

С. Л. Парфенова, В. Н. Долгова, К. А. Безроднова

*Российский НИИ экономики, политики и права
в научно-технической сфере (Москва)*

И. В. Михайленко

ГПНТБ России

Разработка методики для выявления организаций – лидеров публикационной и патентной активности в России на основе данных Scopus, Web of Science и Derwent

Цель представленного в статье исследования состояла в обосновании методического подхода для выявления организаций – лидеров публикационной и патентной активности на основе первичных машиночитаемых в формате XML данных Scopus, Web of Science и Derwent. В основу положен метод фракционного счёта, т.е. определение веса организации по числу научных статей или выданных патентов пропорционально числу аффилиаций, указанных в них. Методология рейтингования организаций основывалась на опыте построения Лейденского рейтинга (в части применения метода фракционного счёта) и рейтинга Academic Ranking of World Universities (в части расчёта взвешенного показателя на одного исследователя, выбора способа нормирования показателей). Подчёркнуто, что принципиальное отличие предложенной в статье методологии рейтингования организаций от существующих методик состоит в расчёте интегрального показателя по сопоставимым библиометрическим параметрам Scopus и Web of Science. Кроме того, предложен способ построения рейтинга организаций по среднему числу выданных патентов по данным Derwent. Результаты исследования показали, что топы-5 российских организаций – лидеров публикационной активности по данным Scopus и Web of Science в значительной мере пересекаются и относятся к государственному сектору науки и сектору высшего образования. Топ-5 российских организаций – лидеров патентной активности преимущественно состоит из организаций коммерческого сектора науки.

Исследование выполнено за счёт средств государственных заданий Министерства науки и высшего образования Российской Федерации: «Анализ востребованности научных (научно-технических) результатов» (730000Ф.99.1.БВ16АА02001).

Ключевые слова: методология, рейтингование, организации-лидеры, публикационная активность, патентная активность.

SCIENTOMETRICS. BIBLIOMETRICS

UDC 002:51

DOI 10.33186/1027-3689-2019-8-38-57

Svetlana Parfenova, Vladislava Dolgova, Karina Bezrodnova

Russian Research Institute of Economics, Politics and Law in Science and Technology, Moscow, Russia

Irina Mikhailenko

Russian National Public Library for Science and Technology, Moscow, Russia

Developing the method for searching corporate leaders in publication and patent activities in Russia based on the Scopus, Web of Science and Derwent data

The authors aim to substantiate the methodical approach to revealing corporate leaders in publishing and patent activities based on the primary Scopus, Web of Science и Derwent data machine-readable in XML format. The approach is based on the fractional calculus method, i. e. determining the organization's weight by the number of its academic papers or issued patents proportionally to the number of specified affiliates. The organizational ranking method is based on the experience of the Leiden ranking (in using fractional calculus method) and the Academic Ranking of World Universities (in calculating the weighted value per one researcher, and choosing quality rating method). The authors emphasize that the fundamental distinctive feature of the proposed ranking methodology is that the integral index is calculated with the comparable Scopus and Web of Science parameters. For this purpose ranking of organizations is developed based on the average number of issued patents as reported by Derwent. The study proves that the top-5 ratings for Russian organization leading in publication activities, as reported by Scopus and Web of Science, overlap substantially and cover the public sector of science and higher education. The top-5 of Russian organizations leading in patent activities comprise primarily the non-government, commercial sector of science.

The study is accomplished within the framework of the state order of the Ministry of Science and Higher Education of the Russian Federation: "Analysis of the relevance of scientific (scientific and technical) results" (730000Ф.99.1.ББ16АА02001).

Keywords: methodology, ranking, leader organizations, publication activity, patent activity.

Scopus, Web of Science and Derwent databases were selected as information resources containing information on the number of publications of organizations conducting research and development, and the number of patents. Rightholders (Elsevier and Clarivate Analytics) supply raw data in XML format. The possibility of operating with raw data allowed to perform the general codification structure of Scopus, Web of Science and Derwent. Correspondence tables of

codifications used in each database were created. The codification of the Web of Science is chosen as the basic one, since it practically corresponds to the popular OECD codification. As key parameters in determining organizations that are leaders in publishing and patent activity, it is proposed to use five main indicators: 1) the average number of scientific articles per researcher/scientific-pedagogical employee of a scientific organization/university; 2) the average number of scientific articles in journals of the first and second quartile per researcher/scientific and pedagogical employee of a scientific organization / university; 3) the average number of citation of scientific articles of the organization; 4) the average number of citation of scientific articles of the organization in journals of the first and second quartiles; 5) the average number of patents granted per researcher/scientific-pedagogical employee of a scientific organization/university. Analysis of organizations – leaders of publication activity according to Scopus and Web of Science allowed to identify the absolute leader in all scientific fields in the field of "Natural and Exact Sciences" – Moscow State University. The frequent coincidence of the lists of organizations according to the Web of Science and Scopus data provides a good base for evaluations. The first positions of ratings of patent activity are mainly occupied by commercial companies (Rosatom, the Space Corporation Energia, etc.); scientific organizations and universities, in the research work of which there is a large proportion of applied research (Moscow State University, All-Russian Scientific Research Institute of Petrochemical Processes, St. Petersburg Mining University, etc.).

В первой части исследования [1] был проведён анализ национальных и локальных систем рейтингования организаций, выполняющих научные исследования и разработки.

Результаты анализа известных методологий рейтингования организаций положены в основу методического подхода, применимого для выбора организаций – лидеров публикационной и патентной активности. Так же, как и методология Лейденского рейтинга, предлагаемый метод рейтингования организаций будет основан на методе фракционного счёта, суть которого – определять вес публикаций и вес патентов, относящихся к конкретной организации, пропорционально числу аффилиаций, указанных в публикациях и патентах.

Подробно метод фракционного счёта и способы расчёта на его основе числа публикаций организации рассмотрены в статье «Публикационный ландшафт российской науки» [2]. Основное достоинство этого метода – «устранение проблемы искажения показателей публикационной активности» организации способом дедубликации. Модель расчёта числа публикаций

организации предполагает: выявление публикаций с аффилиацией искомой организации за определённый период; расчёт доли публикаций, относящихся к анализируемой организации, пропорционально числу аффилиаций, указанных в публикации; расчёт общего числа публикаций организаций путём суммирования долей каждой публикации с аффилиацией искомой организации.

Аналогичный подход предложено применить к расчёту числа патентов организации. С точки зрения авторов статьи, число публикаций и число патентов, рассчитанные методом фракционного счёта, будут более корректно отражать количественные показатели организации. Дальнейшее развитие методики, возможно, потребует от авторов поиска общего знаменателя для публикаций и патентов. Пока эта задача не решена.

Методика проведения исследования

В качестве информационных ресурсов, содержащих сведения о числе публикаций организаций, выполняющих исследования и разработки, и числе патентов, выбраны базы данных *Scopus*, *Web of Science* и *Derwent*. Компании-правообладатели (*Elsevier* и *Clarivate Analytics*) в рамках лицензионных соглашений поставляют первичные – «сырые» – данные в формате XML [3, 4] в Министерство науки и высшего образования Российской Федерации. Отличие передаваемой информации от имеющейся в интерфейсе БД *Scopus*, *Web of Science* и *Derwent* состоит в возможности проведения гибких инвариантных исследований на основе «сырого» массива данных в машиночитаемом формате.

Возможность оперирования первичными данными позволила привести к общей структуре кодификации *Scopus*, *Web of Science* и *Derwent*. Были созданы таблицы соответствия кодификаций, используемых в каждой БД. Как базовая выбрана кодификация *Web of Science*, так как она практически соответствует популярной кодификации OECD, а специалисты считают, что она наиболее полно и сбалансированно отражает актуальные научные направления [5].

В качестве ключевых параметров при определении организаций – лидеров публикационной и патентной активности предложено использовать пять основных показателей:

- 1) среднее число научных статей на одного исследователя/научно-педагогического работника научной организации/вуза;
- 2) среднее число научных статей в журналах первого и второго квартилей на одного исследователя/научно-педагогического работника научной организации/вуза;
- 3) среднее число цитирования научных статей организации;
- 4) среднее число цитирования научных статей организации в журналах первого и второго квартилей;

5) среднее число выданных патентов на одного исследователя/научно-педагогического работника научной организации/вуза.

При выборе показателей публикационной активности авторы статьи ориентировались на количественные (1, 2 и 4) и качественные (2 и 3) показатели. Показатель «среднее число научных статей в журналах первого и второго квартилей на одного исследователя/научно-педагогического работника научной организации/вуза» характеризует одновременно количество научных статей и их качественный уровень (квартили журналов, в которых опубликованы статьи).

Расчёт среднего числа цитирования научных статей организации (C_{cp}) предложено проводить по формуле:

$$C_{cp} = \frac{CЦ}{N_{cm}},$$

где $CЦ$ – совокупное число ссылок на научные статьи, аффилированные конкретной организацией за конкретный период (единицы); N_{cm} – число научных статей, аффилированных с конкретной организацией за конкретный период (единицы).

Расчёт показателей публикационной активности проводился по типу публикации «Article/научная статья» на основе первичных данных *Web of Science* по следующим индексам научного цитирования: *Science Citation Index Expanded (SCI-EXPANDED)*, *Social Sciences Citation Index (SSCI)*, *Arts & Humanities Citation Index (A&HCI)*, *Emerging Sources Citation Index (ESCI)*.

Чтобы проводить арифметические действия со значениями показателей различных единиц измерения, рекомендовано провести их нормирование, суть которого – привести значения показателей к безразмерной единице измерения. Процесс нормирования предполагал следующие шаги:

построение ряда исходных данных;

выбор максимального значения из ряда исходных данных ($\max(X_{ij})$);

нормирование данных по формуле [6]:

$$\bar{X}_{ij} = \frac{X_{ij}}{\max(X_{ij})},$$

где \bar{X}_{ij} – нормированное значение j -го показателя i -й организации; X_{ij} – значение j -го показателя i -й организации.

В результате нормирования значений показателей организаций из определённой выборки они приведены к безразмерной величине в диапазоне от 0 до 1.

Выявлять организации – лидеров публикационной и патентной активности авторы статьи предлагают по алгоритму, представленному на рис. 1.

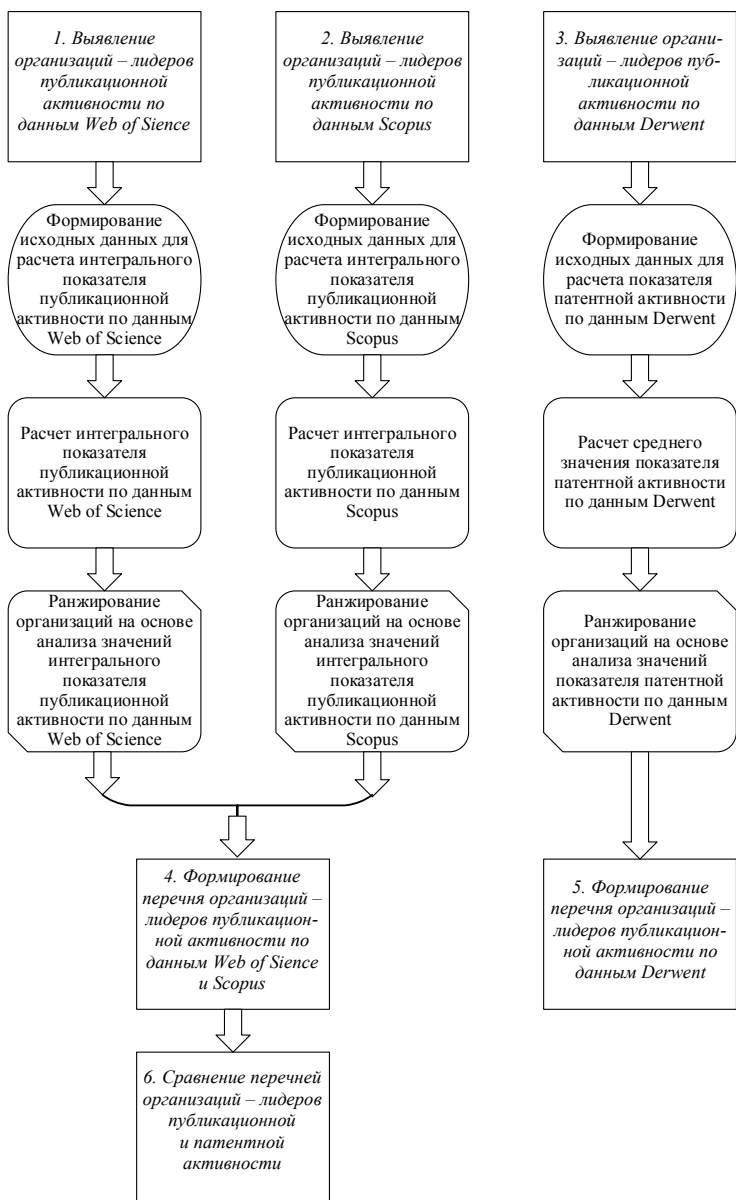


Рис. 1. Алгоритм формирования перечня организаций – лидеров публикационной и патентной активности

Сформулируем порядок расчёта интегрального показателя публикационной активности:

1. Расчёт и суммирование фракций (долей) научных статей, аффилированных с организацией, по каждому научному направлению за год.

2. Расчёт среднего числа научных статей, приходящегося на одного исследователя/научно-педагогического работника научной организации/вуза (N_{cm}^{cp}). Численность исследователей и научно-педагогических работников организаций определена по данным Федеральной системы мониторинга результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы*.

3. Расчёт и суммирование фракций (долей) научных статей (первого и второго квартилей журналов), аффилированных с организацией, по каждому научному направлению за год.

4. Расчёт среднего числа научных статей в журналах первого и второго квартилей, приходящегося на одного исследователя/научно-педагогического работника научной организации/вуза $\left(N \frac{Q1/Q2}{cm} \right)$.

5. Определение совокупного числа ссылок на научные статьи организации по каждому научному направлению за год.

6. Расчёт среднего числа цитирований на одну научную статью организации по каждому научному направлению за год ($СИ_{cp}$).

7. Определение совокупного числа ссылок на научные статьи (первого и второго квартилей) по каждому научному направлению за год.

8. Расчёт среднего числа цитирований на одну научную статью организации по каждому научному направлению за год ($СИ_{cp}^{Q1/Q2}$).

9. Нормирование значений показателей публикационной активности организаций по каждому научному направлению.

10. Расчёт интегрального показателя публикационной активности организации по конкретному научному направлению за год как сумма нормированных значений показателей:

* Федеральная система мониторинга результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы. Режим доступа: <http://www.sciencemon.ru/>

$$\text{ИП} = \frac{\sum_{ni=1}^n \left(N_{cni}^{cp./норм.} + N_{cni}^{\frac{Q1}{Q2}/норм.} + CЦ_{cni}^{норм.} + CЦ_{cni}^{\frac{Q1}{Q2}/норм.} \right)}{n},$$

где n – период расчёта интегрального показателя (годы); $N_{cni}^{cp./норм.}$ – нормированное значение среднего числа научных статей, приходящегося на одного исследователя/научно-педагогического работника научной организации/вуза

(единицы); $N_{cni}^{\frac{Q1}{Q2}/норм.}$ – нормированное значение среднего числа научных статей в журналах первого и второго квартилей, приходящегося на одного исследователя/научно-педагогического работника научной организации/вуза (единицы); $CЦ_{cni}^{норм.}$ – нормированное значение среднего числа цитирований на одну научную статью организации по каждому научному направлению за

год (единицы); $CЦ_{cni}^{\frac{Q1}{Q2}/норм.}$ – нормированное значение среднего числа цитирований на одну научную статью организации по каждому научному направлению за год (единицы).

Расчёт интегрального показателя публикационной активности рекомендуется проводить за период не менее 5 лет, что гарантирует стабильность его значения.

На основе анализа значений интегральных показателей публикационной активности проводится ранжирование организаций. В зависимости от целей исследования составляются топ-20, топ-10 или топ-5 организаций – лидеров публикационной активности. Сравнение организаций – лидеров публикационной активности по данным *Web of Science* и *Scopus* в рамках одного тематического направления позволит повысить объективность результатов исследования.

Преимущества предложенного метода рейтингования по показателям публикационной и патентной активности в сравнении с существующими методиками заключаются в следующем:

сопоставимые показатели публикационной активности рассчитываются и по данным *Scopus*, и по данным *Web of Science*, благодаря чему организации оцениваются на основе данных двух глобальных индексов научного цитирования;

рейтинги организаций составляются по одной системе кодификации, что позволяет проводить их сравнительный анализ по одним и тем же направлениям;

количественные показатели организаций рассчитываются методом фракционного счёта, поэтому нет дублирования их научных статей и показана «чистая» научная производительность организаций;

показатели публикационной активности организаций рассчитываются за пятилетний период, что гарантирует стабильность их значений и устраняет периодические «всплески» или «провалы».

Выявлять организации – лидеров патентной активности мы предлагаем на основе расчёта методом фракционного счёта числа выданных патентов (по данным *Derwent*). Определение среднего числа выданных патентов по организациям (методом средней арифметической) за определённый период (не менее 5 лет) позволит, как и в случае с интегральным показателем публикационной активности, оценить стабильность показателя. На основе среднего числа выданных патентов ранжируются организации. В зависимости от целей исследования составляются топ-20, топ-10 или топ-5 организаций – лидеров патентной активности.

Результаты исследования

Апробация методического подхода к выявлению организаций – лидеров публикационной и патентной активности на основе метода фракционного счёта проведена по данным *Scopus*, *Web of Science* и *Derwent*.

За период 2013–2017 гг. по каждому научному направлению в области «Естественные и точные науки» на основе первичных – «сырых» – данных, поступающих из компаний *Elsevier* и *Clarivate Analytics*, рассчитаны: значения интегральных показателей публикационной активности и составлены топ-5 российских организаций – лидеров публикационной активности (рис. 2, 3); средние значения показателей патентной активности; составлен топ-5 организаций – лидеров патентной активности (рис. 4).

Анализ организаций – лидеров публикационной активности по данным *Scopus* и *Web of Science* позволил выявить абсолютного лидера по всем научным направлениям в области «Естественные и точные науки» – это Московский государственный университет им. М. В. Ломоносова.

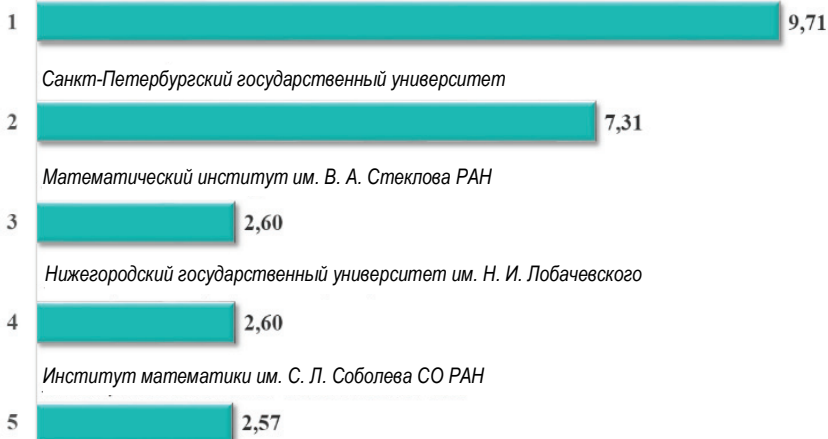
Интересно отметить следующий расклад организаций в разрезе научных направлений:

стоппроцентное совпадение наименований и позиций организаций в рейтингах топ-5 по данным *Web of Science* и *Scopus* в научном направлении «Математика»;

совпадение наименований (три и более) при расхождении по позициям организаций в рейтингах топ-5 по данным *Web of Science* и *Scopus* в научных направлениях «Физика и астрономия», «Химия», «Науки о Земле», «Биология».

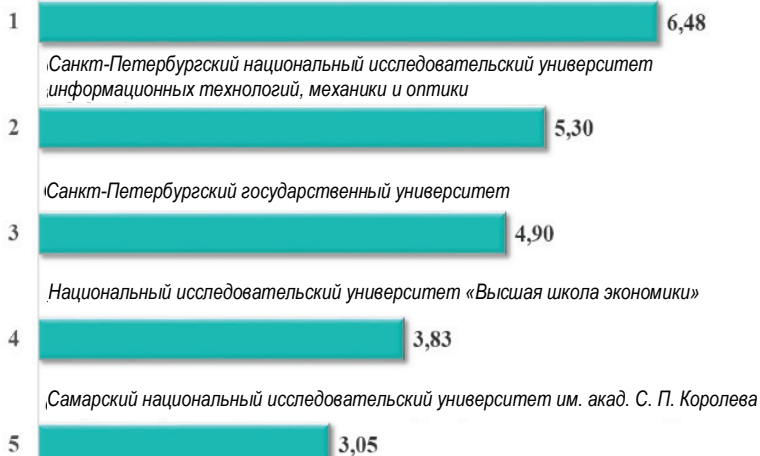
Математика

Московский государственный университет им. М. В. Ломоносова

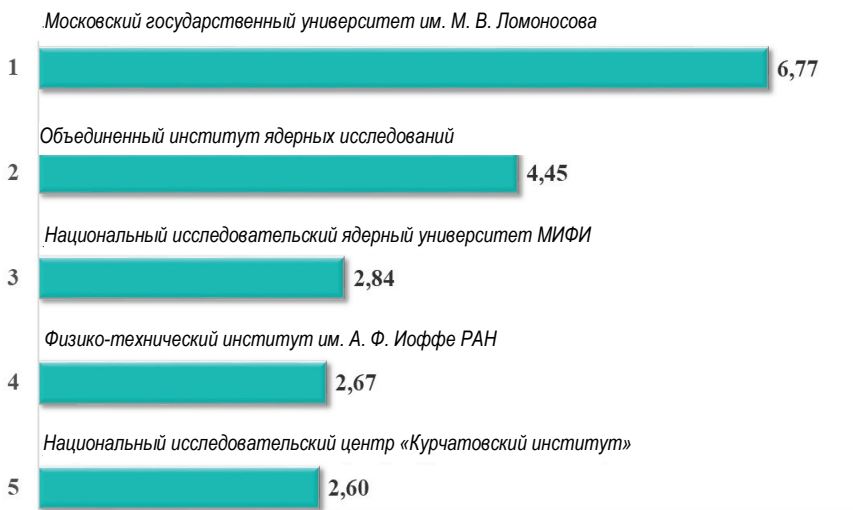


Компьютерные науки

Московский государственный университет им. М. В. Ломоносова



Физика и астрономия



Химия



Науки о Земле



Биология

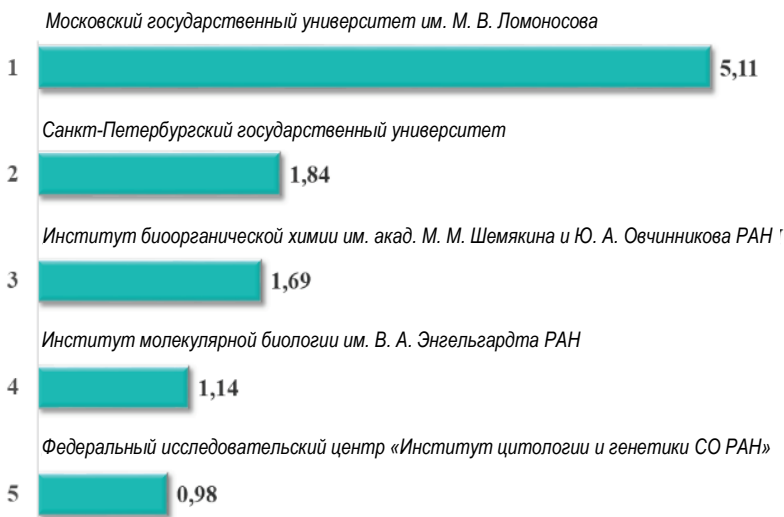
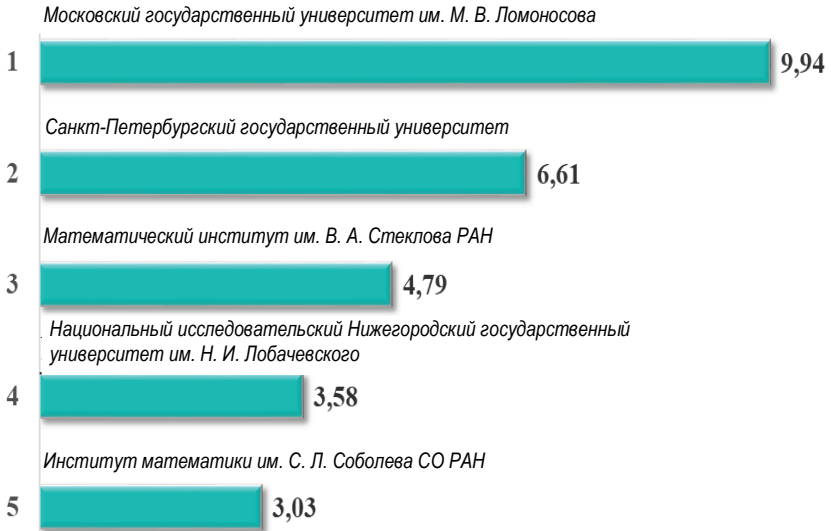
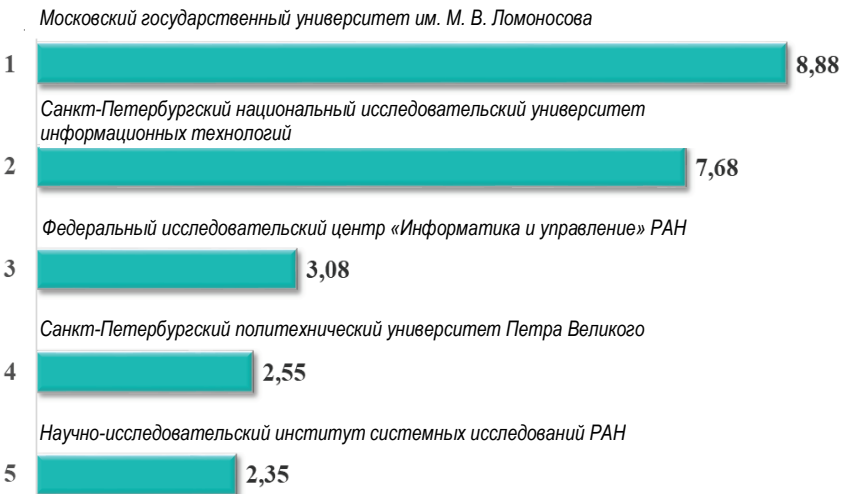


Рис. 2. Организации – лидеры публикационной активности в области «Естественные и точные науки» (по данным Scopus, дата обращения: 26.09.2018)

Математика



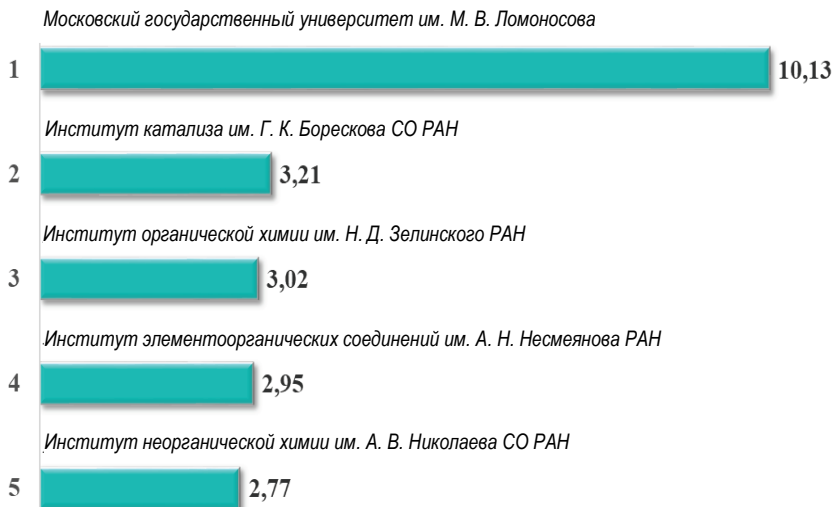
Компьютерные науки



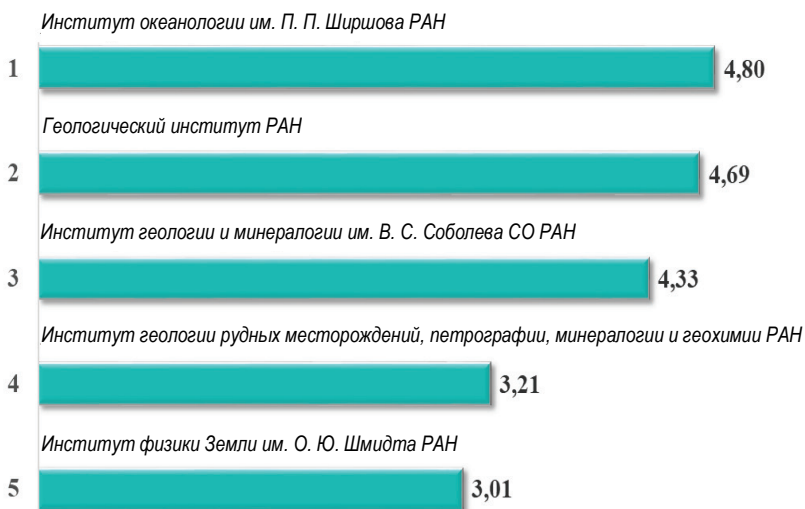
Физика и астрономия



Химия



Науки о Земле



Биология

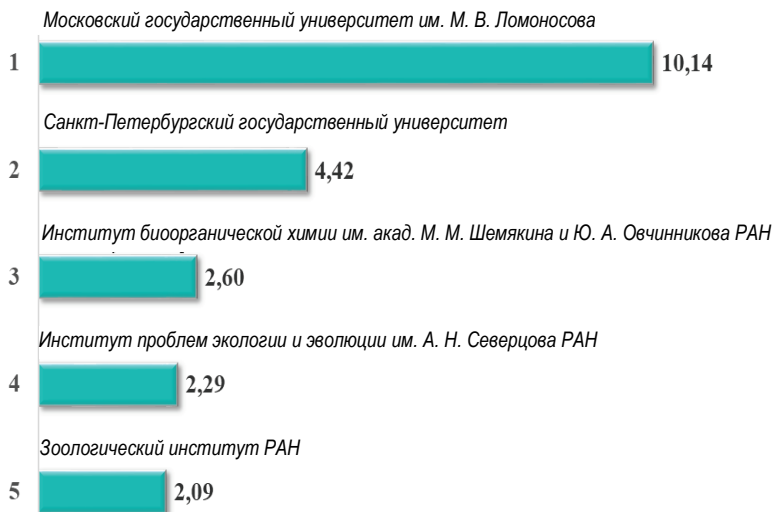
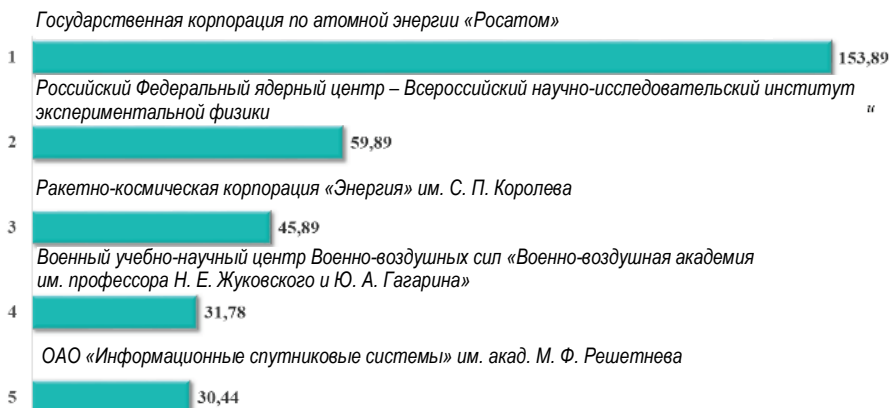


Рис. 3. Организации – лидеры публикационной активности в области «Естественные и точные науки» (по данным *Web of Science*, дата обращения: 26.09.2018)

Компьютерные науки

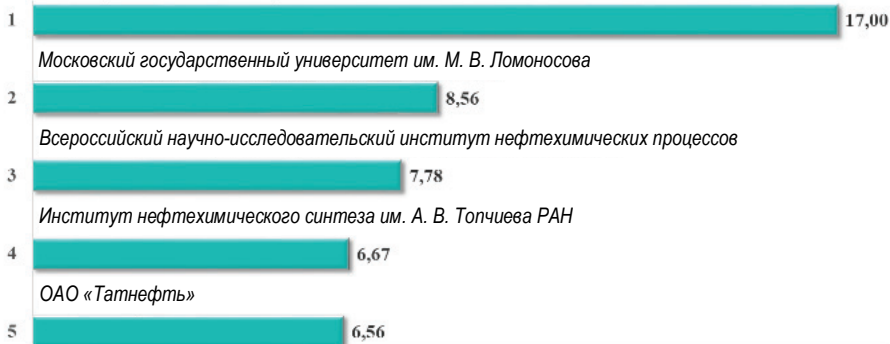


Физика и астрономия



Химия

Институт катализа им. Г. К. Борескова СО РАН



Московский государственный университет им. М. В. Ломоносова

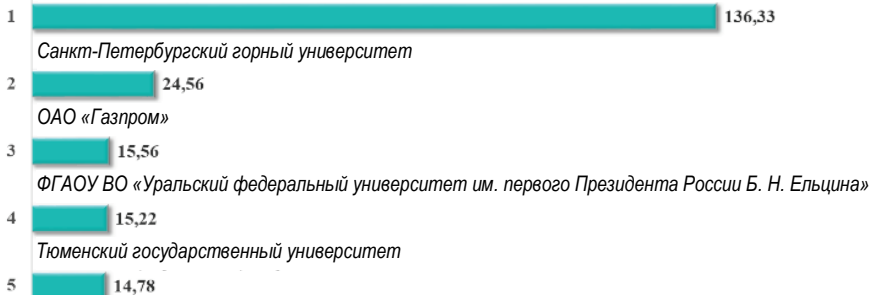
Всероссийский научно-исследовательский институт нефтехимических процессов

Институт нефтехимического синтеза им. А. В. Топчиева РАН

ОАО «Татнефть»

Науки о Земле

ОАО «Татнефть»



Санкт-Петербургский горный университет

ОАО «Газпром»

ФГАОУ ВО «Уральский федеральный университет им. первого Президента России Б. Н. Ельцина»

Тюменский государственный университет

Биология

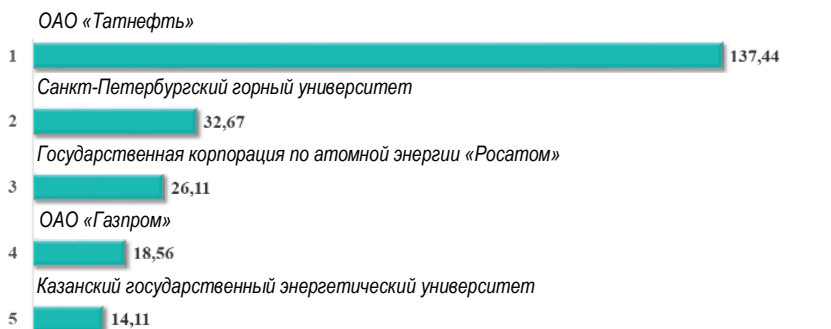


Рис. 4. Организации – лидеры патентной активности в области «Естественные и точные науки» (по данным *Derwent*, дата обращения: 26.09.2018)

Частое совпадение перечней организаций – лидеров публикационной активности в разрезе научных направлений по данным *Web of Science* и *Scopus* позволило сделать вывод: на топовые позиции в рейтингах публикационной активности глобальных индексов научного цитирования попадают практически одни и те же организации, занимающиеся фундаментальными исследованиями. Чего нельзя сказать про организации – лидеры патентной активности. Первые позиции рейтингов патентной активности в сопоставимых научных направлениях занимают в основном коммерческие компании; организации, специализирующиеся на прикладных исследованиях (например, госкорпорация «Росатом», ракетно-космическая корпорация «Энергия» им. С. П. Королёва и др.); научные организации и вузы, в научно-исследовательской работе которых велика доля прикладных исследований (МГУ им. М. В. Ломоносова, Всероссийский научно-исследовательский институт нефтехимических процессов, Санкт-Петербургский горный университет и др.).

Апробация методического подхода для выявления организаций – лидеров публикационной и патентной активности на основе метода фракционного счёта по данным *Scopus*, *Web of Science* и *Derwent* подтвердила адекватность и релевантность полученных данных.

Результаты исследования, содержащие перечни топ-5 по данным *Scopus* и по данным *Web of Science* в разрезе научных направлений шести областей наук, опубликованы в «Дайджете показателей публикационной активности российских исследователей по данным *Web of Science* и *Scopus*» (декабрь 2018 г.) [7].

Выводы

В статье представлена методика выявления организаций – лидеров публикационной и патентной активности по данным *Scopus*, *Web of Science* и *Derwent*. Она основана на методе фракционного счёта, который позволяет исключить дублирование научных статей и патентов в разрезе организаций, выполняющих исследования и разработки, и повысить корректность используемых в процессе проведения исследования данных.

Выбор организаций – лидеров публикационной активности основан на оценке интегрального показателя публикационной активности за 5 лет по данным *Scopus* и *Web of Science*, организаций – лидеров патентной активности – на оценке среднего значения показателя патентной активности за 5 лет по данным *Derwent*.

В соответствии с представленным в статье алгоритмом сформированы перечни топ-5 организаций – лидеров публикационной и патентной активности в разрезе научных направлений области «Естественные и точные науки». Анализ полученных результатов подтвердил их адекватность и релевантность.

Предложенный инструментарий предназначен для определения центров компетенций или организаций – лидеров по направлениям научных исследований в области фундаментальной и прикладной науки.

СПИСОК ИСТОЧНИКОВ

1. **Парфенова С. Л.** Анализ международных и российских рейтингов научно-исследовательских организаций / С. Л. Парфенова, В. Н. Долгова, К. А. Безроднова, И. В. Михайленко // Науч. и техн. б-ки. – 2019. – № 7. – С. 14–24.

Parfenova S. L. Analiz mezhdunarodnyh i rossiyskih reytingov nauchno-issledovatel'skikh organizatsiy / S. L. Parfenova, V. N. Dolgova, K. A. Bezrodnova, I. V. Mihaylenko // Nauch. i tehn. b-ki. – 2019. – № 7. – S. 14–24.

2. **Парфенова С. Л.** Публикационный ландшафт российской науки / С. Л. Парфенова, Е. Г. Гришакина, Д. В. Золотарев, В. В. Богатов // Наука. Инновации. Образование. – 2017. – № 23. – С. 53–79.

Parfenova S. L. Publikatsionnyy landshaft rossiyskoy nauki / S. L. Parfenova, E. G. Grishakina, D. V. Zolotarev, V. V. Bogatov // Nauka. Innovatsii. Obrazovanie. – 2017. – № 23. – S. 53–79.

3. **Scopus Custom Data** [Электронный ресурс] // Elsevier B. V. – 2014. – 08. – URL: https://www.elsevier.com/_data/assets/pdf_file/0009/94878/Scopus-Custom-Data.pdf. – 29.04.2019.

4. **Web of Science Raw Data (XML)** // Clarivate Analytics : User guide for Web of Science raw data. – 2018. – Dec. 13. – URL: <https://clarivate.libguides.com/rawdata>. – 29.04.2019.

5. **Large Scale** Analysis of the Accuracy of the Journal Classification Systems of Web of Science and Scopus / Qi Wang, Ludo Waltman // Journal of Informetrics. – 2016. – May. – Vol. 10. – Iss. 2. – URL: <https://arxiv.org/abs/1511.00735>. – 29.04.2019.

6. **Дмитриев Г. И.** Использование методики ранжирования для оценки результативности научной деятельности вузов / Дмитриев Г. И., Законников Е. А., Мейев В. А., Воронов Ю. В. // Дискурс. – 2016. – № 6. – С. 38–45.

Dmitriev G. I. *Ispolzovanie metodiki ranzhirovaniya dlya otsenki rezul'tativnosti nauchnoy deyatelnosti vuzov / Dmitriev G. I., Zakonnikov E. A., Meyev V. A., Voronov Yu. V. // Diskurs. – 2016. – № 6. – S. 38–45.*

7. **Дайджест** показателей публикационной активности российских исследователей по данным Web of Science/Scopus (декабрь 2018) / Авт.-сост.: Г. В. Трубников, М. Ю. Романовский, И. Е. Ильина, С. Л. Парфенова, В. Н. Долгова и др. – Москва : Буки Веди, 2018. – 56 с. – ISBN 978-5-9909396-9-1.

Daydzhest *pokazateley publikatsionnoy aktivnosti rossiyskikh issledovateley po dannym Web of Science/Scopus (dekabr 2018) / Avt.-sost.: G. V. Trubnikov, M. Yu. Romanovskiy, I. E. Ilyina, S. L. Parfenova, V. N. Dolgova i dr. – Moskva : Buki Vedi, 2018. – 56 s. – ISBN 978-5-9909396-9-1*

Svetlana Parfenova, *Cand. Sc. (Economics), Deputy Director, Head, Socio-Economical Development in Science and Technology Department, Russian Research Institute of Economics, Politics and Law in Science and Technology;*

parfyonova.s.l@yandex.ru

50A, Build. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Vladislava Dolgova, *Cand. Sc. (Economics), Associate Professor, Head of Socio-Economical Development in Science and Technology Sector, Russian Research Institute of Economics, Politics and Law in Science and Technology;*

vlada8@bk.ru

50A, Build. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Karina Bezrodnova, *Researcher Engineer, Russian Research Institute of Economics, Politics and Law in Science and Technology;*

__karina.__@mail.ru

50A, Build. 6, Zemlyanoy Val st., 105064 Moscow, Russia

Irina Mikhailenko, *Head, Scientometrics and Bibliometrics Research Department, Russian National Public Library for Science and Technology;*

irimikhaylenko@gmail.com

17, 3rd Khoroshevskaya st., 123298 Moscow, Russia