

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 5

Москва 2022

ОБЩИЙ РАЗДЕЛ

УДК 004.6–022.59:001:311.311

Е.В. Мельникова

Технология больших данных в наборе методов и средств научного исследования в современной наукометрии*

Рассмотрены проблемы применения технологии больших данных (Big Data) в современном наукометрическом анализе. Важность и актуальность этих проблем обусловлена способностью больших данных существенно повысить эффективность наукометрических исследований путем глубокого анализа мегаобъемов разнородных данных и выявления на этой основе новых смысловых взаимосвязей и закономерностей. Раскрыто основное содержание технологии больших данных; представлен перечень требований к данным, которые позволяют их относить к категории больших данных; отмечено значение этой технологии как передового средства научного исследования в наукометрии. Дана подробная характеристика методов научного исследования. Проанализированы особенности библиометрических, альтметрических, вебометрических и вероятностно-статистических мето-

* Статья подготовлена в рамках исследования по теме FFFU-2021-0002 Государственного задания ВИНТИ РАН и при поддержке Российского фонда фундаментальных исследований – проект № 20-07-00014.

дов. Подчеркнуто важное место технологии больших данных в современном наборе методов и средств научного исследования в наукометрии.

Ключевые слова: технология больших данных, Big Data, наукометрия, альтметрия, вебметрия, киберметрия, наукометрические показатели, методы научного исследования, неструктурированные данные

DOI: 10.36535/0548-0019-2022-05-1

ВВЕДЕНИЕ

Технология больших данных (*Big Data*), разработка которой относится к началу XXI в., применяется для решения конкретных прикладных или теоретических задач в области обработки и анализа огромных объемов информации, накопленной в мире, а также новой информации, формирующейся постоянно, на каждом новом временном отрезке развития человечества и всей нашей планеты в целом. Объемы данных растут экспоненциально. Если в 2020 г. глобальный объем накопленных данных составил порядка 60 зеттабайт¹, то в 2025 г., как ожидается, он уже достигнет уровня в 165-170 зеттабайт [1].

Возможности технологии больших данных позволяют справляться с обработкой растущих объемов информации и оптимизировать различные сферы социальной жизни и отрасли экономики, включая энергетику, логистику, медицину, государственное управление, сферу безопасности, банковскую и биржевую деятельность, область телекоммуникаций, метеорологию и другие сферы. В последние годы эта технология все более широко применяется и в сфере науки, в том числе – в такой научной дисциплине, как наукометрия.

СОВРЕМЕННАЯ НАУКОМЕТРИЯ: МЕТОДЫ И СРЕДСТВА НАУЧНОГО ИССЛЕДОВАНИЯ

Наукометрия – это научная дисциплина, которая на основе метрических исследований изучает «развитие науки как информационного процесса» [2], проявляющего себя через систему научных коммуникаций. Наукометрия использует совокупность методов, основанных на разработке и применении особых числовых показателей (метрик²), для решения задач по оценке результативности научной деятельности, исследования процесса развития науки, ее структуры, тенденций и перспектив развития. Важно отметить, что при анализе результативности науки наукометрия может определять как количественные, так и качественные показатели, характеризующие результаты научной деятельности [3]. На международном уровне наработки в области методологии оценки научной деятельности были изложены в Лейденском манифесте для наукометрии [4], принятом в 2014 г.

Рассматривая средства научного исследования в наукометрии, необходимо выделить несколько их категорий: материальные, информационные, математические, логические и языковые. В условиях развития

процессов информатизации общества, цифровизации инфо-коммуникационной среды значительно возрастает роль *информационных средств* научного исследования. Вычислительная техника, информационные технологии, системы телекоммуникаций, широко внедряемые в различные сферы общественной жизни, включая науку, коренным образом преобразовывают научную деятельность и становятся для науки средствами познания, осмысления окружающей действительности, т. е. средствами научного исследования.

Значимое место в категории информационных средств научного исследования занимают автоматизированные информационные системы и технологии, среди которых различают традиционные и новые, перспективные технологии. В наукометрии в текущий период к наиболее передовым следует отнести технологию больших данных, которая в составе соответствующих информационных систем будет рассмотрена ниже как важное средство научного исследования.

МЕТОДЫ ИССЛЕДОВАНИЯ

В зависимости от области знания, где методы научного исследования изначально разработаны, в наукометрии они делятся на: библиометрические, альтметрические, вебметрические, вероятностно-статистические и метод профессиональных экспертных оценок. Следует сразу подчеркнуть, что всестороннюю объективную оценку результатов научной деятельности может дать только профессиональная экспертиза; она способна в полной мере учесть содержательные аспекты научной работы. В отличие от метода экспертных оценок, все остальные методы основаны на формальных количественных показателях и служат важными инструментами поддержки принятия решений экспертами [5].

Библиометрические методы. Для наукометрии важно, что в рамках библиометрии разработан набор методов для изучения текстов, документальных потоков и массивов информации, которые наукометрия может использовать для решения своих задач, включая оценку результативности науки, определение «веса» ученого в научном сообществе и др. Среди основных методов библиометрии целесообразно выделить анализ пристатейных ссылок и библиографического сочетания документов [6].

Метод пристатейных ссылок (также называют методом научного цитирования) базируется на идее Ю.Гарфилда об использовании ссылок (количества ссылок) на статьи в научных журналах как средства изучения структуры науки и развития ее направлений, равно как и средства информационного поиска. Метод основан на таком показателе, как индекс

¹ Один зеттабайт (ZB) равен 10^{21} байт.

² «Метрика» в переводе с греческого означает «измерение».

научного цитирования, разработанный Ю.Гарфилдом и запущенный в научный оборот в 1964 г. – *Science Citation Index (SCI)* [7]. Сейчас это общепринятый показатель значимости трудов ученого; он представляет собой число ссылок на научные публикации ученого. На базе индекса научного цитирования разработано большое количество производных библиометрических показателей, которые широко используются в современной наукометрии. К ним относятся: индекс Хирша³ [8], g-индекс (вычисляется как корень из суммарного цитирования работ ученого или сотрудников организации) [9], i-индекс (позволяет выделить ядро наиболее научно активных и востребованных авторов, имеющих наиболее высокий индекс Хирша) [10], импакт-фактор (для научных журналов) и некоторые другие показатели.

Метод библиографического сочетания документов (или библиографической связанности)⁴ предусматривает поиск связанных по смыслу документов, авторы которых ссылаются на одни и те же работы [11]. Числом совпадающих ссылок измеряется степень смысловой и тематической связанности документов, что важно для пользователей, осуществляющих поиск публикаций по интересующей их тематике. Библиографическое сочетание образуется между цитируемыми документами. Если два или более документов имеют общие ссылки, то они библиографически связаны. Концепция библиографической связанности помогает исследователям в ретроспективном поиске документов в базах данных.

Альтметрические методы. Альтметрия предлагает метрики, альтернативные традиционным (библиометрическим) метрикам/индексам цитирования. Альтметрики, или альтернативные показатели, формируются на основе обработки информации из социальных сетей, традиционных СМИ, правительственных интернет-порталов, тематических платформ, блогов, профессиональных сетевых сообществ – научной и ненаучной направленности. Альтметрия оценивает результаты научной деятельности не по количеству ссылок на научные статьи, как библиометрия, а по присутствию, упоминанию и использованию статей, имени ученых и их идей в вышеперечисленных информационных ресурсах, по уровню общественного интереса к ним и масштабам реального использования [12]. Альтметрия анализирует их *общественный* вес, предоставляя наукометрии дополнительный материал для оценки результатов деятельности отдельных ученых, реже – коллективов, исследовательских организаций.

³ Индекс Хирша ученого равен h , если ученый опубликовал h статей, на каждую из которых сослались как минимум h раз.

⁴ Метод библиографической связанности документов (Bibliographic Coupling) был введен американским ученым Кеслером в 1963 г. в работе «Библиографическая связанность между научными документами» (Kessler M.M. Bibliographic coupling between scientific papers // American Documentation. – Wiley-Blackwell. – 1963. – Vol. 14, Issue 1. – P. 10-25). Ю.Гарфилд развил и более глубоко проработал эту концепцию применительно к научным публикациям.

Для наукометрии важно, что альтметрики можно использовать как вспомогательный инструмент для оценки уровня внимания участников информационных коммуникаций к ученым, их идеям и публикациям, для определения степени их воздействия на социум, а также для характеристики степени влияния ученых в сообществе. Альтметрики, кроме того, можно применять к научным журналам, книгам, научным конференциям, презентациями т. д. Альтметрики эффективно работают с новыми формами представления научных результатов: теперь это не только статьи в академических журналах, но и видео ролики, записи в блогах, аудиоданные, фотографии и пр. Их обсуждение выходит за границы академического сообщества; к обсуждению подключаются представители внеакадемического социума [13]. Традиционные методы оценки не умеют работать с таким многообразием форм и каналов передачи информации, альтметрические – умеют.

Альтметрики формируют достаточно многочисленную группу и подразделяются на несколько видов: просмотры (количество просмотров статей, информации об ученом, научной идее) и скачивания – отражают уровень внимания к результатам научного труда; обсуждения/комментарии (в блогах и на форумах, упоминание в новостях, репосты в социальных сетях) – характеризуют потенциальное влияние ученого/научной идеи в социальной среде; сохранения/закладки – свидетельствуют об общественном интересе и степени воздействия ученого/идеи на пользователей; цитирования (например, ссылка на научную публикацию в экспертных заключениях, в правительственных документах) – отражают воздействие ученого/научной статьи на участников коммуникаций, и некоторые другие.

Следует отметить, что существуют определенные ограничения на использование альтметрических показателей в рамках наукометрии. Так, у части научного сообщества вызывает сомнение достоверность данных альтметрии. Сомнение вызвано тем, что сбор значительного объема информации для дальнейшей обработки альтметрии производит на информационных ресурсах ненаучной направленности. Еще одно ограничение связано с тем, что альтметрики могут отразить степень влияния, например, статьи на участников коммуникаций, но при этом не могут ответить на вопрос, является ли отношение участников к данной статье позитивным или негативным. Эти и некоторые другие особенности альтметрик обуславливают необходимость осторожного отношения к применению альтметрических методов для решения задач наукометрии.

Вебометрические методы. Вебометрия исследует количественные аспекты конструирования и использования информационных ресурсов, структур и технологий в пространстве современного веба – на интернет-сайтах, порталах, интернет-форумах и т.д. Одной из разновидностей веб-пространства является интернет вещей, в котором концентрируются данные датчиков, собирающих и передающих информацию от различных устройств, создавая «вещевые сети». Вебометрия формирует свои методы на базе библиометрических, которые она использует в сетевом режиме. Суть вебометрических методов заключается в

том, что собираемая информация обрабатывается с использованием математических, статистических процедур, включая определение среднеарифметического по каждому вебметрическому индикатору. Получаемые графы, построенные на основе взаимного цитирования, анализируются на базе теории графов и методов анализа социальных сетей *SNA (Social network analysis)* [14], который включает исследование плотности и интенсивности возникающих в ходе социальной коммуникации связей/сетей.

Значимость вебметрических методов для наукометрии заключается в том, что результаты анализа плотности и интенсивности сетевых связей, относящихся к конкретному информационному ресурсу, могут дать характеристику степени воздействия ученого или научной статьи на коммуникационный социум и свидетельствовать об их значимости для участников вэб-коммуникаций. Необходимо также отметить, что особенности анализа данных в сетевом пространстве позволяют решать исследовательские задачи наукометрии с использованием новых, отличных от традиционных средств научного исследования, включая технологию больших данных.

Общее, что объединяет рассмотренные метрические методы научного исследования, состоит в том, что в качестве базы они используют библиометрические показатели и уже на их основе в каждой из сфер разрабатываются свои метрики, которые в наибольшей степени соответствуют потребностям именно этой сферы: в библиометрии – показатели, характеризующие движение потоков документов, в альтметрии – движение преимущественно недокументальных потоков, в вебметрии – смешанных информационных потоков в интернет-пространстве. Кроме того, в киберметрии, например, исследуются информационные потоки в цифровой среде. Помимо этого в научном сообществе в последние годы рассматриваются идеи о выделении в самостоятельную область сетеметрии, занимающейся потоками информации в сетевой среде, медиаметрии [15], работающей с информационными потоками в средствах массовой информации, и некоторых других метрических областей. В каждой из них может осуществляться движение научной информации, которая там обрабатывается существующими в этой среде методами. Поэтому все перечисленные области и их методы исследования представляют практический интерес для наукометрии в решении стоящих перед ней задач.

Вероятностно-статистические методы. Базируются на эмпирических закономерностях, нашедших свое отражение в законах Брэдфорда, Лотки, Ципфа и некоторых других ученых. Законы в равной мере относятся к различным сферам системы научных коммуникаций. Сущность закономерности С. Брэдфорда [16] заключается в следующем. Если журналы расположить в порядке убывания количества помещенных в них статей по определенной теме и полученный список разделить на три зоны с одинаковой численностью статей по этой теме, то количество наименований журналов в зонах растет в геометрической прогрессии (например, 10:100:1000) [13].

Подобная закономерность имеет место и в других сферах системы научных коммуникаций. Так, А. Лотка [17] выявил аналогичный характер распределения ученых по публикационной активности, а Дж. Ципф [18] – распределения слов в тексте по частоте их употребления⁵. В 60-е годы XX в. был установлен примерный этим вероятностно-статистическим закономерностям феномен масштабной инвариантности, т.е. свойство сохранять форму уравнений, которые их описывают, при произвольных изменениях объемов информационных массивов и потоков. На основе этого общего свойства – закономерности были объединены в рамках наукометрии в группу вероятностно-статистических методов.

Таким образом, с учетом вышеизложенного можно констатировать, что широкий спектр методов научного исследования в современной наукометрии позволяет проводить мониторинг развития науки, используя все многообразие форм представления информации на ресурсах различных типов, и отражать этот процесс с различных ракурсов, а также разрабатывать и применять оценочные метрические показатели, которые характеризуют результаты научной деятельности и служат инструментом поддержки экспертных групп в оценке результативности науки.

ТЕХНОЛОГИЯ BIG DATA КАК ПЕРЕДОВОЕ СРЕДСТВО НАУЧНОГО ИССЛЕДОВАНИЯ В НАУКОМЕТРИИ

Общая характеристика технологии больших данных

Технология *Big Data* включает особые подходы и методы *глубинного поиска, обработки, хранения и анализа данных*. Глубинный поиск и анализ данных объединяются в понятие *Data Mining*. Функционирование этой технологии [19] осуществляется на основе особо крупных и быстро растущих цифровых баз данных или виртуальных массивов, содержащих структурированные и неструктурированные данные. Традиционные решения перестают работать при высоких значениях объема и скорости поступления данных. Способность того или иного технологического приложения обрабатывать большие массивы данных, поступающих на высоких скоростях, из разнообразных источников и в различных форматах является главным критерием отнесения приложения к технологии больших данных.

Преимущества технологии *Big Data* заключаются в том, что она позволяет раскрывать смысловой потенциал мегамассивов данных за счет поиска ценных закономерностей и фактов путем объединения и глубинного анализа больших объемов данных, которые на первый взгляд не связаны между собой по смыслу. Человеческий мозг не может обнаружить такие закономерности, какие выявляет мощный компьютер в комплексе с технологией больших данных, находя совершенно неожиданные смысловые взаимосвязи.

⁵ Если все слова достаточно длинного текста упорядочить по частотности их использования, то частотность n -го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру n .

Следует отметить, что для использования больших данных в наукометрии есть некоторые ограничения. Главные из них – это особенности физического доступа к большим данным, так как такими данными зачастую владеют крупные коммерческие компании, для которых запросы исследователей не являются приоритетными. Более простой путь состоит в использовании данных, доступ к которым имеет меньше ограничений: это большие данные из открытых общественных коммуникаций, включая записи научных дискуссий на социальных порталах, обсуждения в соцсетях отдельных научных идей и ученых, комментарии на интернет-форумах, контент веб-страниц, научные статьи, опубликованные в цифровом формате, оцифрованные архивные документы и т.д.

Формула «Шесть V». Для эффективного применения технологии *Big Data* данные должны обладать рядом качеств в соответствии с формулой «Шесть V». К этим качествам относятся [20]: неоднородность и многообразие данных (*Variety*); общий объем данных, поступающих в обработку (*Volume*); скорость прироста данных (*Velocity*); достоверность данных, обеспечивающая точность результатов их обработки (*Veracity*); изменчивость данных, предполагающая многовариантность их интерпретации в зависимости от контекста (*Variability*); ценность данных (*Value*) – общественная полезность результатов обработки данных, которую технология *Big Data* обеспечивает для пользователей.

Необходимо уточнить, что под неоднородностью и многообразием данных подразумевается, что данные поступают в разных форматах, из различных источников (внутренних и внешних) и разной степени структурированности. Многие наукометрические задачи требуют совместной обработки данных различных форматов и степени структурированности. Это как раз позволяет осуществлять технологии больших данных. По мере роста многообразия данных инструменты *Big Data* становятся все более эффективными.

Структурированные и неструктурированные данные. Это две основные группы данных, которыми может быть в целом представлена вся генерируемая информация. Структурированные данные (СД) составляют 20% от общего объема информации, неструктурированные данные (НД) – 80%. Наукометрия имеет дело с обеими группами данных, у каждой из которых есть свои отличительные черты.

Структурированные данные – это хорошо организованные и точно отформатированные данные, которые в виде букв/текста и чисел хорошо вписываются в связанные строки и столбцы таблиц, подобных, например, файлам *Excel*, *Google Sheets*. СД часто называют количественными данными; это означает, что их объективный и заранее определенный характер позволяет выражать данные в числах, легко их подсчитывать и измерять. Структурированные данные существуют и хранятся в виде таблиц в формате реляционных баз данных; для их хранения не требуется много места. СД легче поддаются автоматизированной обработке, чем неструктурированные.

Неструктурированные данные – это данные, которые не имеют заранее определенной структуры и хранятся в своих собственных, неструктурированных

(исходных) форматах. Существует большое разнообразие форматов неструктурированных данных, что дает наукометрии возможность определять различные характеристики объектов исследования, оценивать их с разных ракурсов. Примерами таких данных являются: текстовые файлы, например, документы в форматах *Word*, *PDF*; переписка в электронной почте, сообщения в социальных сетях, изображения, видео, аудио файлы, данные датчиков интернета вещей, собирающих и передающих информацию от различных устройств, формируя «вещевые сети», и т.д. Неструктурированные данные также называются качественными данными: это означает, что они имеют субъективный и интерпретирующий характер, их можно разделить на категории в зависимости от характеристик и свойств. В связи с неструктурированностью эти данные не могут быть обработаны и проанализированы с помощью традиционных методов и инструментов. Наиболее распространенный формат существования неструктурированных данных – в рамках нереляционных баз данных. Для хранения таких данных требуется много места; они обычно хранятся в «озерах» данных, репозиториях хранения в необработанных форматах.

Возможности технологии *Big Data* позволяют производить глубокий поиск, обработку и анализ необходимых для наукометрии данных во всем их многообразии – структурированных и неструктурированных данных, данных в различных форматах и из различных источников, перечень которых определяется методологическими потребностями наукометрии.

Технология больших данных в когнитивных и аналитических системах

Технология *Big Data* эффективно применяется в когнитивных⁶ и аналитических информационных системах. Обработка структурированной и неструктурированной информации в них и выдача результатов происходит на скоростях, которые намного выше, чем это может делать человек. Объемы данных, с которыми работают системы, также значительно превышают возможности естественного интеллекта. При этом когнитивные системы могут обеспечивать достаточно высокий уровень точности ответов на вопросы, которые пользователи излагают на естественном языке.

Когнитивные системы – это информационные системы, использующие инновационные методы обработки данных, сходные с мыслительными процессами человека. Такие системы обладают способностью автоматизированного самообучения (машинного обучения [21]). Используя технологию *Big Data*, когнитивные системы производят анализ особо крупных и динамически растущих объемов неоднородных данных и выявляют определенные закономерности, которые не мог получить человек/ученый на основе возможностей только естественного интеллекта или на основе применения традиционных программных продуктов об-

⁶ «Когнитивный» от лат. *Cognitio*; означает «обладающий способностью познания, осмысления».

работки данных [22]. Зачастую с помощью технологии больших данных могут выявляться совершенно неожиданные закономерности или характеристики исследуемых объектов, что позволяет находить принципиально новые, более рациональные, менее затратные варианты решения существующих задач.

Аналитические информационные системы – это особый класс компьютерных систем, предназначенных для аналитической обработки и сохранения данных. Системы объединяют информацию, извлекаемую из внутренних баз данных организации и из внешних источников, анализируют ее и хранят как единое целое.

Примерами применения технологии больших данных в аналитических информационных системах являются мировые индексы цитирования, точнее – их аналитические надстройки [13]. В системе индексации и цитирования *Web of Science* функционирует аналитическая надстройка *InCites*, в системе *Scopus* – *SciVal*. В *InCites* сравниваются сводные библиометрические показатели стран и организаций за разные промежутки времени и по разным областям знания. В *SciVal* выполняется кластеризация научных публикаций и графическое представление кластеров в виде «Колеса науки». Надстройки делают аналитические вычисления для поддержки экспертных оценок научных результатов, а также для выявления тенденций и перспектив развития науки.

Аналитическую надстройку, сходную по своим базовым задачам с выше приведенными, имеет российская система индексации и цитирования РИНЦ [23]. Ее аналитическая надстройка *ScienceIndex* позволяет: 1) оценивать результативность научных организаций, отдельных ученых, определять импакт-фактор научных журналов и т.д. и 2) получать общее представление об отраслевом и региональном распределении отечественной науки.

Технология *Big Data* в когнитивных и аналитических системах достаточно эффективно обеспечивает решение задач наукометрического анализа, обрабатывая большие объемы разнородных данных на высоких скоростях и выявляя скрытые смысловые взаимосвязи и закономерности.

Практическое применение технологии больших данных в наукометрическом анализе

Характерной особенностью наукометрических исследований на основе *Big Data* является их междисциплинарность: они выполняются учеными из разных дисциплин, преимущественно – из компьютерных и инженерных наук, медицины, естественных наук (включая физику, химию, биологию и некоторые другие области).

Практическим примером может служить исследование индийских ученых *Keshav Singh* и *Sandeep Kumar*, опубликованное в 2021 г. [24]. В работе представлена методика наукометрического исследования на основе технологии больших данных для систем индексации и цитирования, библиотечных коллекций, научных репозитариев, оцифрованных архивов, баз данных особо крупных размеров. Ученые исполь-

зуют библиометрические данные за последнее десятилетие, полученные из системы *Scopus* в CSV-файлах (файлах в формате изображения). Рассматриваются библиометрические особенности документов, проиндексированных системой *Scopus*. Для выявления скрытой информации из загруженного набора данных авторы исследования анализируют плотность библиометрических сетей и интенсивность связей между публикациями, число цитирований и перекрестных цитирований, социализирований и самоцитирований. Ученые делают акцент на сборе и анализе данных о темпах роста числа публикаций, их тематических категориях, географическом распределении, особенностях цитирования. С помощью инструмента *VOSviewer* в исследовании проводится оценка частоты использования ключевых слов. На основе автоматизированного глубинного анализа выявляются высоко цитируемые публикации, наиболее интересные для научного сообщества авторы, авторитетные журналы, влиятельные институты и исследовательские коллаборации. Анализ всего объема данных позволяет определить наиболее «горячие» темы научных исследований в заданный период, выявить новые тенденции в развитии изучаемого исследовательского ландшафта, обозначить наиболее востребованные и перспективные направления будущих исследований и в упреждающем порядке оказать финансовую и организационную поддержку соответствующим областям фундаментальных исследований.

ЗАКЛЮЧЕНИЕ

Таким образом, можно констатировать, что разнообразие методов исследования в современной наукометрии, включая библиометрические, вебометрические и другие, а также неоднородность и многообразие данных, которые обрабатываются в рамках этих методов и обладают перечнем характеристик, соответствующих требованиям технологии больших данных, формируют благоприятную основу для применения этой технологии в наукометрических исследованиях. Такая естественная корреляция между существующими в наукометрии условиями, с одной стороны, и требованиями технологии больших данных – с другой, позволяют сделать позитивный прогноз о возможностях открытия на базе *Big Data* новых значимых наукометрических закономерностей, которые позволят значительно повысить качество исследований в данной научной сфере и увеличить точность наукометрических оценок.

СПИСОК ЛИТЕРАТУРЫ

1. Доклады “DataAge 2020” и “DataAge 2025” аналитической компании International Data Corporation (США) / Корпоративный сайт www.idc.com (дата обращения 23.02.2022).
2. Налимов В.В., Мульченко З.М. Наукометрия: изучение развития науки как информационного процесса. – Москва: Наука. – 1969. – 192 с.
3. Гиляревский Р.С., Мельникова Е.В. Отказ от приоритетности международных индексов научного цитирования при оценке труда ученых в

- Китае // Научно-техническая информация. Сер. 1. – 2020. – № 9. – С.19-24; Gilyarevski R.S., Melnikova E.V. Rejection of the Priority of International Science Citation Indexes in the Evaluation of Results of Scientific Activity in China // Scientific and Technical Information Processing. – Springer. – 2020. – Vol. 47, № 3. – P. 194-199.
4. Hicks D., Wouters P., Waltman L., Rijcke S., Rafols I. Bibliometrics: The Leiden Manifesto for research metrics // Nature. – 2015. – Vol. 520. – P. 429-431.
 5. Мельникова Е.В. Сравнительный анализ современных подходов России и Китая к оценке результатов научной деятельности // Проблемы национальной стратегии. – 2022. – № 1(70). – С. 153-162.
 6. Москалева О.В. Развитие наукометрии: основные вехи // В монографии «Руководство по наукометрии: индикаторы развития науки и технологии» / Под ред. М.А. Акоева. – Екатеринбург: Изд-во Урал. ун-та. – 2021. – 358 с.
 7. Гиляревский Р.С., Мельникова Е.В. Институт научной информации США: идеология, преобразования, продукты // Научно-техническая информация. Сер. 1. – 2017. – № 10. – С. 26-31.
 8. Hirsch J.E. An index to quantify an individual's scientific research output // Proceedings of the National Academy of Sciences of the United States of America. – 2005. – № 102(46). – P. 16569-16572.
 9. Egghe Leo. Expansion of the field of informetrics: origins and consequences // Information Processing & Management. – 2005. – Vol. 41, № 6. – P. 1311-1316.
 10. Prathap G. Hirsch-type indices for ranking institutions' scientific research output // Current Science. – 2006. – Vol. 91, № 11. – P. 1439.
 11. Garfield Eugene. "Science citation index" – a new dimension in indexing // Science. – 1964. – Vol. 144. – P. 649-654.
 12. Цветкова В.А., Калашникова Г.В. Альтметрические показатели в оценке региональной публикационной активности // Информационные ресурсы России. – 2021. – № 4(182). – С. 20-23.
 13. Симоненко Т.В. Наукометрия: объект, предмет, методология // Наукометрия: методология, инструменты, практическое применение: сб. науч. ст. / Ред. А.И. Груша и др. – Минск: Белорусская наука. – 2018. – 343 с.
 14. Даденко В.А., Даденко С.В. Метрические исследования как форма анализа научной продуктивности // Аналитика и научное проектирование. – 2019. – № 2. – С. 130-136.
 15. Мицкевич А.К. К вопросу о сущности и истоках политической медиаметрии // Философско-гуманитарные науки: сб. науч. ст. / Ред. В.А. Гайсёнок и др. – Минск: Изд-во РИВШ. – 2017. – 420 с.
 16. Bredford S.C. Sources of information on specific subjects // Engineering. – 1934. – Vol.137. – P. 85-86.
 17. Lotka A.J. The frequency distribution of scientific productivity // Journal of the Washington Academy of Science. – 1926. – № 12. – P. 317-323.
 18. Zipf G.K. Selected Studies of the Principle of Relative Frequency in Language. – Cambridge, MA: Harvard University Press. – 1932. – 51 p.
 19. Румянцев Д.М. Социальная инженерия и технология Big Data // Шестая межд. науч.-практ. конф. «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск, Республика Беларусь, 20-21 мая 2020 г. / Сб. материалов. Ч 3. – Минск: Бестпринт. – 2020. – 458 с.
 20. Nor Asiakin et al. Exploring big data traits and data quality dimensions for big data analytics application // Springer Science and Business Media / Journal of Big Data. – 2021. – Vol. 8. – P. 1-15.
 21. Elshawi R., Sakr S., Talia D. Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service // Big Data Research. – 2018. – Vol. 14. – P. 1-11.
 22. Мельникова Е.В. Особенности наполнения научных баз данных для эффективного применения технологии Big Data // Информационные ресурсы России. – 2021. – № 4(182). – С. 6-11.
 23. Губа К.С. Большие данные в исследовании науки: новое исследовательское поле // Социологические исследования. – 2021. – № 6. – С. 24-33.
 24. Keshav Singh R., Sandeep Kumar S. Emerging trends and global scope of big data analytics: a scientometric analysis // Quality & Quantity. – 2021. – Vol. 55, № 2. – P. 1-26.

Материал поступил в редакцию 09.03.22.

Сведения об авторе

МЕЛЬНИКОВА Елена Владимировна – кандидат технических наук, старший научный сотрудник Отделения теоретических и прикладных проблем информатики ВИНТИ РАН, Москва
e-mail: verden.mel@yandex.ru