

Измерение информационного общества

ИССЛЕДОВАНИЯ И РАЗРАБОТКИ В ОБЛАСТИ ТЕХНОЛОГИЙ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ

Хохлов Юрий Евгеньевич

*Кандидат физико-математических наук, доцент
Институт развития информационного общества, председатель совета директоров
РЭУ имени Г.В. Плеханова, научный руководитель базовой кафедры цифровой экономики ИРИО
Москва, Российская Федерация
yuri.hohlov@iis.ru*

Шапошник Сергей Борисович

*Карельский научный центр РАН, ОКНИ, лаборатория цифровых технологий регионального развития,
старший научный сотрудник
Петрозаводск, Российская Федерация
sergei.shaposhnik@gmail.com*

Аннотация

Разработаны концептуальная схема и набор показателей для мониторинга исследований и разработок как фактора развития и использования технологий работы с большими данными. В концептуальной схеме мониторинга российских исследований и разработок в области больших данных выделены три основных компонента: масштабы, уровень, а также инновационные и международные связи. Произведен расчет показателей для России и сделан сравнительный анализ полученных результатов с другими странами.

Ключевые слова

большие данные, технологии работы с большими данными; исследования и разработки; публикационная активность; патентная активность; наукометрические показатели; мониторинг и оценка

Введение

Инновационность является характерной чертой цифровой экономика, поскольку она связана с развитием и внедрением современных цифровых технологий, которое сопровождается преобразованием деловых процессов, выводом на рынок новых товаров и услуг, изменением бизнес-моделей.

Опыт гигантов и стартапов цифровой индустрии говорит о том, что научный фундамент для современных инноваций в области информационно-коммуникационных технологий (ИКТ) становится обязательным и определяет успех развития компаний, даже если в период их создания интенсивные научные исследования не проводились. Характерно, что в базе данных Web of Science (WoS), в которой индексируются наиболее значимые научные публикации (прежде всего в области фундаментальных исследований), довольно быстро растет число публикаций сотрудников крупных цифровых компаний. Например, в первые годы существования компании Facebook это были единичные публикации, тогда как в последние годы (2017–2020) ежегодно в WoS индексируется несколько сотен публикаций (в т. ч. по технологиям работы с большими данными), подготовленных сотрудниками исследовательских подразделений. Исследования нужны компании Facebook, чтобы улучшать сервисы социальной сети, использовать таргетированную рекламу как один из основных способов монетизации, развивать новые направления бизнеса и т. д.

Наличие развитой сферы исследований и разработок в области технологий хранения и анализа больших данных необходимо также для обеспечения национальной безопасности – в ряде

© Хохлов Ю.Е. Шапошник С.Б., 2021.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>
https://doi.org/10.52605/16059921_2021_04_90

чувствительных областей возможно использование только отечественных решений для работы с большими данными.

В данной статье представлены концептуальная схема и методология мониторинга исследований и разработок в области технологий работы с большими данными, а также результаты оценки этой области в 2020 году, проведенной на основе разработанной методологии (и части предложенных показателей).

1 Определение предметной области мониторинга и оценки

Предметом мониторинга и оценки в данном направлении выступает состояние исследований и разработок в области технологий работы с большими данными в Российской Федерации.

В документах стратегического планирования, которые приняты в последнее время (федеральный проект «Цифровые технологии», дорожные карты по развитию сквозных цифровых технологий), заложены традиционные представления об инновационной цепочке: результаты отечественных научных исследованиях должны использоваться при разработке технологических решений, которые, в свою очередь, коммерциализируются, выводятся на российский и зарубежные рынки и внедряются в различные сферы деятельности, приводя к их трансформации. Хотя эта логика не покрывает всех механизмов инноваций – научные исследования не являются их единственным источником, – в ней представлена значительная часть инноваций, наиболее существенных и прорывных. Поэтому данную предметную область мониторинга можно определить как исследования и разработки в области технологий работы с большими данными, находящиеся на начальных уровнях готовности цифровых технологий (УГТ 1–4 в терминах шкалы уровней готовности технологий, см. [1, 2]). Результаты исследований и разработок в этой области как правило представлены в виде научных публикаций и патентов, измеримые характеристики которых используются при мониторинге и оценке.

Для операционализации определения предметной области «Исследования и разработки в области технологий работы с большими данными» необходимо охарактеризовать ее с помощью действующих классификаторов. Сразу отметим, что анализ подобных (достаточно узких направлений) исследований и разработок затруднен тем, что существующие классификаторы (научных областей, патентов) достаточно «грубы» и не предоставляют возможности выделить предметную область путем формального перечисления кодов существующих классификаторов. В данном случае для мониторинга исследований в области технологий работы с большими данными, основанного на анализе массивов публикаций и патентов был реализован подход, включающий два следующих этапа.

1. Определение более широкой, формально определяемой предметной области исследований и разработок, в рамках которой будет выделяться более узкая область, связанная с технологиями работы с большими данными. Такой широкой предметной областью для целей настоящей работы являются исследования и разработки в области ИКТ.

Организация экономического сотрудничества и развития (ОЭСР) разработала классификатор областей науки и технологий (Fields of Science and Technology, FOS 2007) [3], с которым в настоящее время гармонируются отечественные классификаторы. В рамках данной работы для определения массива публикаций в области ИКТ использованы две области FOS 2007 на 2-м уровне классификатора: “1.2 Computer and information sciences” и “2.2 Electrical engineering, Electronic engineering, Information engineering”.

Для выделения массива патентов в области ИКТ предлагается использовать Международную патентную классификацию и группировку ее кодов, относящихся к ИКТ, которая предложена ОЭСР для мониторинга развития информационного общества [4].

2. Для выделения узкой предметной области, связанной с исследованиями и разработками по технологиям работы с большими данными, сформирован перечень ключевых слов и поисковых образов, позволяющих выбрать в международных базах данных публикаций и патентов релевантные документы (см. раздел 4).

Массив публикаций и патентов, полученный на основе сформированных поисковых образов, анализируется в соответствии с набором показателей из концептуальной схемы мониторинга (см. раздел 2).

2 Концептуальная схема и показатели мониторинга исследований и разработок в области работы с большими данными

2.1 Обзор международных стандартов и показателей мониторинга сферы исследований и разработок

Международные стандарты статистического наблюдения (мониторинга) сферы исследований и разработок были предложены и начали использоваться еще в 1960-х годах. В 1963 г. ОЭСР впервые опубликовала методологию сбора статистических данных об исследованиях и разработках – Руководство Фраскати [5]. Это руководство обязано своим названием итальянскому городу, где в 1962 году Рабочая группа национальных экспертов ОЭСР по показателям науки и техники впервые согласовала общий подход к измерению и представлению статистических данных о НИОКР. Руководство пересматривалось семь раз для решения новых задач и учета возникающих интересов пользователей, 7-е издание опубликовано в 2015 г. Руководство Фраскати в своей статистической практике используют не только страны ОЭСР, но и большинство стран мира; российское статистическое наблюдение за сферой исследований и разработок также гармонизировано со стандартами ОЭСР. В руководстве даны определение и указания по измерению таких аспектов науки, как типы исследований и разработок (фундаментальные исследования, прикладные исследования, разработки), научный персонал (исследователи, техники, вспомогательный персонал), области науки и сектора проведения научных исследований (бизнес, вузы, государство, частные некоммерческие организации). Показатели, основанные на Руководстве Фраскати, получили впоследствии название показатели «входа», так как характеризуют кадровые, финансовые и организационные ресурсы, выделяемые на проведение исследований и разработок.

В 1960-е годы начал активно развиваться еще один подход к мониторингу и исследованию научной деятельности: разработка и использование библиометрических показателей, характеризующих научный «выход» – результаты исследований и разработок, представленные в научных публикациях и патентах. Важным стимулом развития этого направления стало создание в 1961 г. Института научной информации (ISI, США), который вышел на рынок научной информации с библиографической системой «Индекс научного цитирования» (Science Citation Index). В настоящее время она принадлежит компании Clarivate Analytics и известна под торговой маркой Web of Science [6].

Возможность анализа науки с точки зрения «выхода» и значимость для развития этого направления индексов научного цитирования основана на специфических особенностях научной деятельности. В этой связи достаточно упомянуть следующие обстоятельства.

1. Основными «продуктами» научной деятельности являются научные публикации и патенты, в которых фиксируются и обосновываются результаты и методы исследований и разработок. Достоверное отражение полученных результатов научной деятельности в информационной системе науки является социальной нормой, соблюдение которой достаточно строго контролируется.

2. Большие возможности для реализации подхода дает отражение в библиографических системах научных ссылок индексируемых публикаций. В любой научной работе используются результаты предыдущих исследований: каждый научный результат влетает в единую ткань развивающегося научного знания. При подготовке публикации научный работник обязательно помещает свои результаты в контекст развития определенной исследовательской области, ссылаясь на наиболее существенные для него научные публикации. Адекватное отражение в библиографических ссылках использованного в работе познавательного аппарата и результатов предыдущих исследований регулируется довольно жесткими нормами. Поэтому «систему библиографических ссылок» можно рассматривать как особую, очень информативную семиотическую систему, пригодную для решения целого ряда задач, в том числе для оценки состояния исследований, определения структуры и приоритетов переднего фронта исследований, определения публикаций и журналов, которые оказывают наибольшее воздействие на развитие науки, анализа инновационных связей и т. д.

Показатели «входа» и «выхода» широко используются в настоящее время в национальных и международных системах мониторинга, а также в международных рейтингах развития сферы научно-исследовательских и опытно-конструкторских работ (НИОКР). Эти показатели можно объединить в группы для характеристики:

- масштабов национальных научных систем (доля затрат на исследования и разработки в ВВП страны, доля исследователей в населении, количественные и нормированные показатели публикационной и патентной активности);
- секторальной и дисциплинарной структуры затрат и научного персонала (доля в расходах на НИОКР государства, бизнеса, некоммерческого сектора в т.ч. в разрезе научных дисциплин и направлений);
- уровня исследований (цитируемость публикаций – средняя цитируемость, доля высокоцитируемых статей, публикации в ведущих рецензируемых журналах из квартилей Q1 и Q2 и др.);
- международного сотрудничества (доля публикаций в международном соавторстве);
- инновационных связей (цитирование научных публикаций в патентах, сотрудничество организаций академического сектора и бизнеса).

В контексте данной работы важно отметить, что в системах мониторинга и оценки технологического развития показатели состояния НИОКР используются как характеристика одного из факторов развития. Это относится как к оценкам общего технологического и инновационного развития (например, в Глобальном индексе инноваций [7]), так и определенного направления технологического развития (например, в индексах развития и использования технологий искусственного интеллекта [8]). Публикационная активность в них измеряется, как правило, на основе библиографических баз данных Web of Science и Scopus, патентная активность – на основе баз данных патентов, а для международных сравнений – на основе наиболее полных международных систем патентной информации (например, Patentscope Всемирной организации интеллектуальной собственности [9]).

Если речь идет о более узких технологических направлениях, таких как развитие и использование технологий работы с большими данными, определенную проблему составляет его выделение в статистике «входа» и «выхода» – действующие классификаторы, как правило, достаточно грубы и не позволяют охарактеризовать его напрямую. Несколько подходов к решению этой проблемы для показателей «выхода» рассмотрены в следующих разделах.

2.2 Концептуальная схема предметной области

Результаты анализа из предыдущих разделов позволяют задать концептуальную схему мониторинга исследований и разработок в области технологий работы с большими данными в виде трех основных компонентов: (1) масштабы исследований и разработок; (2) уровень национальных исследований по сравнению с мировым; (3) международные и инновационные связи, трансфер технологий (см. таблицу 1).

Таблица 1 - Концептуальная схема мониторинга исследований и разработок в области технологий работы с большими данными

Масштабы исследований и разработок	Уровень исследований и разработок	Международные и инновационные связи
Публикационная активность	Цитируемость работ	Сотрудничество научных организаций и бизнеса
Патентная активность	Публикации в ведущих журналах	Использование результатов исследований в технических изобретениях
Количество исследователей		Международное сотрудничество

Масштабы исследований и разработок традиционно измеряются показателями вооруженности экономики знаниями и техническими изобретениями (массивы публикаций и патентов страны, отнесенные к размеру национальной экономики) и ее обеспеченности исследовательскими кадрами и финансовыми ресурсами, выделяемыми на НИОКР (доля расходов

на НИОКР в ВВП, доля исследователей в занятом населении) [7]. В данной работе речь будет идти о масштабах исследований и разработок в области технологий работы с большими данными.

Помимо масштаба исследований и разработок важно оценить их «качество». Уровень научных исследований характеризуется общепринятыми (в том числе принятыми органами управления наукой) показателями, связанными с цитируемостью работ и публикациями в ведущих рецензируемых журналах.

Для развития цифровой экономики важно использование фундаментальных знаний для разработки технологий и решений, поэтому для мониторинга важно оценить инновационные связи и трансфер технологий, то есть понять, насколько проводимые исследования используются для технических изобретений, а разработки – бизнесом. Этот компонент предметной области мониторинга в свою очередь содержит два направления оценки: (1) научное сотрудничество научно-образовательных организаций (вузов и НИИ) с бизнесом, (2) использование результатов исследований в технических изобретениях. Важным аспектом в этом компоненте предметной области служит характеристика международных научных связей, которые являются каналом обмена знаниями, исследовательскими практиками и техническими идеями.

2.3 Показатели масштабов исследований и разработок

Научные публикации и патенты являются основными результатами научных исследований и разработок, поэтому измерение их количественных характеристик регулярно используется при мониторинге – от оценки объемов проводимых исследований до анализа структуры переднего фронта исследований и оценки результатов научной деятельности. Для целей данной работы в мониторинг масштабов исследований в области технологий работы с большими данными включены библиометрические показатели публикаций российских ученых, индексированных в WoS, и российских патентных заявок, поданных в соответствии с договором о патентной кооперации (Patent Cooperation Treaty, PCT) и опубликованных в международных базах данных.

Патенты – это право интеллектуальной собственности, выданное уполномоченными органами изобретателям, позволяющее им использовать свои изобретения и эксплуатировать их в течение ограниченного периода времени (обычно 20 лет). Патенты выдаются частным лицам, фирмам или другим юридическим лицам при условии, что изобретение соответствует определенным критериям: оно должно быть новым, включать изобретательский уровень (то есть быть неочевидным) и быть пригодным для промышленного применения.

Патенты являются одним из ключевых показателей инновационной деятельности. Они используются для измерения результатов НИОКР, распространения знаний, изобретательских достижений в качестве инструмента для оценки направления исследований и стратегических целей компаний и стран. Анализ массива патентов может также дать представление о связях технических изобретений и научных исследований.

Для характеристики масштабов исследований в области технологий работы с большими данными предлагаются показатели, аналогичные включаемым в международные рейтинги и системы мониторинга состояния научно-инновационной сферы.

1. Прежде всего, это «Доля российских публикаций в области технологий работы с большими данными, %» в системе WoS.
2. Для корректных международных сопоставлений публикационная активность относится к размеру национальной экономики – используется удельный показатель «Количество публикаций по технологиям работы с большими данными на 1 млрд ВВП». Этот показатель можно трактовать как характеристику обеспеченности экономики исследованиями в области технологий работы с большими данными.
3. Для характеристики масштабов разработок в области технологий работы с большими данными и проведения международных сопоставлений, как и в случае с публикациями, используется показатель «Количество патентных заявок по технологиям работы с большими данными, поданных по процедуре PCT, на 10 млрд ВВП».

Наукоёмкость цифровой экономики характеризует еще один показатель, связанный с обеспеченностью исследований и разработок (в данном случае в области технологий работы с большими данными) кадровыми ресурсами. Учитывая, что в классификаторах областей науки, используемых для статистических наблюдений за кадровым составом науки, отсутствует специализация в области больших данных, для измерения этого важного аспекта сферы

исследований и разработок применен подход, основанный на анализе публикационной активности в этой области. В рамках этого подхода в качестве прокси-метрики для измерения обеспеченности исследователями в области больших данных предлагается использовать следующий показатель:

4. Количество уникальных авторов публикаций по технологиям работы с большими данными в WoS на 100 тыс. занятых.

2.4 Показатели уровня исследований и разработок

При мониторинге и оценке состояния научных исследований важно иметь не только объемные показатели, но и уровень исследований. Для этого предлагается использовать следующие показатели цитируемости публикаций.

1. «Средняя цитируемость публикаций, нормализованная по области исследований, году и типу публикаций». Этот показатель рассчитывается как среднее значение нормализованных показателей цитируемости публикаций страны, организации и т. п. Нормализация осуществляется делением числа цитирований, полученных публикацией, на среднюю по миру цитируемость работ в данной области того же года и типа. Такой подход позволяет при оценке убрать зависимость числа цитирований от года выхода публикации (более ранние в среднем цитируются выше – набирают больше ссылок), типа (например, обзоры цитируются чаще) и области исследований (в более «крупных» областях средняя цитируемость выше). При интерпретации значений данного показателя исходят из следующего: если показатель меньше 1, то цитируемость ниже среднемировой, если больше – выше. Тем самым показатель позволяет оценить, насколько востребованность результатов исследований страны (или организации) отличается от среднемировой.
2. «Доля высокоцитируемых публикаций среди всех публикаций по технологиям работы с большими данными». Присутствие высокоцитируемых публикаций свидетельствует о наличии прорывных исследований, оказывающих большое воздействие на отечественную и мировую науку.

2.5 Показатели международных и инновационных связей

Возможности научных исследований генерировать новые технологии характеризуют наличие устойчивых связей научных исследований и изобретательской активности, связей научных организаций с предпринимательским сектором экономики, а также международное сотрудничество, которое является важным каналом обмена знаниями, исследовательскими практиками и возможностями. Для мониторинга и оценки этого компонента предметной области введены следующие показатели.

1. «Доля публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией». Показатель характеризует трансфер технологий из академического сектора в коммерческий. Он является одной из проблемных зон российской инновационной системы – в целом по компьютерным наукам доля российских научных публикаций, выполненных в соавторстве с индустрией, на порядок меньше, чем у стран с эффективными инновационными системами, высокой инновационной активностью предприятий и успешной коммерциализацией разработок (Швеция, Швейцария, Израиль, США) [23].
2. «Доля публикаций, на которые ссылаются патенты в области технологий работы с большими данными». Данный показатель отражает способность научных исследований служить основой для разработки новых технологий.
3. «Доля публикаций по технологиям работы с большими данными, опубликованных в международном соавторстве». Участие в международном научном сотрудничестве обогащает потенциал национальной науки, позволяет обмениваться исследовательскими практиками, которые в значительной мере являются «неявным знанием» и не представлены в научных публикациях. Сотрудничество с ведущими мировыми исследовательскими центрами является и косвенным свидетельством уровня исследований. Совместная научная деятельность обычно проявляется в совместных публикациях, поэтому использование показателя публикаций с международным соавторством является наиболее естественной операционализацией понятия международного научного сотрудничества.

3 Методология измерения показателей

3.1 Обзор библиометрических исследований, посвященных анализу публикационной и патентной активности в области технологий работы с большими данными

Как уже отмечалось, проблемой для измерения показателей исследований и разработок в области технологий работы с большими данными является выделение этой области с точки зрения имеющихся классификаторов. Действующие классификаторы (областей науки, патентов, публикаций в библиографических базах данных) достаточно «грубы», не отражают эту узкую область и не позволяют измерять описанный выше спектр показателей на основе действующего статистического наблюдения (показатели «входа») и стандартных поисковых инструментов баз данных (показатели «выхода»).

В этой ситуации одним из основных подходов является опора на показатели «выхода», для которых указанная проблема разрешима за счет использования ключевых слов, характеризующих данную область. С их использованием возможно выделение массива публикаций и патентов, которые относятся к технологиям работы с большими данными, что позволяет измерить не только библиометрические показатели «выхода», но и в некоторых случаях «ресурсные» показатели. Примером вычисления показателей «входа» на основе библиометрических показателей является подсчет уникальных авторов публикаций в анализируемой узкой области, что может выступать заменой показателя числа исследователей в этой области.

Такой подход к измерению исследований и разработок в области технологий работы с большими данными (и в других аналогичных узких областях), основанный на выделении массива публикаций с использованием ключевых слов и дальнейшем изучении его библиометрических характеристик, достаточно распространен.

За последние 20 лет (2001–2020) в базе данных WoS было проиндексировано 754 публикации, в названиях, ключевых словах и аннотациях которых встречаются одновременно термины, характеризующие технологии работы с большими данными (см. ниже раздел «Методология») и термин «bibliometrics» в различных вариантах написания. Основной массив публикаций (676) появился в последние 10 лет, когда, собственно, и начались активные исследования в области работы с большими данными. При этом половина публикаций последних 10 лет приходится на 2019 и 2020 гг. (144 и 193 публикации соответственно). Отметим, что при этом в большинстве указанных публикаций библиометрические инструменты использованы для обзора литературы при изучении отдельных аспектов работы с большими данными.

Для выделения собственно библиометрических работ, посвященных измерению исследований и разработок в области технологий работы с большими данными, параметры поиска были уточнены по следующим направлениям.

1. Поиск проводился только по названиям и ключевым словам публикаций последних 10 лет.

2. Для поиска использовался только термин «big data», а не весь набор специализированных терминов, так как интерес представляла методика выделения данной области исследований (279 публикаций за последние 10 лет). Дополнительно проводился поиск публикаций, у которых в названиях и ключевых словах одновременно используются термины “big data” и “patent” – для выявления статей, анализирующих патентную активность в области технологий работы с большими данными, но не использующих термин «bibliometrics» (46 публикаций за последнее десятилетие).

3. Проводился поиск только научных статей, содержащих развернутое описание методологии библиометрических исследований. Обзорные публикации не рассматривались, так как чаще всего они посвящены отдельным аспектам работы с большими данными в различных сферах деятельности.

4. В результате уточнения поискового запроса по приведенным выше критериям были найдены 58 публикаций, а после ознакомления с полными текстами были отобраны 12 релевантных публикаций [10–21], посвященных библиометрическим исследованиям публикационной и патентной активности в области больших данных. Анализ 46 публикаций, в которых используются термины «big data» и «patent», позволил отобрать еще одну публикацию, посвященную анализу патентной активности [22]; еще одна статья входила в ранее отобранные 12 релевантных публикаций; большинство же из указанных 46 публикаций посвящены использованию методов работы с большими данными для анализа патентной информации.

Анализ отобранных публикаций с точки зрения интересующих нас поисковых стратегий показал следующее.

1. Для выделения с помощью ключевых слов исследований и разработок в области технологий работы с большими данными используется, как правило, только ключевое слово «big data» (см., например [10, 11]), в одном случае – с добавлением синонима «mega data» [12]. В то же время, в статье Лью и соавторов [13] при формировании набора ключевых слов для поискового образа проводилось специальное исследование, включающее предварительный анализ релевантных публикаций и формирование исходного перечня специфичных для данной области ключевых слов с участием экспертов, а затем проведение широкого опроса экспертов с целью верифицировать и уточнить список. В результате был сформирован следующий перечень ключевых слов для построения поискового образа: big data, data analytics, data center, data science, data warehouse, Hadoop, large dataset, machine learning, MapReduce, NoSQL, predictive analytics, sentiment analysis, unstructured data.

2. Для анализа публикационной активности в области больших данных используются, как правило, базы данных WoS и Scopus. Первая отмечается как содержащая тщательно отобранные более качественные источники, вторая – как более широкая по охвату публикаций [14]. Для исследований, содержащих показатели цитирования, чаще используется база данных WoS. Для анализа патентной активности используются базы данных с наибольшим охватом патентных документов (например, Lens Open Source Platform [20]).

3. Поиск по ключевым словам проводится в различных поисковых полях. Используются три основных подхода: (а) поиск по теме исследования (одновременно в названиях, ключевых словах, аннотациях [11]); (б) поиск в названиях и ключевых словах [15]; (в) поиск только в названиях публикаций [12]. Наиболее часто проводится поиск в названиях и ключевых словах, а также по теме исследования. В патентных базах данных обычно используется поиск по текстовым полям – например, одновременно в названиях, рефератах и формуле изобретения [20].

4. При выделении полного массива публикаций важным является выбор типа публикаций для анализа. В отмеченных статьях [10–21] присутствует несколько подходов: (а) анализ только научных статей [13]; (б) анализ нескольких типов публикаций (например, “article”, “proceedings paper”, “editorial material”, “review”, “meeting abstract” [16]). Наиболее часто для анализа публикационной активности выбирают статьи, особенно при изучении совместных исследований, связей областей исследования или тематических трендов (см. также [17]).

3.2 Методология измерения показателей мониторинга исследований и разработок в области работы с большими данными

Выбор WoS для измерения и мониторинга публикационной активности связан с тем, что эта система является наиболее авторитетной международной библиографической базой, для индексации в которой тщательно отбираются рецензируемые журналы и другие научные издания. Показатели, основанные на анализе публикаций, индексированных в WoS, используются при оценке результатов научной деятельности и проведении широкого круга науковедческих исследований. Важным является то обстоятельство, что для анализа публикаций, представленных в WoS, разработаны аналитические инструменты, позволяющие получать широкий спектр библиометрических показателей (InCites). Наряду с этим, для целей данной работы также использовались библиографическая база данных Scopus и соответствующий аналитический инструментарий SciVal. База данных Scopus является конкурентом WoS, близка ей по функциональным возможностям и целям разработки, при этом в ней индексируется более широкий спектр источников. Выбор базы данных Patentscope Всемирной организации интеллектуальной собственности (ВОИС) для измерения показателей патентной активности определялся тем, что ВОИС является официальным публикатором международных патентных заявок, поданных по процедуре PCT, а используемые показатели основаны на их подсчете.

Ключевым вопросом для наукометрического анализа исследований в области технологий работы с большими данными является выделение публикаций, относящихся к данной предметной области, из всего публикационного потока, индексированного в используемых библиографических базах данных. Как отмечалось ранее при обзоре проведенных исследований других авторов, обычным подходом является использование ключевого слова “big data” для отбора публикаций, имеющих это словосочетание в названиях или ключевых словах. Подобный подход существенно сужает выборку и приводит к потере значительного количества релевантных публикаций.

Отказ в данной работе от использовавшегося ранее подхода связан как с общими соображениями, так и с результатами проведенных экспериментов и экспертных процедур. Общая гипотеза, подтвержденная в ходе поисковых экспериментов, состоит в том, что в достаточно развитых и обширных областях исследования публикации, посвященные специальной узкой тематике (или технологии), зачастую перестают упоминать в названиях и ключевых словах более широкую «материнскую» область исследований, но характеризуют тематику исследования специфическими для нее терминами. Тем самым, работа, оставаясь в рамках обширной предметной области исследований (в данном случае – больших данных), теряет в метаописаниях (ключевых словах, аннотациях, названиях) прямую терминологическую связь с ней и характеризуется более специальным набором терминов. Одним из примеров в нашем случае является такая специфическая технология работы с большими данными, как Hadoop, – из 1794 публикаций, проиндексированных за последние 5 лет в WoS и содержащих в названиях или ключевых словах термин “Hadoop”, более половины (1022) не содержат одновременно термина “big data”.

С учетом сказанного, для формирования массива публикаций в области технологий работы с большими данными был реализован следующий подход:

1. На первом этапе на основе консультаций с экспертами в области технологий работы с большими данными и анализа публикаций был сформирован исходный список ключевых слов, характеризующих данную область. Основными критериями отбора терминов для ключевых слов были:
 - a. релевантность – термин используется в рамках области технологий работы с большими данными;
 - b. полезность для целей данного исследования – термин расширяет искомый массив и позволяет идентифицировать публикации, которые не имеют в названии или других метаданных термина “big data”;
 - c. отсутствие «шума» в результатах поиска – т.е. основная часть публикаций, содержащих предлагаемый термин в названии, ключевых словах или аннотации, должна относиться к указанной области исследований и разработок.
2. На втором этапе предложенный список был верифицирован в ходе опроса экспертов по работе с большими данными. Предложенный набор ключевых слов был подтвержден, наибольшее число баллов, помимо “big data”, набрали ключевые слова “Hadoop” и “MapReduce”.
3. На основе разработанного списка ключевых слов был сформирован запрос для поиска в библиографических базах данных, учитывающий различные варианты написания терминов.

Дополнительно было принято решение использовать для поиска не только названия публикаций и ключевые слова, но и аннотации. Проведенные эксперименты и экспертная оценка поисковых выдач показали, что подавляющее число публикаций, имеющих в аннотациях искомые ключевые слова, относятся к области технологий работы с большими данными.

В итоге был сформирован следующий поисковый образ для публикаций в области технологий работы с большими данными:

“big data*” OR bigdata OR “large dataset*” OR “massive data*” OR “data science” OR “data* mining” OR “datamining” OR “text mining” OR “Hadoop*” OR “MapReduce” OR “Map Reduce” OR “unstructured data*” OR “semistructured data*” OR “semi-structured data*” OR “data analytic*” OR “descriptive analytic*” OR “diagnostic analytic*” OR “predictive analytic*” OR “prescriptive analytic*”

Указанный поисковый образ использовался для формирования массива публикаций за 2016–2020 годы. Поиск проводился в поле «Тема» (IS – поиск по названиям, ключевым словам и аннотациям) в базе данных WoS, в аналогичном поле в базе данных Scopus и в текстовых полях патентной базы Patentscope.

Библиометрический анализ исследований и разработок в области технологий работы с большими данными был проведен на массиве публикаций, выделенных по описанной выше методике и относящихся только к двум областям науки, которые связаны с ИКТ (по классификатору FOS 2007): “1.2 Computer and information sciences” и “2.2 Electrical engineering, Electronic engineering, Information engineering”.

4 Результаты

Описанная выше методология мониторинга исследований и разработок в области технологий работы с большими данными была протестирована в ходе пилотной реализации, в рамках которой было измерено большинство показателей «выхода», основанных на анализе публикационной активности. Измерение проводилось в основном на массиве публикаций за 2019 год, т.к. индексирование работ, опубликованных в 2020 году, на момент подготовки публикаций не было еще завершено. Анализ полученных результатов представлен в данном разделе.

4.1 Масштабы исследований и разработок

В 2019 году Россия занимала 17-е место среди стран по общему числу публикаций в области технологий работы с большими данными. Доля России от общего числа публикаций – 1,53%, что в двадцать раз меньше, чем у стран-лидеров (Китая и США – см. рис. 1). При этом Россия отстает по этому показателю от таких стран, как Бразилия, Пакистан и Саудовская Аравия.

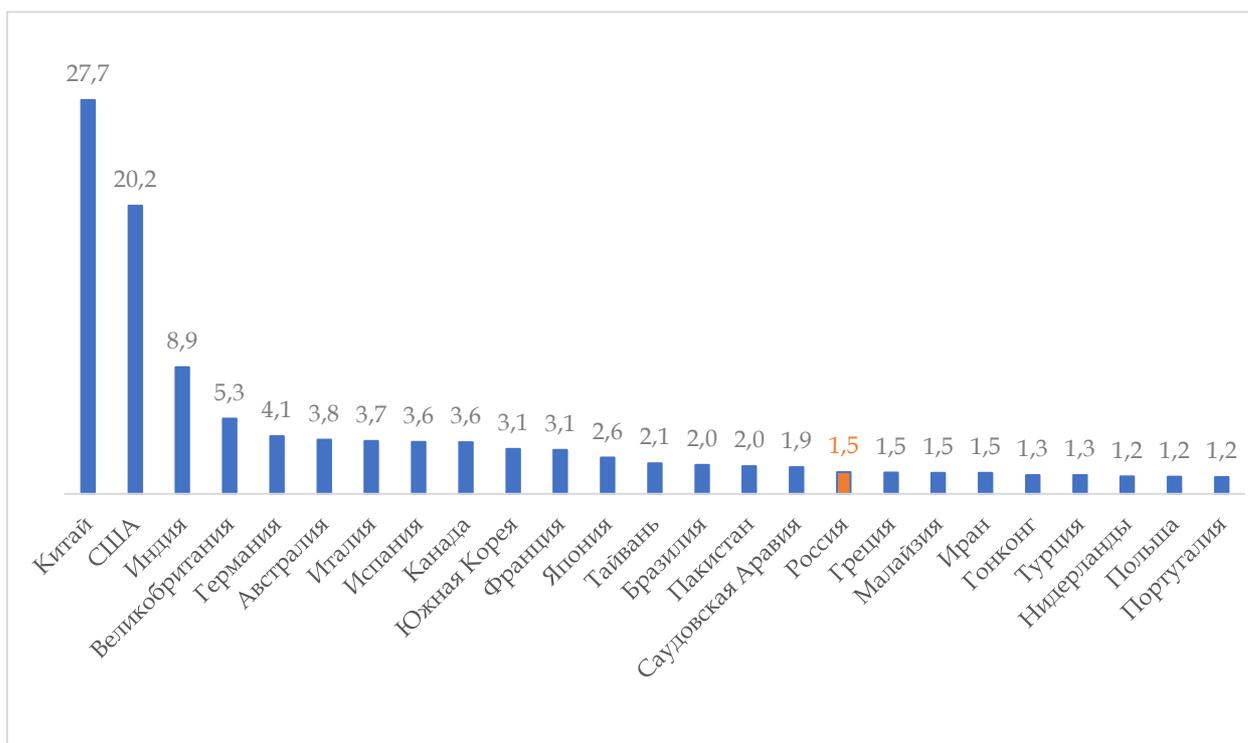


Рисунок 1. Доля публикаций России и стран мира от общего числа публикаций в области технологий работы с большими данными, индексируемых в WoS (25 стран-лидеров по числу публикаций, 2019 г.), %

Если посмотреть на динамику (рис. 2), то доля российских работ по сравнению с 2016 г. немного выросла (была 1,32%). В целом с точки зрения числа публикаций в области технологий работы с большими данными у пятерки стран-лидеров динамика разнонаправленная. У Китая продолжается устойчивый рост доли публикаций (более чем на 3 процентных пункта с 2016 г. по 2019 г.); США несколько снизили свое присутствие (на 1 процентный пункт с 2016 г. по 2019 г.); падает доля Индии; отмечается небольшой рост у Великобритании; доля Германии сначала уменьшилась, но последние 3 года медленно, но стабильно, растет. По неполным данным в 2020 г. эти тенденции в целом сохраняются и даже усиливаются – Китай существенно нарастил долю своих публикаций на фоне явно выраженного падения данного показателя у США.

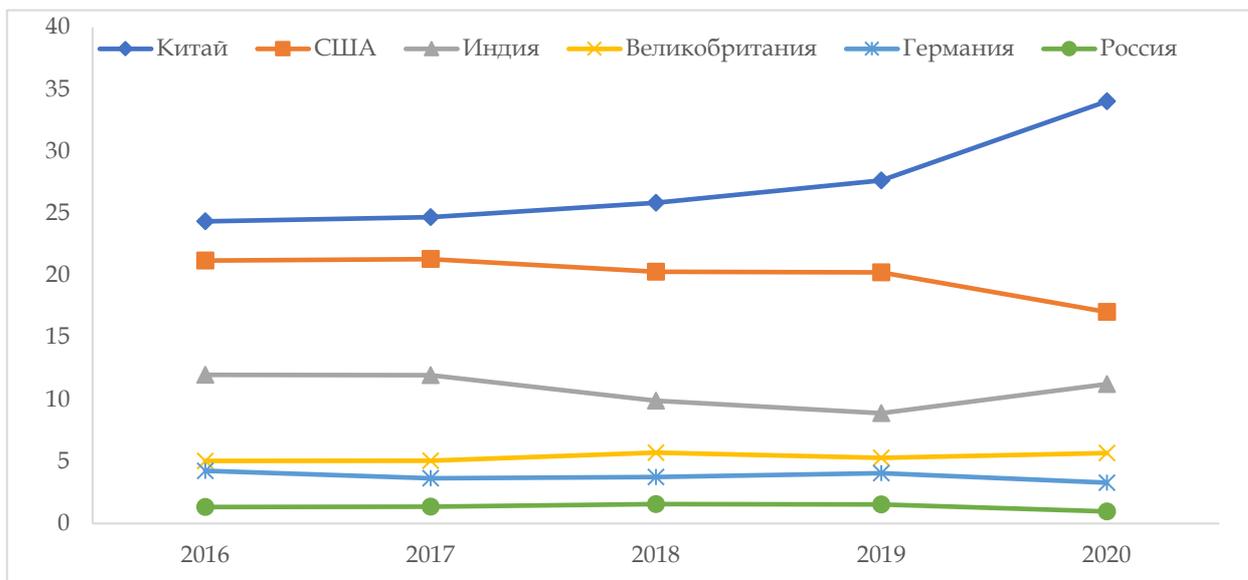


Рисунок 2. Динамика доли публикаций России и 5 стран-лидеров по числу публикаций в области технологий работы с большими данными, индексируемых в WoS, 2016–2020 гг., %

Для более корректных международных сопоставлений, учитывающих масштабы экономики сравниваемых стран, при мониторинге используется удельный показатель «Количество публикаций по технологиям работы с большими данными на 1 млрд ВВП». На рисунке 3 представлены рассчитанные для 2019 г. значения этого показателя для 25 стран с наибольшим числом публикаций в области технологий работы с большими данными. Для расчета использовались данные МВФ о размерах национальных экономик, рассчитанных по паритету покупательной способности (ППС).

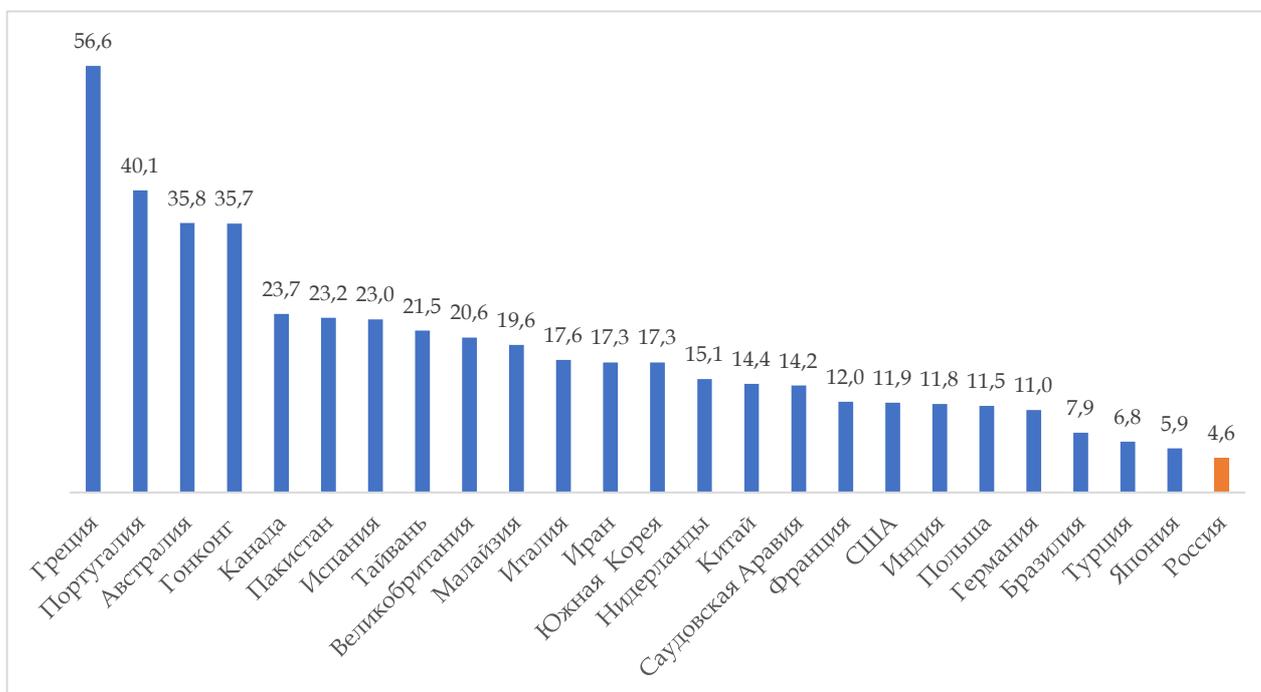


Рисунок 3. Количество публикаций в области технологий работы с большими данными на 1 млрд. ВВП по ППС, индексируемых в WoS (25 стран-лидеров по числу публикаций, 2019 г.), шт.

Россия в 2019 г. на порядок отставала от стран-лидеров по этому показателю и существенно уступала даже странам с большими экономиками (США, Китай, Индия, Германия).

Для мониторинга патентной активности были проанализированы сведения по международным патентным заявкам, поданным по процедуре РСТ. Поиск проводился в патентной базе данных Patentscope ВОИС. Учитывая, что заявки по процедуре РСТ могут подаваться на национальных языках участников договора (включая русский), для корректных международных сопоставлений поиск российских заявок осуществлялся на русском и английском языках, а для сравнения выбирались англоязычные страны. Поиск проводился по ключевым словам во всех текстовых частях патентных заявок. Для нормализации показателей использовались данные МВФ о размерах ВВП в международных долларах по паритету покупательной способности [24].

На рисунке 4 приведены рассчитанные за 2016–2020 гг. значения показателя «Количество патентных заявок, поданных по процедуре РСТ, по технологиям работы с большими данными на 10 млрд ВВП» для России и англоязычных стран.

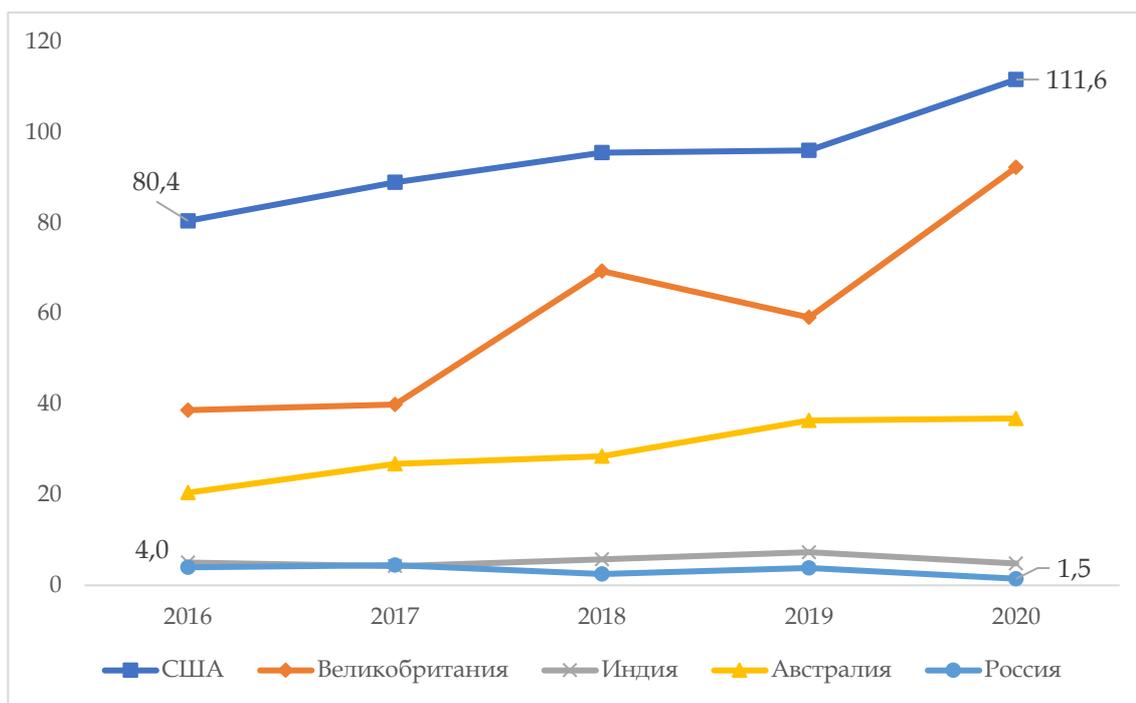


Рисунок 4. Количество международных патентных заявок в области технологий работы с большими данными, поданных по процедуре РСТ, на 10 млрд долларов ВВП по ППС (Россия и англоязычные страны, 2016–2020 гг.), шт.

Россия в 2020 г. почти на два порядка отставала от США по этому показателю (в 76 раз), в остальные годы разрыв варьировался от 20 до 38 раз. Значения показателя у России во все анализируемые годы были близки к индийским и, в отличие от других сравниваемых стран, не имели тенденции к росту – значения показателя у США за рассматриваемый период выросли почти в 1,5 раза, а у Великобритании – более чем в 2 раза.

Как и в случае публикационной активности, в последние годы в области подачи патентных заявок по технологиям работы с большими данными доминирует Китай. На рисунке 5 приведены данные за 2020 г. по доле патентных заявок, поданных по процедуре РСТ, у которых в названии или реферате заявки встречаются разработанные для данного исследования ключевые слова (см. выше раздел 4). Ограничение поиска только по названиям и рефератам в данном случае вызвано необходимостью проведения международных сопоставлений – во всех заявках РСТ, даже поданных на национальных языках, есть название и реферат на английском языке (отметим еще раз, что приведенные выше на рисунке 4 данные – результат поиска по всем текстовым полям заявок, поэтому данные по странам различаются).

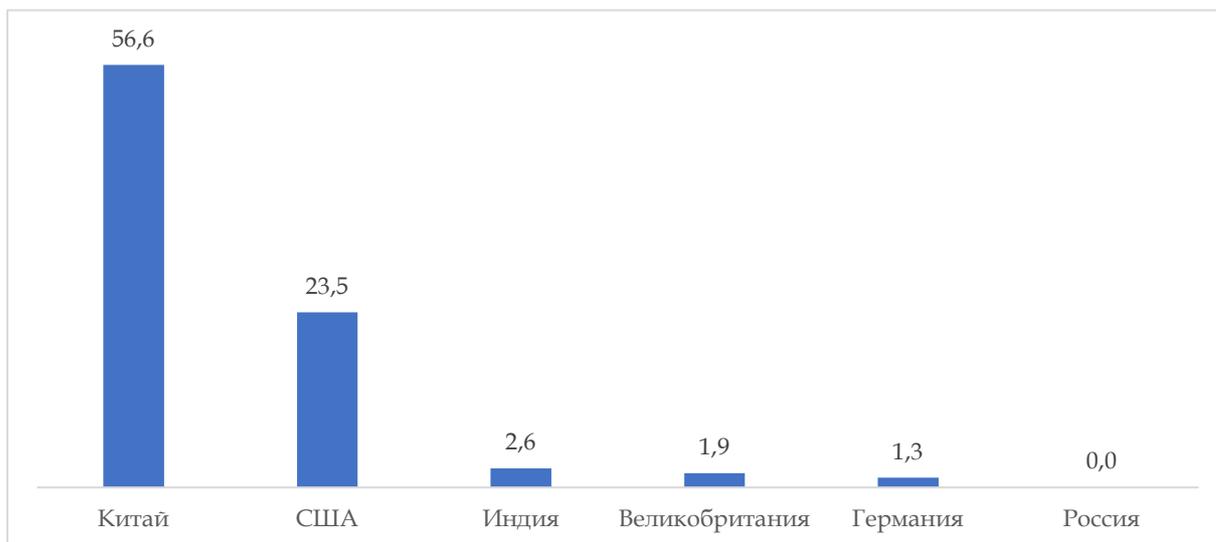


Рисунок 5. Доля международных патентных заявок, поданных по процедуре РСТ, в названии и реферате которых встречаются ключевые слова технологий работы с большими данными (2020 г.), %

5.2 Уровень исследований

Среди 25 стран с наибольшим числом публикаций в области технологий работы с большими данными Россия в 2019 г. занимает 16-е место по средней цитируемости публикаций, нормализованной по области исследования, году и типу публикаций. Ее показатель равен 1.21, а это означает, что цитируемость российских работ выше среднемировой (см. рис. 6).

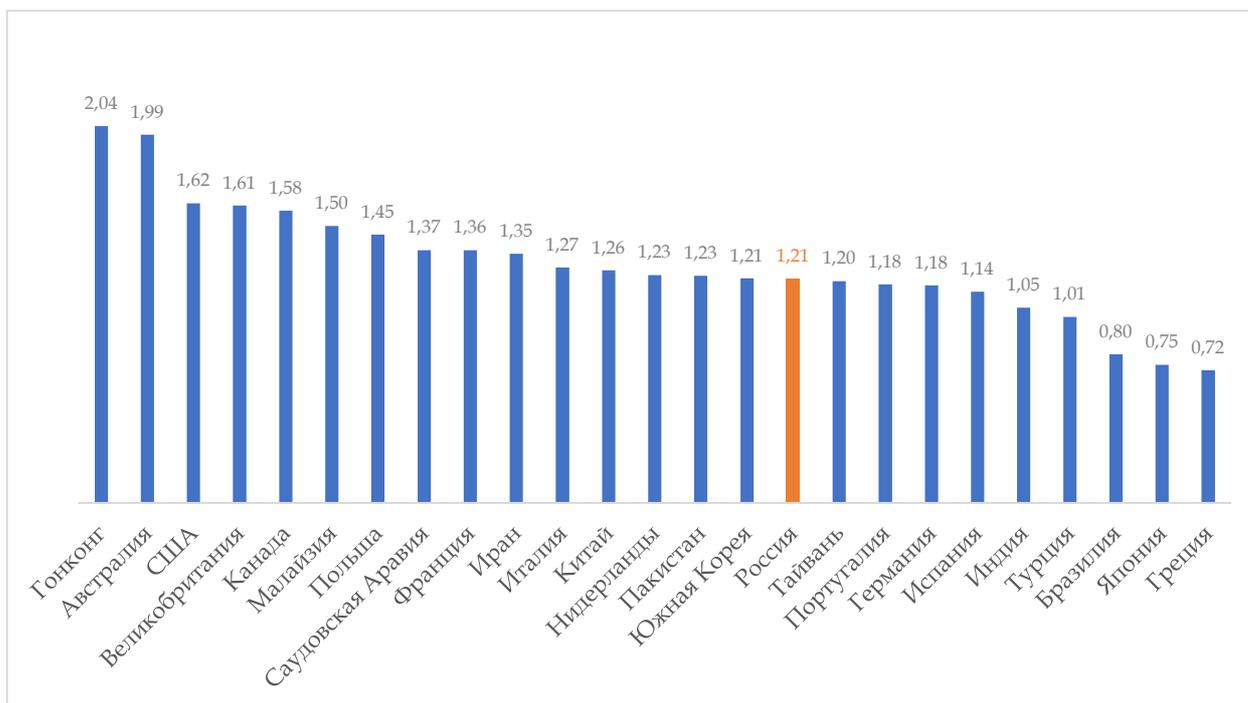


Рисунок 6. Средняя цитируемость публикаций, нормализованная по области исследований, году и типу публикаций (25 стран-лидеров по числу публикаций, 2019 г.)

Среди тенденций последних лет для пятерки самых крупных (по числу публикаций в области технологий работы с большими данными) стран и России можно отметить падение значений показателя у США, Великобритании и (наиболее выраженное) Германии; небольшой, но стабильный рост средней нормализованной цитируемости публикаций Китая, Индии и России (рис. 7).

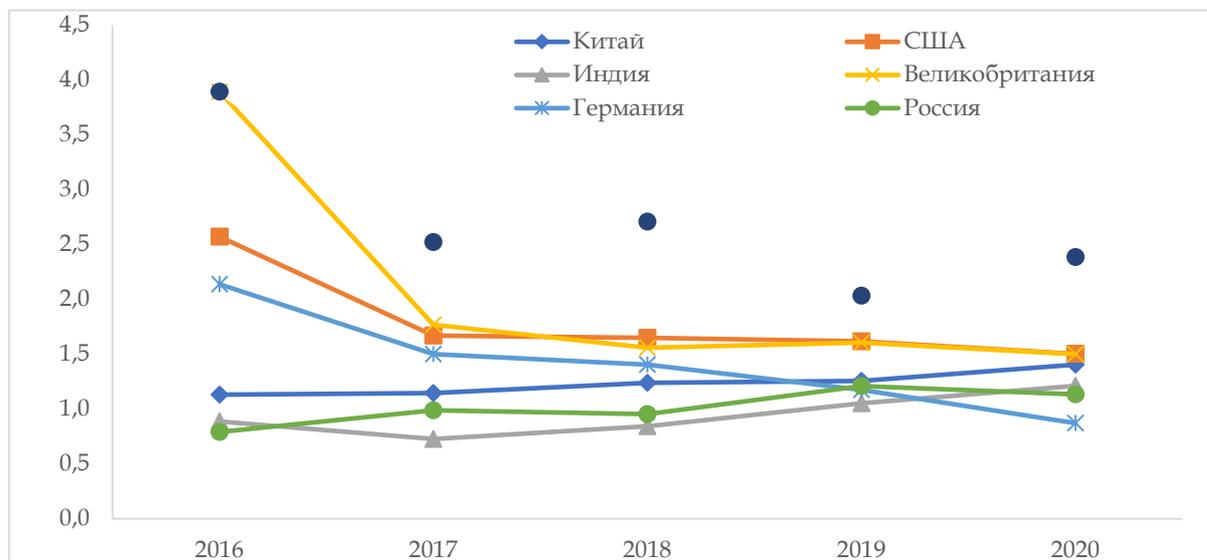


Рисунок 7. Динамика средней цитируемости публикаций, нормализованной по области исследований, году и типу публикаций (5 стран-лидеров по числу публикаций, Россия и лидер по показателю, 2016–2020 гг.)

5.3 Международные и инновационные связи

Сотрудничество академических исследователей с корпоративным сектором является важным каналом трансфера технологий и индикатором инновационных связей на первых этапах инновационного цикла. Метрика этого сотрудничества («Доля публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией») отражает потенциальную практическую значимость проводимых академических исследований и характеризует эффективность национальной инновационной системы – не случайно лидерами по этому показателю являются страны с высокой инновационной активностью предприятий и высоким уровнем технологического развития (рис. 8; из не вошедших в анализируемые 25 стран-лидеров по числу публикаций в области технологий работы с большими данными в этой связи можно отметить еще Израиль со значением этого показателя 10% и Швейцарию – со значением 14,81%). Россия по этому показателю в 2019 году занимала 19-е место среди 25 стран с наибольшим числом публикаций в области технологий работы с большими данными.

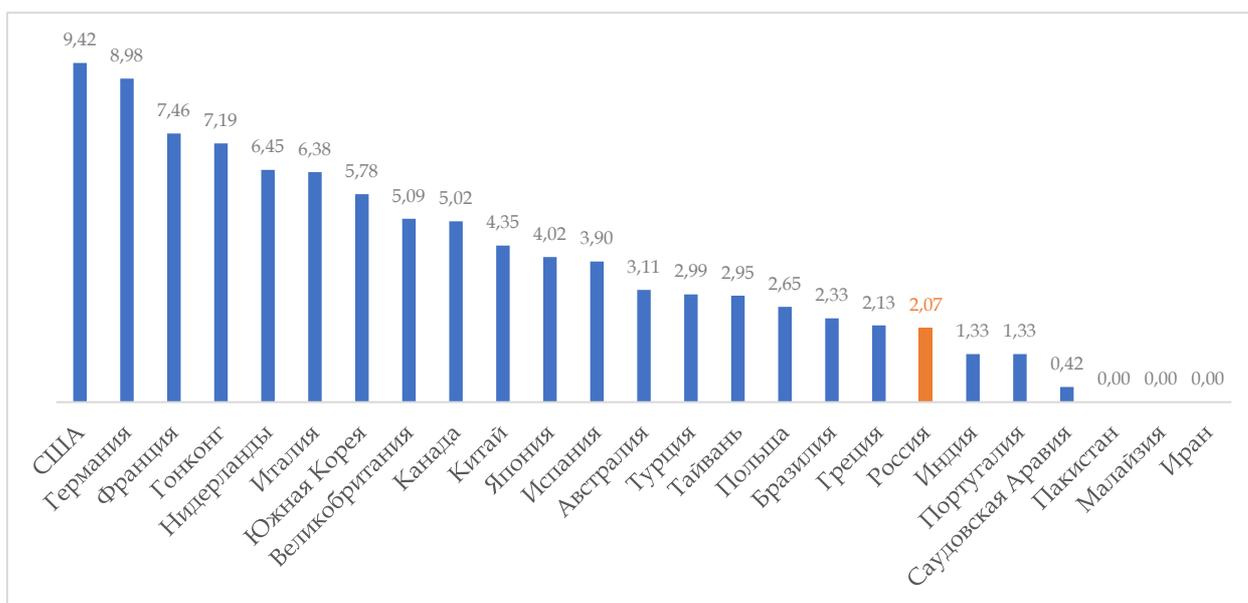


Рисунок 8. Доля публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией (25 стран-лидеров по числу публикаций, 2019 г.), %

В последние годы среди пяти стран, наиболее активно публикующихся в области технологий работы с большими данными, Германия и США становились лидерами и по этому показателю (рис. 9). Среди анализируемых стран за последние 5 лет можно отметить небольшой тренд на снижение значения показателя у Германии, США, Великобритании и Китая и на увеличение сотрудничества с индустрией у Индии и России. При этом у России и Индии этот показатель остается ниже, чем у лидеров, а сами тенденции не являются выраженными.

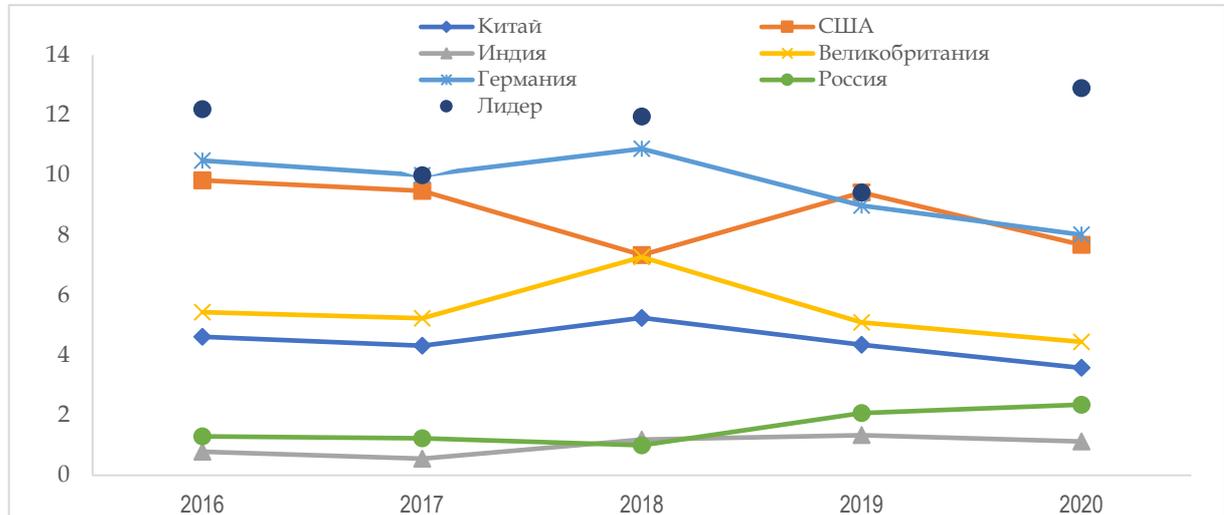


Рисунок 9. Динамика доли публикаций по технологиям работы с большими данными, опубликованных в сотрудничестве с индустрией (5 стран-лидеров по числу публикаций, Россия и лидер по показателю, 2016–2020 гг.), %

Международное научное сотрудничество – важный фактор развития национальных исследований и разработок. Совместная работа ученых стимулирует обмен научными знаниями и навыками исследования, позволяет получить доступ к уникальному научному оборудованию, весь спектр которого не может позволить себе ни одна из стран.

На рисунке 10 приведены данные о международном соавторстве публикаций 25 стран, представленных наибольшим числом публикаций в WoS. Россия занимает среди них 22-е место, публикуя в международном соавторстве около 31% работ.

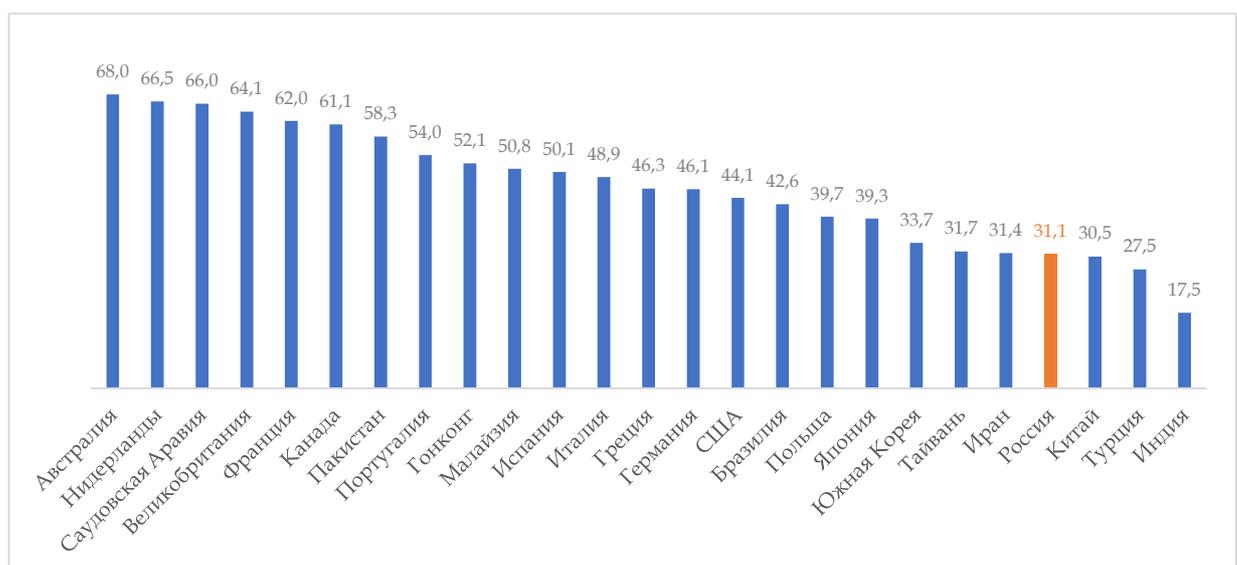


Рисунок 10. Доля публикаций по технологиям работы с большими данными, опубликованных в международном соавторстве (25 стран-лидеров по числу публикаций, 2019 г.), %

Следует отметить, что интернационализация научной деятельности — одна из ярких и характерных тенденций развития современной науки. Рост масштабов и разнообразия форм международного научного сотрудничества существенно изменил за последние десятилетия облик науки. Эта тенденция — динамичный и устойчивый рост числа совместных публикаций исследователей из разных стран — отмечается в последние годы и для исследований в области технологий работы с большими данными (рис. 11). Россия в последние два года догнала Китай по числу совместных работ и опережает Индию. При интерпретации данного показателя следует принять во внимание, что исследователи из «больших» научных стран, как правило, меньше публикуются в международном соавторстве, находя партнеров внутри страны. Россия уступает по показателю международного сотрудничества почти всем странам, имеющим в WoS большее число публикаций, что является негативным симптомом.

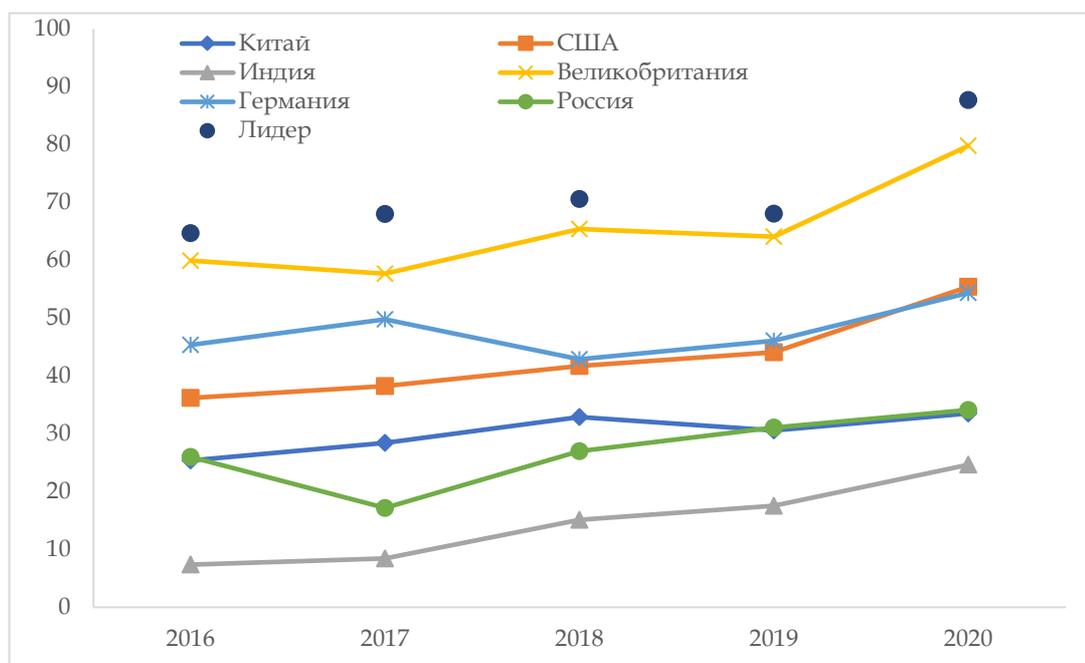


Рисунок 11. Динамика доли публикаций по технологиям работы с большими данными, опубликованных в международном соавторстве (5 стран-лидеров по числу публикаций, Россия и лидер по показателю, 2016–2020 гг.), %

Заключение

По результатам пилотной реализации системы мониторинга исследований и разработок в области технологий работы с большими данными можно сделать несколько выводов:

- 1) разработанная методология адекватно отражает данную предметную область и может быть использована при проведении полномасштабных мониторинговых исследований;
- 2) предложенная концептуальная схема мониторинга и система показателей масштабируемы и позволяют включать в систему мониторинга (без изменения ее концептуальных рамок и методических основ) новые обследования и показатели;
- 3) опыт разработки и реализации системы мониторинга исследований и разработок в области технологий работы с большими данными может быть распространен на другие сквозные технологии цифровой экономики.

Полученные результаты свидетельствуют о том, что для российских исследований и разработок в области технологий работы с большими данными характерен ряд проблем:

- 1) масштабы публикационной активности российских исследователей, представленные в ведущих международных библиографических системах, не соответствуют масштабам экономики и научной деятельности в стране;

2) показатели уровня российских исследований близки к среднемировым, постепенно растут, но значительно уступают странам с наиболее масштабными исследованиями в области технологий работы с большими данными;

3) инновационные связи проводимых в стране исследований и разработок, их востребованность в коммерческом секторе остаются на низком уровне – Россия в разы (в отдельных случаях на порядок) уступает странам с развитой инновационной системой по показателям сотрудничества академических исследователей с коллегами из индустрии.

Благодарности

В работе использованы результаты проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по договору МГУ имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований в рамках проекта № 18–29–03086.

Работа частично выполнена в рамках государственного задания Карельского научного центра РАН.

Литература

1. ГОСТ Р 57194.1 – 2916 Трансфер технологий. Общие положения.
2. ГОСТ Р 58048 – 2017 Трансфер технологий. Методические указания по оценке уровня зрелости технологий.
3. OECD (2007) Revised Field of Science and Technology (FOS) Classification in the Frascati Manual. URL: <http://www.oecd.org/science/innovationinsciencetechnologyandindustry/38235147.pdf> (дата обращения 15.06.2021)
4. OECD Guide to Measuring the Information Society 2011. <https://doi.org/10.1787/9789264113541-en>
5. OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. <https://doi.org/10.1787/9789264239012-en>
6. Web of Science. Clarivate Analytics. URL: <https://csi.webofknowledge.com/CSI/exit.do?Func=Exit&SID=C1LIQAIG3E5mZMtoMhE> (дата обращения 15.06.2021)
7. Cornell University, INSEAD, and WIPO (2020). The Global Innovation Index 2020: Who Will Finance Innovation? URL: <https://www.globalinnovationindex.org/> (дата обращения 15.06.2021)
8. Daniel Zhang, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault, “The AI Index 2021 Annual Report” AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021. URL: <https://aiindex.stanford.edu/report/> (дата обращения 15.06.2021)
9. Patentscope – WIPO. URL: <https://patentscope.wipo.int/> (дата обращения 15.06.2021)
10. López Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A.-J., & Parra-González, M. E. (2020). Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science. *Symmetry*, 12(4), 495. <https://doi.org/10.3390/sym12040495>
11. Gupta, D., & Rani, R. (2018). A study of big data evolution and research challenges. *Journal of Information Science*, 016555151878988. <https://doi.org/10.1177/0165551518789880>
12. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. A bibliometric approach to tracking big data research trends. *J Big Data* 4, 30 (2017). <https://doi.org/10.1186/s40537-017-0088-1>

13. Liu, X., Sun, R., Wang, S. and Wu, Y.J. (2019). The research landscape of big data: a bibliometric analysis, *Library Hi Tech*, Vol. 38 No. 2, pp. 367-384. <https://doi.org/10.1108/LHT-01-2019-0024>
14. Raban, D.R., Gordon, A. The evolution of data science and big data research: A bibliometric analysis. *Scientometrics* 122, 1563–1581 (2020). <https://doi.org/10.1007/s11192-020-03371-2>
15. Bidosola, I., Garechana, G., Zarrabeitia, E. et al. Characterization of strategic emerging technologies: the case of big data. *Cent Eur J Oper Res* 28, 45–60 (2020). <https://doi.org/10.1007/s10100-018-0597-9>
16. Peng, Y., Shi, J., Fantinato, M. et al. A study on the author collaboration network in big data. *Inf Syst Front* 19, 1329–1342 (2017). <https://doi.org/10.1007/s10796-017-9771-1>
17. Parlina A, Ramli K, Murfi H. Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. *Information*. 2020; 11(2):69. <https://doi.org/10.3390/info11020069>
18. Ting-Peng Liang, Yu-Hsi Liu. Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, Volume 111, 2018, Pages 2-10. <https://doi.org/10.1016/j.eswa.2018.05.018>
19. Furlan, Patricia Kuzmenko; Laurindo, Fernando Jose Barbin (2017). Epistemological grouping of published articles on big data analytics. *TransInformacao*, 29(1), P. 91-100. <https://doi.org/10.1590/2318-08892017000100009>
20. Saheb, T., Saheb, T. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. *J Big Data* 7, 12 (2020). <https://doi.org/10.1186/s40537-020-00287-9>
21. Ahmad, I., Ahmed, G., Shah, S.A.A. et al. A decade of big data literature: analysis of trends in light of bibliometrics. *J Supercomput* 76, 3555–3571 (2020). <https://doi.org/10.1007/s11227-018-2714-x>
22. Gui, B., Liu, Y., Bai, X., & Zhang, J. (2017). Longitudinal Patent Analysis for Big Data Technology. 2017 Portland International Conference on Management of Engineering and Technology (PICMET). doi:10.23919/picmet.2017.81254
23. Шапошник С.Б. Международное научное сотрудничество и публикационная активность российских ученых в Computer Science в 1993-2017 гг.: междисциплинарный и межстрановой сравнительный анализ // Информационное общество, 2018, №6, С. 39-45.
24. World Economic Outlook Database: April 2021. URL: <https://www.imf.org/en/Publications/WEO/weo-database/2021/April/weo-report?c=512,914,612,614,311,213,911,314,193,122,912,313,419,513,316,913,124,339,638,514,218,963,616,223,516,918,748,618,624,522,622,156,626,628,228,924,233,632,636,634,238,662,960,423,935,128,611,321,243,248,469,253,642,643,939,734,644,819,172,132,646,648,915,134,652,174,328,258,656,654,336,263,268,532,944,176,534,536,429,433,178,436,136,343,158,439,916,664,826,542,967,443,917,544,941,446,666,668,672,946,137,546,674,676,548,556,678,181,867,682,684,273,868,921,948,943,686,688,518,728,836,558,138,196,278,692,694,962,142,449,564,565,283,853,288,293,566,964,182,359,453,968,922,714,862,135,716,456,722,942,718,724,576,936,961,813,726,199,733,184,524,361,362,364,732,366,144,146,463,528,923,738,578,537,742,866,369,744,186,925,869,746,926,466,112,111,298,927,846,299,582,487,474,754,698,&s=PPPPC,&sy=2016&ey=2020&ssm=0&scsm=1&scs=0&ssd=1&ssc=0&sic=0&sort=countr y&ds=.&br=1> (дата обращения 15.06.2021)

RESEARCH AND DEVELOPMENT IN BIG DATA

Hohlov, Yuri Evgenyevich

*Candidate of physical and mathematical sciences, associate professor
Institute of the Information Society, chairman of the Board of directors
Plekhanov Russian University of Economics, IIS-Based Digital economy department, scientific advisor
Moscow, Russian Federation
yuri.hohlov@iis.ru*

Shaposhnik, Sergei Borisovich

*Karelian Scientific Center of the Russian Academy of Sciences, OKNI, Laboratory of digital technologies for regional development, senior researcher
Petrozavodsk, Russian Federation
sergei.shaposhnik@gmail.com*

Abstract

A conceptual framework and a set of indicators have been developed for monitoring R&D as a factor in the development and use of big data technologies. The framework has three main components: national scale of research and development, level of research and development in comparison with other countries, and international cooperation as well as cooperation between academia and business. The calculation of indicators and a comparative analysis of the results obtained with other countries were carried out.

Keywords

big data; big data technologies; research and development; publication activity; patent activity; scientometric indicators; monitoring and evaluation

References

1. GOST R 57194.1 – 2916 Transfer tekhnologiy. Obshchiye polozheniya.
2. GOST R 58048 – 2017 Transfer tekhnologiy. Metodicheskiye ukazaniya po otsenke urovnya zrelosti tekhnologiy.
3. OECD (2007) Revised Field of Science and Technology (FOS) Classification in the Frascati Manual. URL: <http://www.oecd.org/science/innovationinsciencetechnologyandindustry/38235147.pdf> (accessed on 15.06.2021)
4. OECD Guide to Measuring the Information Society 2011. <https://doi.org/10.1787/9789264113541-en>
5. OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris. <https://doi.org/10.1787/9789264239012-en>
6. Web of Science. Clarivate Analytics. URL: <https://csi.webofknowledge.com/CSI/exit.do?Func=Exit&SID=C1LIQAIG3E5mZMtoMhE> (data obrashcheniya 15.06.2021)
7. Cornell University, INSEAD, and WIPO (2020). The Global Innovation Index 2020: Who Will Finance Innovation? URL: <https://www.globalinnovationindex.org/> (accessed on 15.06.2021)
8. Daniel Zhang, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault, “The AI Index 2021 Annual Report” AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021. URL: <https://aiindex.stanford.edu/report/> (accessed on 15.06.2021)
9. Patentscope – WIPO. URL: <https://patentscope.wipo.int/> (accessed on 15.06.2021)
10. López Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A.-J., & Parra-González, M. E. (2020). Machine Learning and Big Data in the Impact Literature. A Bibliometric Review with Scientific Mapping in Web of Science. *Symmetry*, 12(4), 495. <https://doi.org/10.3390/sym12040495>
11. Gupta, D., & Rani, R. (2018). A study of big data evolution and research challenges. *Journal of Information Science*, 016555151878988. <https://doi.org/10.1177/0165551518789880>

12. Kalantari, A., Kamsin, A., Kamaruddin, H.S. et al. A bibliometric approach to tracking big data research trends. *J Big Data* 4, 30 (2017). <https://doi.org/10.1186/s40537-017-0088-1>
13. Liu, X., Sun, R., Wang, S. and Wu, Y.J. (2019). The research landscape of big data: a bibliometric analysis, *Library Hi Tech*, Vol. 38 No. 2, pp. 367-384. <https://doi.org/10.1108/LHT-01-2019-0024>
14. Raban, D.R., Gordon, A. The evolution of data science and big data research: A bibliometric analysis. *Scientometrics* 122, 1563–1581 (2020). <https://doi.org/10.1007/s11192-020-03371-2>
15. Bidosola, I., Garechana, G., Zarrabeitia, E. et al. Characterization of strategic emerging technologies: the case of big data. *Cent Eur J Oper Res* 28, 45–60 (2020). <https://doi.org/10.1007/s10100-018-0597-9>
16. Peng, Y., Shi, J., Fantinato, M. et al. A study on the author collaboration network in big data. *Inf Syst Front* 19, 1329–1342 (2017). <https://doi.org/10.1007/s10796-017-9771-1>
17. Parlina A, Ramli K, Murfi H. Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. *Information*. 2020; 11(2):69. <https://doi.org/10.3390/info11020069>
18. Ting-Peng Liang, Yu-Hsi Liu Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, Volume 111, 2018, Pages 2-10. <https://doi.org/10.1016/j.eswa.2018.05.018>
19. Furlan, Patricia Kuzmenko; Laurindo, Fernando Jose Barbin (2017). Epistemological grouping of published articles on big data analytics. *TransInformacao*, 29(1), P. 91-100. <https://doi.org/10.1590/2318-08892017000100009>
20. Saheb, T., Saheb, T. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. *J Big Data* 7, 12 (2020). <https://doi.org/10.1186/s40537-020-00287-9>
21. Ahmad, I., Ahmed, G., Shah, S.A.A. et al. A decade of big data literature: analysis of trends in light of bibliometrics. *J Supercomput* 76, 3555–3571 (2020). <https://doi.org/10.1007/s11227-018-2714-x>
22. Gui, B., Liu, Y., Bai, X., & Zhang, J. (2017). Longitudinal Patent Analysis for Big Data Technology. 2017 Portland International Conference on Management of Engineering and Technology (PICMET). doi:10.23919/picmet.2017.81254
23. Shaposhnik S.B. Mezhdunarodnoye nauchnoye sotrudnichestvo i publikatsionnaya aktivnost' rossiyskikh uchenykh v Computer Science v 1993-2017 gg.: mezhdistsiplinarnyy i mezhstranovoy sravnitel'nyy analiz // *Informatsionnoye obshchestvo*, 2018, №6, S. 39-45.
24. World Economic Outlook Database: April 2021. URL: <https://www.imf.org/en/Publications/WEO/weo-database/2021/April/weo-report?c=512,914,612,614,311,213,911,314,193,122,912,313,419,513,316,913,124,339,638,514,218,963,616,223,516,918,748,618,624,522,622,156,626,628,228,924,233,632,636,634,238,662,960,423,935,128,611,321,243,248,469,253,642,643,939,734,644,819,172,132,646,648,915,134,652,174,328,258,656,654,336,263,268,532,944,176,534,536,429,433,178,436,136,343,158,439,916,664,826,542,967,443,917,544,941,446,666,668,672,946,137,546,674,676,548,556,678,181,867,682,684,273,868,921,948,943,686,688,518,728,836,558,138,196,278,692,694,962,142,449,564,565,283,853,288,293,566,964,182,359,453,968,922,714,862,135,716,456,722,942,718,724,576,936,961,813,726,199,733,184,524,361,362,364,732,366,144,146,463,528,923,738,578,537,742,866,369,744,186,925,869,746,926,466,112,111,298,927,846,299,582,487,474,754,698,&s=PPPPC,&sy=2016&ey=2020&ssm=0&scsm=1&sc=0&ssd=1&ssc=0&sic=0&sort=country&ds=.&br=1> (accessed on 15.06.2021)