimensions отмечена применяемая в ней классификация на уровне публикаций, а не журналов, что существенно расширяет возможности тематического анализа информации.

В сравнительном исследовании Scopus и Dimensions на выборке в 10 тыс. статей показано индексирование 97% публикаций из Scopus в базе данных Dimensions, а также сопоставимый уровень цитирований этих публикаций, что позволяет рассматривать Dimensions как достойную альтернативу Scopus [64]. Анализ охвата экономических дисциплин в Dimensions и Crossref в сравнении с другими известными системами также показал их потенциальные преимущества как в поисковых, так и в библиометрических целях с учётом их открытости [65].

При сравнении Dimensions со Scopus и Google Scholar на уровне журналов, публикаций и цитирований были продемонстрированы преимущества Dimensions перед Scopus в объёме охвата научной литературы и перед Google Scholar – в функциональности обработки и выгрузки данных [66]. Отдельно Dimensions и Web of Science исследовались с точки зрения точности отображения данных о финансировании научных разработок [67]. При сопоставлении Dimensions, Crossref, Scopus и Web of Science показана существенная корреляция данных между Crossref и зависимой от неё Dimensions, а также больший охват документов в Dimensions в сравнении со Scopus [68]. В то же время отмечена низкая научная ценность дополнительно индексируемых документов, например, таких как список рецензентов определённого журнала.

При сравнении Dimensions и Web of Science с точки зрения полноты индексации статей в открытом доступе сделан вывод о преимуществах Dimensions в охвате полностью открытых публикаций [69]. Проводятся исследования на выбор наилучшей системы для анализа национальной науки. Так, например, немецкая наука лучше представлена и имеет лучшие библиометрические показатели в Dimensions и Scopus в сравнении с Web of Science [70]. Преимущества Dimensions перед Microsoft Academic, Crossref и Semantic Scholar показаны при анализе цитирований в этих системах корпуса публикаций тематической базы Pubmed [71].

В меньшей степени сравнение проводится в отношении систем Scilit, AMiner, Lens, хотя они активно используются как источники данных в различных обзорах и метаанализах. Так, возможности базы данных Lens на выборке патентной литературы по коронавирусам представлены в работе [72]. Система Scilit нередко используется как источник данных при написании медицинских обзоров [73; 74]. Только что открытые системы Exaly и OpenAlex ещё ожидают своих исследователей.

ЗАКЛЮЧЕНИЕ

Резюмируя сказанное, можно сделать вывод о наличии целого ряда открытых библиографических ресурсов, систематически и оперативно индексирующих потоки международной научной информации, которые в информационно-поисковых и библиометрических целях могут стать полноценной заменой недоступным по разным причинам коммерческим системам. Каждая из баз данных имеет уникальный набор свойств, инструментов и функций, которые в сочетании могут дать даже лучшие результаты при решении различных задач. Особенно это касается аналитических возможностей, которые занимают лидирующее положение по отношению к метаданным как таковым в конкурентной борьбе за конечного пользователя информационных систем. В частности,

- практически все системы открытого доступа могут эффективно использоваться для поиска научной литературы, необходимого для ознакомления с новейшими исследованиями в различных областях знаний, написания обзорных материалов, поиска коллег для совместных работ и определения конкурирующих научных коллективов;
- с помощью Dimensions и Lens можно эффективно решить вопрос сбора сведений о публикациях сотрудников научно-образовательных организаций, особенно опубликованных в зарубежных источниках, что насущно для научных библиотек и информационных служб, ведущих репозитории и институциональные базы данных. Поскольку многие сотрудники напрямую не заинтересованы в отслеживании своей публикационной активности, зачастую не предоставляют в организацию публикации по грантам или без аффилиации, крайне важным является их автоматизированный сбор во внешних системах;
- широко представленные в международных базах данных публикации научных сотрудников являются важным ресурсом для популяризации их достижений. С этой целью, например, в Dimensions, Semantic Scholar и Lens представлены эффективные инструменты для создания и поддержки авторских профилей;
- библиометрические задачи, включая оптимизацию комплектования, оценку публикационной активности отдельных авторов, а также более сложные задачи анализа научных регионов, журналов, организаций, научных фронтов и пр. могут решаться в Lens и Dimensions, а при соответствующих навыках работы с API и в OpenAlex. В свою очередь Semantic Scholar, scite или Exaly позволяют решать традиционные библиометрические задачи с совершенно новых позиций благодаря глубокому семантическому анализу полных текстов научных публикаций.

ЛИТЕРАТУРА

1. Web of Science Journal Evaluation Process and Selection Criteria // Clarivate : [сайт]. 2023. URL: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/core-collection/editorial-selection-process/editorial-selection-process/> (дата обращения: 21.04.2023).
2. Content Policy and Selection // Elsevier : [сайт]. 2023. URL: <https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection> (дата обращения: 21.04.2023).
3. *Vera-Baceta M. A.* Web of Science and Scopus language coverage / M. A. Vera-Baceta, M. Thelwall, K. Kousha // *Scientometrics*. 2019. Vol. 121, № 3. P. 1803–1813. DOI 10.1007/s11192-019-03264-z.
4. *Mongeon P.* The journal coverage of Web of Science and Scopus: a comparative analysis / P. Mongeon, A. *Paul-Hus* // *Scientometrics*. 2016. Vol. 106, № 1. P. 213–228. DOI 10.1007/s11192-015-1765-5.
5. *Martín-Martín A.* Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison / A. Martín-Martín, E. Orduna-Malea, E. Delgado López-Cózar // *Scientometrics*. 2018. Vol. 116, № 3. P. 2175–2188. DOI 10.1007/s11192-018-2820-9.
6. Microsoft Academic Graph: When experts are not enough / K. Wang, Z. Shen, C. Huang [et al.] // *Quantitative Science Studies*. 2020. Vol. 1, № 1. P. 396–413. DOI 10.1162/qss_a_00021.
7. *Herzog C.* Dimensions: Bringing down barriers between scientometricians and data / C. Herzog, D. Hook, S. Konkiel // *Quantitative Science Studies*. 2020. Vol. 1, № 1. P. 387–395. DOI 10.1162/qss_a_00020.
8. *Penfold R.* Using the Lens database for staff publications // *Journal of the Medical Library Association*. 2020. Vol. 108, № 2. P. 341–344. DOI 10.5195/jmla.2020.918.
9. *Семёнов Е. В.* Позади – год напрасных ожиданий, впереди – год надежд и неотложных дел // *Управление наукой: теория и практика*. 2023. Т. 5, № 1. С. 10–13. EDN ANLSAF.
10. *Мазов Н. А.* IPGGTR Труды сотрудников ИНГГ СО РАН (реферативно-полнотекстовая библиография): Свидетельство о государственной регистрации программы для ЭВМ / Н. А. Мазов, В. Н. Гуреев // Свид-во о прогр. 2020621025; RU; № 2020620872, заявл. 10.06.2020, опубли. 19.06.2020, ИНГГ СО РАН. URL: https://new.fips.ru/registers-doc-view/fips_servlet?DB=DB&DocNumber=2020621025&Type-File=html (дата обращения: 21.04.2023).
11. *Hendricks G.* Crossref: The sustainable source of community-owned scholarly metadata / G. Hendricks, D. Tkaczyk, J. Lin, P. Feeney // *Quantitative Science Studies*. 2020. Vol. 1, № 1. P. 414–427. DOI 10.1162/qss_a_00022.
12. A Review of Microsoft Academic Services for Science of Science Studies / K. Wang, Z. Shen, C. Huang [et al.] // *Frontiers in Big Data*. 2019. Vol. 2. Art. no. 45. DOI 10.3389/fdata.2019.00045.
13. ORCID: a system to uniquely identify researchers / L. L. Haak, M. Fenner, L. Paglione [et al.] // *Learned Publishing*. 2012. Vol. 25, № 4. P. 259–264. DOI 10.1087/20120404.
14. *Chawla D. S.* Five-year campaign breaks science’s citation paywall // *Nature*. 2022. September 13. DOI 10.1038/d41586-022-02926-y.
15. *Peroni S.* OpenCitations, an infrastructure organization for open scholarship / S. Peroni, D. Shotton // *Quantitative Science Studies*. 2020. Vol. 1, № 1. P. 428–444. DOI 10.1162/qss_a_00023.
16. SJR and SNIP: two new journal metrics in Elsevier’s Scopus / L. Colledge, F. De Moya-Anegón, V. Guerrero-Bote [et al.] // *Serials*. 2010. Vol. 23, № 3. P. 215–221. DOI 10.1629/23215.

17. Master Journal List // Clarivate : [сайт]. 2023. URL: <https://mjl.clarivate.com/home> (дата обращения: 21.04.2023).
18. Web of Science Core Collection // Clarivate : [сайт]. 2023. URL: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/web-of-science-core-collection/> (дата обращения: 21.04.2023).
19. Scopus. Expertly curated abstract & citation database // Elsevier : [сайт]. 2023. URL: <https://www.elsevier.com/solutions/scopus> (дата обращения: 21.04.2023).
20. Google Scholar // Google : [сайт]. 2023. URL: <https://scholar.google.com/intl/ru/scholar/about.html> (дата обращения: 21.04.2023).
21. *Gusenbauer M.* Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases // *Scientometrics*. 2019. Vol. 118, № 1. P. 177–214. DOI 10.1007/s11192-018-2958-5.
22. Российский индекс научного цитирования // Научная электронная библиотека : [сайт]. 2023. URL: https://www.elibrary.ru/project_risc.asp (дата обращения: 21.04.2023).
23. Russian Index of Science Citation: Overview and review / O. Moskaleva, V. Pislyakov, I. Sterligov [et al.] // *Scientometrics*. 2018. Vol. 116, № 1. P. 449–462. DOI 10.1007/s11192-018-2758-y.
24. AMiner // Tsinghua University : [сайт]. 2023. URL: <https://www.aminer.org/manual#aminer-manual-subTitle-research-feed> (дата обращения: 21.04.2023).
25. *Tang J.* AMiner: Toward Understanding Big Scholar Data // *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM'16)* (22–25 February 2016, San Francisco). 2016. P. 467.
26. ArnetMiner: Extraction and Mining of Academic Social Networks / J. Tang, J. Zhang, L. Yao [et al.] // *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)* (14–27 August 2008, Las Vegas). Association for Computing Machinery, 2008. P. 990–998.
27. About The Lens // Cambia : [сайт]. 2023. URL: <https://about.lens.org/> (дата обращения: 21.04.2023).
28. The Lens MetaRecord and LensID: An open identifier system for aggregated meta-data and versioning of knowledge artefacts / O. A. Jefferson, D. Koellhofer, B. Warren, R. Jefferson // *Researchgate* : [сайт]. 2019. URL: <https://www.researchgate.net/publication/337496596> (дата обращения: 15.05.2023). DOI 10.31229/osf.io/t56yh.
29. Scilit brochure // MDPI : [сайт]. 2023. URL: <https://app.scilit.net/scilit-brochure.pdf> (дата обращения: 21.04.2023).
30. About Semantic Scholar // Allen Institute for AI : [сайт]. 2023. URL: <https://www.semanticscholar.org/about> (дата обращения: 05.04.2023).
31. The Semantic Scholar Open Data Platform / R. Kinney, C. Anastasiades, R. Authur [et al.] // 2023. P. 1–8. DOI 10.48550/arXiv.2301.10140.
32. Construction of the Literature Graph in Semantic Scholar / W. Ammar, D. Groeneveld, C. Bhagavatula [et al.] // *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2–4 June 2018, New Orleans). 2018. Vol. 3. P. 84–91. DOI 10.18653/v1/n18-3011.
33. The data in Dimensions – from idea to impact // *Digital Science* : [сайт]. 2023. URL: <https://www.dimensions.ai/dimensions-data/> (дата обращения: 05.04.2023).
34. scite: A smart citation index that displays the context of citations and classifies their intent using deep learning / J. M. Nicholson, M. Mordaunt, P. Lopez [et al.] // *Quantitative Science Studies*. 2021. V. 2, № 3. P. 882–898. DOI 10.1162/qss_a_00146.
35. OpenAlex API documentation // OurResearch : [сайт]. 2023. URL: <https://docs.openalex.org/api-entities/works> (дата обращения: 21.04.2023).
36. *Priem J.* OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts / J. Priem, H. Piwowar, R. Orr // *26th International Conference on*

Science and Technology Indicators “From Global Indicators to Local Applications” (STI 2022) (7–9 September, Granada, Spain), 2022. P. 1–5.

37. Exaly Statistics // Exaly : [сайт]. 2023. URL: <https://exaly.com/statistics.html> (дата обращения: 21.04.2023).

38. Mazov N. A. The role of unique identifiers in bibliographic information systems / N. A. Mazov, V. N. Gureev // Scientific and Technical Information Processing. 2014. Vol. 41, № 3. P. 206–210. DOI 10.3103/S0147688214030101.

39. Mazov N. A. Publication Databases of Research Organizations as a Tool for Information Studies / N. A. Mazov, V. N. Gureev // Scientific and Technical Information Processing. 2022. Vol. 49, № 2. P. 108–118. DOI 10.3103/s0147688222020071.

40. Мазов Н. А. Библиографическая база данных трудов сотрудников организации: цели, функции, сфера использования в наукометрии / Н. А. Мазов, В. Н. Гуреев // Вестник Дальневосточной государственной научной библиотеки. 2016. № 2. С. 84–87. EDN YFMJGN.

41. Prancutè R. Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today’s Academic World // Publications. 2021. Vol. 9, № 1. Art. no. 12. DOI 10.3390/publications9010012.

42. Delgado López-Cózar E. The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators / E. Delgado López-Cózar, N. Robinson-García, D. Torres-Salinas // Journal of the Association for Information Science and Technology. 2014. Vol. 65, № 3. P. 446–454. DOI 10.1002/asi.23056.

43. Тихонова Е. В. Культура цитирования: поведение цитирующих авторов vs доверие к результатам научных исследований / Е. В. Тихонова, О. В. Кириллова // Научный редактор и издатель. 2022. Т. 7, № 2. С. 166–181. DOI 10.24069/SEP-22-58.

44. Бобров Л. К. Достоверность ссылок на научные издания: пример порождения мифов и неточностей // Научные и технические библиотеки. 2022. № 5. С. 47–65. DOI 10.33186/1027-3689-2022-5-47-65. EDN TIASTV.

45. Valenzuela M. Identifying meaningful citations / M. Valenzuela, V. Ha, O. Etzioni // 29th AAAI Conference on Artificial Intelligence, AAAI 2015 (25–30 January 2015, Austin, United States). AI Access Foundation, 2015. P. 21–26.

46. New WoS April 29 Release Notes: Enriched cited references, Export to Publons, and more // Clarivate : [сайт]. 2022. URL: <https://clarivate.com/webofsciencegroup/release-notes/wos/new-wos-april-29-release-notes/> (дата обращения: 21.04.2023).

47. Evaluating Research Impact Based on Semantic Scholar Highly Influential Citations, Total Citations, and Altmetric Attention Scores: The Quest for Refined Measures Remains Illusive / L. A. Dardas, M. Sallam, A. Woodward [et al.] // Publications. 2023. Vol. 11, № 1. Art. no. 5. DOI 10.3390/publications11010005.

48. Hook D. W. Dimensions: Building Context for Search and Evaluation / D. W. Hook, S. J. Porter, C. Herzog // Frontiers in Research Metrics and Analytics. 2018. Vol. 3. Art. no. 23. DOI 10.3389/frma.2018.00023.

49. Web of Science Release Notes, April 13 2023: Automatic updates to claimed profiles... // Clarivate : [сайт]. 2023. URL: <https://clarivate.com/webofsciencegroup/release-notes/wos/web-of-science-release-notes-april-13-2023-2/> (дата обращения: 21.04.2023).

50. Ramos M. A. Citation behavior in popular scientific papers: What is behind obscure citations? The case of ethnobotany / M. A. Ramos, J. G. Melo, U. P. Albuquerque // Scientometrics. 2012. Vol. 92, № 3. P. 711–719. DOI 10.1007/s11192-012-0662-4.

51. Деревянко А. П. Проблема качественного анализа археологических публикаций / А. П. Деревянко, Ю. П. Холюшкин // Методология и методика археологических реконструкций: Сборник научных трудов. Новосибирск : СО РАН, 1994. С. 24–32.

52. *Simkin M. V.* Read before you cite! / M. V. Simkin, V. P. Roychowdhury // *Complex Systems*. 2003. Vol. 14, № 3. P. 262–274. DOI 10.25088/ComplexSystems.14.3.269.

53. *Лазарев В. С.* «Цитируемость нобелевского класса» и понятия, выражающие характеристики и свойства цитируемых научных документов. Тамбов ; Москва ; С.-Петербург ; Баку ; Вена ; Гамбург ; Стокгольм ; Буаке : Международный Информационный Нобелевский Центр (МИНЦ), 2018. 70 с. EDN XWCYBV.

54. *Писляков В. В.* Самоцитирование и его влияние на оценку научной деятельности: обзор литературы. Часть I // *Научные и технические библиотеки*. 2022. № 2. С. 49–70. DOI 10.33186/1027-3689-2022-2-49-70. EDN GWNIEW.

55. *Писляков В. В.* Самоцитирование и его влияние на оценку научной деятельности: обзор литературы. Часть II // *Научные и технические библиотеки*. 2022. № 3. С. 85–104. DOI 10.33186/1027-3689-2022-3-85-104. EDN YEHRNX.

56. *González-Pereira B.* A new approach to the metric of journals' scientific prestige: The SJR indicator / B. González-Pereira, V. P. Guerrero-Bote, F. Moya-Anegón // *Journal of Informetrics*. 2010. Vol. 4, № 3. P. 379–391. DOI 10.1016/j.joi.2010.03.002.

57. *Мохначева Ю. В.* Типы документов, индексируемых в базах данных WoS и Scopus: сходства, различия и их значение при анализе публикационной активности // *Научно-техническая информация. Серия 1: Организация и методика информационной работы*. 2023. № 1. P. 38-43. DOI 10.36535/0548-0019-2023-01-4. EDN KCNIMA.

58. *Harzing A. W.* Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison / A. W. Harzing, S. Alakangas // *Scientometrics*. 2016. Vol. 106, № 2. P. 787–804. DOI 10.1007/s11192-015-1798-9.

59. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories / A. Martin-Martin, E. Orduna-Malea, M. Thelwall, E. D. Lopez-Cozar // *Journal of Informetrics*. 2018. Vol. 12, № 4. P. 1160–1177. DOI 10.1016/j.joi.2018.09.002.

60. The difference in referencing in Web of Science, Scopus, and Google Scholar / M. S. Anker, S. Hadzibegovic, A. Lena, W. Haverkamp // *Esc Heart Failure*. 2019. Vol. 6, № 6. P. 1291–1312. DOI 10.1002/ehf2.12583.

61. *Chapman K.* An evaluation of Web of Science, Scopus and Google Scholar citations in operations management / K. Chapman, A. E. Ellinger // *International Journal of Logistics Management*. 2019. Vol. 30, № 4. P. 1039–1053. DOI 10.1108/ijlm-04-2019-0110.

62. *Лутай А. В.* Сравнение качества метаданных в БД CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection / А. В. Лутай, Е. Э. Любушко // *Национальная подписка : [сайт]*. 2023. URL: https://podpiska.rfbr.ru/storage/reports2021/2022_meta_quality.html (дата обращения: 21.04.2023).

63. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis / V. K. Singh, P. Singh, M. Karmakar [et al.] // *Scientometrics*. 2021. Vol. 126, № 6. P. 5113–5142. DOI 10.1007/s11192-021-03948-5.

64. *Thelwall M.* Dimensions: A competitor to Scopus and the Web of Science? // *Journal of Informetrics*. 2018. Vol. 12, № 2. P. 430–435. DOI 10.1016/j.joi.2018.03.006.

65. *Harzing A.-W.* Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? // *Scientometrics*. 2019. Vol. 120, № 1. P. 341–349. DOI 10.1007/s11192-019-03114-y.

66. *Orduna-Malea E.* Dimensions: re-discovering the ecosystem of scientific information / E. Orduna-Malea, E. Delgado-Lopez-Cozar // *Profesional De La Informacion*. 2018. Vol. 27, № 2. P. 420–431. DOI 10.3145/epi.2018.mar.21.

67. Is Dimensions a reliable data source of funding and funded publications? / L. Zhang, Y. X. Zheng, W. J. Zhao, Y. Huang // *Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI2021) (12–15 July 2021, Leuven, Belgium)*. Belgium : KU Leuven, 2021. P. 1573–1574.

68. *Visser M.* Large-scale comparison of bibliographic data sources: Web of Science, Scopus, Dimensions, and Crossref / M. Visser, N. J. van Eck, L. Waltman // 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.

69. *Basso I.* Data sources and their effects on the measurement of open access. Comparing Dimensions with the Web of Science / I. Basso, M. A. Simard, Z. A. Ouangre [et al.] // Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI2021) (12–15 July 2021, Leuven, Belgium). Belgium : KU Leuven, 2021. P. 93–98.

70. *Stahlschmidt S.* From indexation policies through citation networks to normalized citation impacts: Web of Science, Scopus, and Dimensions as varying resonance chambers / S. Stahlschmidt, D. Stephen // *Scientometrics*. 2022. Vol. 127, № 5. P. 2413–2431. DOI 10.1007/s11192-022-04309-6.

71. Finding citations for PubMed: a large-scale comparison between five freely available bibliographic data sources / Z. T. Liang, J. Mao, K. Lu, G. Li // *Scientometrics*. 2021. Vol. 126, № 12. P. 9519–9542. DOI 10.1007/s11192-021-04191-8.

72. *Velayos-Ortega G.* Most cited journals in coronavirus patents according to Lens.org / G. Velayos-Ortega, R. Lopez-Carreno // *Profesional De La Informacion*. 2020. Vol. 29, № 5. Art. no. e290519. DOI 10.3145/epi.2020.sep.19.

73. *Ruan Z., Jiang Y. X., Shi H. H.* Real-world clinical effectiveness of once-weekly semaglutide in patients with type 2 diabetes: a systematic literature review / Z. Ruan, Y. X. Jiang, H. H. Shi [et al.] // *Expert Review of Clinical Pharmacology*. 2023. Vol. 16, № 2. P. 161–176. DOI 10.1080/17512433.2023.2174099.

74. Sinonasal pathophysiology of SARS-CoV-2 and COVID-19: A systematic review of the current evidence / I. Gengler, J. C. Wang, M. M. Speth, A. R. Sedaghat // *Laryngoscope Investigative Otolaryngology*. 2020. Vol. 5, № 3. P. 354–359. DOI 10.1002/lio2.384.

Статья поступила в редакцию 24.04.2023.

Одобрена после рецензирования 11.05.2023. Принята к публикации 22.05.2023.

СВЕДЕНИЯ ОБ АВТОРАХ

Гуреев Вадим Николаевич *GureyevVN@ipgg.sbras.ru*

Кандидат педагогических наук, ведущий научный сотрудник лаборатории информационно-системного анализа, Государственная публичная научно-техническая библиотека Сибирского отделения РАН; доцент кафедры геофизических систем, Новосибирский государственный технический университет, Новосибирск, Россия

AuthorID РИНЦ: 663665

ORCID: 0000-0002-3460-0157

Мазов Николай Алексеевич *MazovNA@ipgg.sbras.ru*

Кандидат технических наук, ведущий научный сотрудник лаборатории информационно-системного анализа, Государственная публичная научно-техническая библиотека Сибирского отделения РАН; ведущий научный сотрудник информационно-аналитического центра, Институт нефтегазовой геологии и геофизики им. А. А. Трофимука Сибирского отделения РАН, Новосибирск, Россия

AuthorID РИНЦ: 98887

ORCID: 0000-0003-4607-1122

DOI 10.19181/sntp.2023.5.2.4

INCREASED ROLE OF OPEN BIBLIOGRAPHIC DATA IN THE CONTEXT OF RESTRICTED ACCESS TO PROPRIETARY INFORMATION SYSTEMS

Vadim N. Gureyev^{1,2}, Nikolay A. Mazov^{1,3}

¹ State Public Scientific Technological Library, Siberian Branch of the RAS, Novosibirsk, Russia

² Novosibirsk State Technical University, Novosibirsk, Russia

³ Trofimuk Institute of Petroleum Geology and Geophysics Siberian Branch of the RAS, Novosibirsk, Russia

For citation: Gureyev, V. N., Mazov, N. A. (2023). Increased Role of Open Bibliographic Data in the Context of Restricted Access to Proprietary Information Systems. *Science Management: Theory and Practice*. Vol. 5, no. 2. P. 49–76. DOI 10.19181/sntp.2023.5.2.4.

Abstract. The paper presents a review of function capabilities and coverage of sources in open bibliographic databases that can be useful in the limited access to proprietary information systems. Databases were primarily evaluated with regard to their usefulness for researchers and research libraries who solve the problems of information and patent search, bibliometric assessment of authors, promotion of papers in international information space, searching collaborators or conducting bibliometric studies. We focused on multidisciplinary databases covering wide range of international scientific literature. Based on our own experience and literature review, we concluded on possibility in principle to solve almost all information-retrieval and bibliometric tasks using current open bibliographic databases and their web-tools. Furthermore, large volumes of metadata are now regarded as a basic and non-unique feature of different databases, while analytical characteristics are taking centre stage.

Keywords: bibliographic databases, information retrieval, open access, AMiner, Dimensions, Exaly, Google Scholar, Lens, OpenAlex, Scilit, scite, Semantic Scholar, RSCI

Acknowledgement: The study was carried out under the projects of SPSTL SB RAS (122040600059-7) and IPGG SB RAS (FWZZ-2022-0028).

REFERENCES

1. Web of Science Journal Evaluation Process and Selection Criteria. *Clarivate*. 2023. URL: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/core-collection/editorial-selection-process/editorial-selection-process/> (accessed: 21.04.2023).
2. Content Policy and Selection. *Elsevier*. 2023. URL: <https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection> (accessed: 21.04.2023).
3. Vera-Baceta, M. A., Thelwall, M. and Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*. Vol. 121, no. 3. P. 1803–1813. DOI 10.1007/s11192-019-03264-z.

4. Mongeon, P. and Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*. Vol. 106, no. 1. P. 213–228. DOI 10.1007/s11192-015-1765-5.
5. Martín-Martín, A., Orduna-Malea, E. and Delgado López-Cózar, E. (2018). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics*. Vol. 116, no. 3. P. 2175–2188. DOI 10.1007/s11192-018-2820-9.
6. Wang, K., Shen, Z., Huang, C. [et al.] (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*. Vol. 1, no. 1. P. 396–413. DOI 10.1162/qss_a_00021.
7. Herzog, C., Hook, D. and Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*. Vol. 1, no. 1. P. 387–395. DOI 10.1162/qss_a_00020.
8. Penfold, R. (2020). Using the Lens database for staff publications. *Journal of the Medical Library Association*. Vol. 108, no. 2. P. 341–344. DOI 10.5195/jmla.2020.918
9. Semenov, E. V. (2023). Behind Is the Year of Vain Expectations, ahead is the Year of Hopes and Urgent Matters. *Science Management: Theory and Practice*. Vol. 5, no. 1. P. 10–13. (In Russ.).
10. Mazov, N. A. and Gureyev, V. N. (2020). *IPGGTR Proceedings of the staff of INGG SB RAS (abstract-full-text bibliography)*: Certificate of state registration of a computer program 2020621025; RU; No. 2020620872. URL: https://new.fips.ru/registers-doc-view/fips_servlet?DB=DB&DocNumber=2020621025&TypeFile=html (accessed: 21.04.2023).
11. Hendricks, G., Tkaczyk, D., Lin, J. and Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*. Vol. 1, no. 1. P. 414–427. DOI 10.1162/qss_a_00022.
12. Wang K., Shen Z., Huang C. [et al.] (2019). A Review of Microsoft Academic Services for Science of Science Studies. *Frontiers in Big Data*. Vol. 2. Art. no. 45. DOI 10.3389/fdata.2019.00045.
13. Haak, L. L., Fenner, M., Paglione, L. [et al.] (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing*. Vol. 25, no. 4. P. 259–264. DOI 10.1087/20120404.
14. Chawla, D. S. (2022). Five-year campaign breaks science’s citation paywall. *Nature*. September 13. DOI: 10.1038/d41586-022-02926-y.
15. Peroni, S. and Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*. Vol. 1, no. 1. P. 428–444. DOI 10.1162/qss_a_00023.
16. Colledge, L., De Moya-Anegón, F., Guerrero-Bote, V. [et al.] (2010). SJR and SNIP: two new journal metrics in Elsevier’s Scopus. *Serials*. Vol. 23, no. 3. P. 215–221. DOI 10.1629/23215.
17. Master Journal List (2023). *Clarivate*. URL: <https://mjl.clarivate.com/home> (accessed: 21.04.2023).
18. Web of Science Core Collection. (2023). *Clarivate*. URL: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/web-of-science/web-of-science-core-collection/> (accessed: 21.04.2023).
19. Scopus. Expertly curated abstract & citation database. (2023). *Elsevier*. URL: <https://www.elsevier.com/solutions/scopus> (accessed: 21.04.2023).
20. Google Scholar. (2023). *Google*. URL: <https://scholar.google.com/intl/ru/scholar/about.html> (accessed: 21.04.2023).
21. Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*. Vol. 118, no. 1. P. 177–214. DOI 10.1007/s11192-018-2958-5.

22. Russian Science Citation Index (2023). *eLibrary*. URL: https://www.elibrary.ru/project_risc.asp (accessed: 21.04.2023).
23. Moskaleva, O., Pisyakov, V., Sterligov, I. [et al.] (2018). Russian Index of Science Citation: Overview and review. *Scientometrics*. Vol. 116, no. 1. P. 449–462. DOI 10.1007/s11192-018-2758-y.
24. AMiner (2023). *Tsinghua University*. URL: <https://www.aminer.org/manual#aminer-manual-subTitle-research-feed> (accessed: 21.04.2023).
25. Tang, J. (2016). AMiner: Toward Understanding Big Scholar Data. In: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM'16)* (22–25 February 2016, San Francisco). P. 467.
26. Tang, J., Zhang, J., Yao, L. (at al). (2008). ArnetMiner: Extraction and Mining of Academic Social Networks. In: *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)* (14–27 August 2008, Las Vegas). Association for Computing Machinery. P. 990–998.
27. About The Lens (2023). *Cambia*. URL: <https://about.lens.org/> (accessed: 21.04.2023).
28. Jefferson, O. A., Koellhofer, D., Warren, B. and Jefferson, R. (2019). The Lens MetaRecord and LensID: An open identifier system for aggregated metadata and versioning of knowledge artefacts. *Researchgate*. URL: <https://www.researchgate.net/publication/337496596> (accessed: 15.05.2023). DOI 10.31229/osf.io/t56yh.
29. Scilit brochure (2023). *MDPI*. URL: <https://app.scilit.net/scilit-brochure.pdf> (accessed: 21.04.2023).
30. About Semantic Scholar (2023). *Allen Institute for AI*. URL: <https://www.semanticscholar.org/about> (accessed: 05.04.2023).
31. Kinney, R., Anastasiades, C., Authur, R. [et al.] (2023). The Semantic Scholar Open Data Platform. P. 1–8. DOI 10.48550/arXiv.2301.10140.
32. Ammar, W., Groeneveld, D., Bhagavatula, C. [et al]. (2018). Construction of the Literature Graph in Semantic Scholar. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2–4 June 2018, New Orleans). Vol. 3. P. 84–91. DOI 10.18653/v1/n18-3011.
33. The data in Dimensions – from idea to impact (2023). *Digital Science*. URL: <https://www.dimensions.ai/dimensions-data/> (accessed: 05.04.2023).
34. Nicholson, J. M., Mordaunt, M., Lopez, P. [et al.] (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*. Vol. 2, no. 3. P. 882–898. DOI 10.1162/qss_a_00146.
35. OpenAlex API documentation (2023). *OurResearch*. URL: <https://docs.openalex.org/api-entities/works> (accessed: 21.04.2023).
36. Priem, J., Piwowar, H. and Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. In: *26th International Conference on Science and Technology Indicators “From Global Indicators to Local Applications” (STI 2022)* (7–9 September, Granada, Spain). P. 1–5.
37. Exaly Statistics (2023). *Exaly*. URL: <https://exaly.com/statistics.html> (accessed: 21.04.2023).
38. Mazov, N. A. and Gureev, V. N. (2014). The role of unique identifiers in bibliographic information systems. *Scientific and Technical Information Processing*. Vol. 41, no. 3. P. 206–210. DOI 10.3103/S0147688214030101.
39. Mazov, N. A. and Gureyev, V. N. (2022). Publication Databases of Research Organizations as a Tool for Information Studies. *Scientific and Technical Information Processing*. Vol. 49, no. 2. P. 108–118. DOI 10.3103/s0147688222020071.
40. Mazov, N. A. and Gureyev, V. N. (2016). Bibliograficheskaya baza dannykh trudov sotrudnikov organizatsii: tseli, funktsii, sfera ispol'zovaniya v naukometrii [Bibliographic

database of the works of the organization's employees: goals, functions, scope of use in scientometrics]. *Vestnik Dal'nevostochnoi gosudarstvennoi nauchnoi biblioteki*. No. 2. P. 84–87. (In Russ.).

41. Prancutè, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*. Vol. 9, no 1. Art. no. 12. DOI 10.3390/publications9010012.

42. Delgado López-Cózar, E., Robinson-García, N. and Torres-Salinas, D. (2014). The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*. Vol. 65, no. 3. P. 446–454. DOI 10.1002/asi.23056.

43. Tikhonova, E. V. and Kirillova, O. V. (2022). Citation culture: Citing authors behaviour vs trust in research results. *Science Editor and Publisher*. Vol. 7, no. 2. P. 166–181. DOI 10.24069/SEP-22-58 (In Russ.).

44. Bobrov, L. K. (2022). Reliability of references to scientific publications: where myths and inaccuracies originate. *Scientific and Technical Libraries*. No. 5. P. 47–65. DOI 10.33186/1027-3689-2022-5-47-65 (In Russ.).

45. Valenzuela, M., Ha, V. and Etzioni, O. (2015). Identifying meaningful citations. *29th AAAI Conference on Artificial Intelligence, AAAI 2015 (25–30 January 2015, Austin, United States)*. AI Access Foundation. P. 21–26.

46. New WoS April 29 Release Notes: Enriched cited references, Export to Publons, and more (2022). *Clarivate*. URL: <https://clarivate.com/webofsciencegroup/release-notes/wos/new-wos-april-29-release-notes/> (accessed: 21.04.2023).

47. Dardas, L. A., Sallam, M., Woodward, A. [et al.] (2023). Evaluating Research Impact Based on Semantic Scholar Highly Influential Citations, Total Citations, and Altmetric Attention Scores: The Quest for Refined Measures Remains Illusive. *Publications*. Vol. 11, no. 1. Art. no. 5. DOI 10.3390/publications11010005.

48. Hook, D. W., Porter, S. J. and Herzog, C. (2018). Dimensions: Building Context for Search and Evaluation. *Frontiers in Research Metrics and Analytics*. Vol. 3. Art. no. 23. DOI 10.3389/frma.2018.00023.

49. Web of Science Release Notes, April 13 2023: Automatic updates to claimed profiles... (2023). *Clarivate*. URL: <https://clarivate.com/webofsciencegroup/release-notes/wos/web-of-science-release-notes-april-13-2023-2/> (accessed: 21.04.2023).

50. Ramos, M. A., Melo, J. G. and Albuquerque, U. P. (2012). Citation behavior in popular scientific papers: What is behind obscure citations? The case of ethnobotany. *Scientometrics*. Vol. 92, no. 3. P. 711–719. DOI 10.1007/s11192-012-0662-4.

51. Derevyanko, A. P. and Kholyushkin, Yu. P. (1994). Problema kachestvennogo analiza arkheologicheskikh publikatsii [The problem of qualitative analysis of archaeological publications]. In: *Metodologiya i metodika arkheologicheskikh rekonstruktsii: Sbornik nauchnykh trudov*. Novosibirsk: SB RAS. P. 24–32. (In Russ.).

52. Simkin, M. V. and Roychowdhury, V. P. (2003). Read before you cite! *Complex Systems*. Vol. 14, no. 3. P. 262–274. DOI 10.25088/ComplexSystems.14.3.269.

53. Lazarev V. S. (2018). *Nobel class citedness level and the notions that designate characteristics and properties of cited scientific documents*. Tambov; Moscow; St.-Petersburg; Baku; Vienna; Gamburg; Stockholm; Buake: Mezhdunarodnyi Informatsionnyi Nobelevskii Tsentri (MINTs). 70 p. (In Russ.).

54. Pislyakov, V. V. (2022). Self-citation and its impact on scientific workflow assessment: The review of publications. Part I. *Scientific and Technical Libraries*. No. 2. P. 49–70. DOI 10.33186/1027-3689-2022-2-49-70 (In Russ.).

55. Pislyakov, V. V. (2022). Self-citation and its impact on research evaluation: Literature review. Part II. *Scientific and Technical Libraries*. No. 3. P. 85–104. DOI 10.33186/1027-3689-2022-3-85-104 (In Russ.).

56. González-Pereira, B., Guerrero-Bote, V. P., Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics*. Vol. 4, no. 3. P. 379–391. DOI 10.1016/j.joi.2010.03.002.
57. Mokhnacheva, Ju. V. (2023). Document types in WoS and Scopus: similarities, differences and their significance in the analysis of publication activity. *Nauchno-tehnicheskaya informatsiya. Seriya 1: Organizatsiya i metodika informatsionnoi raboty*. No. 1. P. 38–43. DOI 10.36535/0548-0019-2023-01-4. (In Russ.).
58. Harzing, A. W. and Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*. Vol. 106, no. 2. P. 787–804. DOI 10.1007/s11192-015-1798-9.
59. Martin-Martin, A., Orduna-Malea, E., Thelwall, M. and Lopez-Cozar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*. Vol. 12, no. 4. P. 1160–1177. DOI 10.1016/j.joi.2018.09.002.
60. Anker, M. S., Hadzibegovic, S., Lena, A., Haverkamp, W. (2019). The difference in referencing in Web of Science, Scopus, and Google Scholar. *Esc Heart Failure*. Vol. 6, no. 6. P. 1291–1312. DOI 10.1002/ehf2.12583.
61. Chapman, K. and Ellinger, A. E. (2019). An evaluation of Web of Science, Scopus and Google Scholar citations in operations management. *International Journal of Logistics Management*. Vol. 30, no. 4. P. 1039–1053. DOI 10.1108/ijlm-04-2019-0110.
62. Lutai, A. V. and Lyubushko, E. E. (2023). Sravnenie kachestva metadannykh v BD CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection [Comparison of metadata characteristics in BD CrossRef, Lens, OpenAlex, Scopus, Semantic Scholar, Web of Science Core Collection]. *Natsionalnaya podpiska*. URL: https://podpiska.rfbr.ru/storage/reports2021/2022_meta_quality.html (accessed: 21.04.2023).
63. Singh, V. K., Singh, P., Karmakar, M. [et al.]. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*. Vol. 126, no. 6. P. 5113–5142. DOI 10.1007/s11192-021-03948-5.
64. Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*. Vol. 12, no. 2. P. 430–435. DOI 10.1016/j.joi.2018.03.006.
65. Harzing, A.-W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*. Vol. 120, no. 1. P. 341–349. DOI 10.1007/s11192-019-03114-y.
66. Orduna-Malea, E. and Delgado-Lopez-Cozar, E. (2018). Dimensions: re-discovering the ecosystem of scientific information. *Profesional De La Informacion*. Vol. 27, no. 2. P. 420–431. DOI 10.3145/epi.2018.mar.21.
67. Zhang, L., Zheng, Y. X., Zhao, W. J. and Huang, Y. (2021). Is Dimensions a reliable data source of funding and funded publications? In: *Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI2021)* (12–15 July 2021, Leuven, Belgium). Belgium: KU Leuven. P. 1573–1574.
68. Visser, M., van Eck, N. J. and Waltman, L. (2019). Large-scale comparison of bibliographic data sources: Web of Science, Scopus, Dimensions, and Crossref. In: *Proceedings of the 17th International Conference on Scientometrics & Informetrics (ISSI2019)* (2–5 September 2019, Rome, Italy). Vol. 2. Rome: Edizioni Efesto. P. 2358–2369.
69. Basso, I., Simard, M. A., Ouangre, Z. A. [et al.]. (2021). Data sources and their effects on the measurement of open access. Comparing Dimensions with the Web of Science. In: *Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI2021)* (12–15 July 2021, Leuven, Belgium). Belgium: KU Leuven. P. 93–98.
70. Stahlschmidt, S. and Stephen, D. (2022). From indexation policies through citation networks to normalized citation impacts: Web of Science, Scopus, and Dimensions as

varying resonance chambers. *Scientometrics*. Vol. 127, no. 5. P. 2413–2431. DOI 10.1007/s11192-022-04309-6.

71. Liang, Z. T., Mao, J., Lu, K. and Li, G. (2021). Finding citations for PubMed: a large-scale comparison between five freely available bibliographic data sources. *Scientometrics*. Vol. 126, no. 12. P. 9519–9542. DOI 10.1007/s11192-021-04191-8.

72. Velayos-Ortega, G. and Lopez-Carreno, R. (2020). Most cited journals in coronavirus patents according to Lens.org. *Profesional De La Informacion*. Vol. 29, no. 5. Art. no. e290519. DOI 10.3145/epi.2020.sep.19.

73. Ruan, Z., Jiang, Y. X., Shi, H. H. [et al.]. (2023). Real-world clinical effectiveness of once-weekly semaglutide in patients with type 2 diabetes: a systematic literature review. *Expert Review of Clinical Pharmacology*. Vol. 16, no. 2. P. 161–176. DOI 10.1080/17512433.2023.2174099.

74. Gengler, I., Wang, J. C., Speth, M. M. and Sedaghat, A. R. (2020). Sinonasal pathophysiology of SARS-CoV-2 and COVID-19: A systematic review of the current evidence. *Laryngoscope Investigative Otolaryngology*. Vol. 5, no. 3. P. 354–359. DOI 10.1002/lio2.384.

The article was submitted on 24.04.2023.

Approved after reviewing 11.05.2023. Accepted for publication 22.05.2023.

INFORMATION ABOUT THE AUTHORS

Gureyev Vadim *GureyevVN@ipgg.sbras.ru*

Candidate of Pedagogics, Leading Researcher, Laboratory of Information and System Analysis, State Public Scientific Technological Library, Siberian Branch of the RAS; Associate Professor of the Department of Geophysical Systems, Novosibirsk State Technical University, Novosibirsk, Russia

ORCID: 0000-0002-3460-0157

AuthorID RSCI: 663665

Mazov Nikolay *MazovNA@ipgg.sbras.ru*

Candidate of Technical Sciences, Leading Researcher, Laboratory of Information and System Analysis, State Public Scientific Technological Library, Siberian Branch of the RAS; leading researcher of the information and analytical center, Trofimuk Institute of Petroleum Geology and Geophysics, Siberian Branch of the RAS, Novosibirsk, Russia

ORCID: 0000-0003-4607-1122

AuthorID RSCI: 98887