

Инструменты анализа научно-технологических заделов России*

И.А. ТИХОМИРОВ, Н.В. ТОГАНОВА, М.И. АНАНЬЕВА

Аннотация. Выявление приоритетных направлений исследований - актуальная практическая задача для научных фондов и иных организаций, финансирующих российскую науку. В статье представлен обзор западных и российских наукометрических инструментов пригодных для анализа научно-технологических заделов. Авторы приходят к выводу, что использование исключительно западных инструментов для оценки научно-технологических заделов России невозможно, необходимо развивать отечественные. Приведен пример анализа состояния дел в России по направлению «фотоника» при помощи системы Exactus Expert, демонстрирующий как интеграция на одной платформе существующих российских ресурсов, находящихся в открытом доступе, может обеспечить анализ задела по отдельно взятому направлению. В заключении сделаны выводы о перспективности отечественного инструментария и использования наукометрических систем.

Ключевые слова: анализ заделов, управление наукой, инновации, наукометрия, анализ полных текстов, приоритизация исследований.

Введение

В настоящее время в виду ограниченности ресурсов на исследования и разработки ни одна страна в мире не может вести полномасштабные исследования сразу по всем научным направлениям, как следствие, требуется их приоритизация [5]. Перед отдельными российскими фондами и государством в целом стоит задача выявить перспективные направления, которые могли бы обеспечить развитие конкурентоспособных высокотехнологических отраслей [6]. В тоже время мы нередко наблюдаем подмену реально достижимых приоритетов, сформулированных в результате анализа научно-технических заделов именно нашей страны, на приоритеты, списанные с зарубежных стратегических документов. Это приводит к тому, что традиционно сильным в России направлениям науки уделяется недостаточно внимания, а доверие научного сообщества к отечественным стратегическим документам подрывается.

Ситуация во многом объясняется тем, что степень информированности лиц, принимающих решение, о реальном состоянии научных заделов недостаточна. Привлекаемые к оценке заделов эксперты не всегда обладают полнотой знания в рассматриваемом поле современных научных исследований и его ретроспективе. Неверно сформулированные темы конкурсов в российских научных фондах оборачиваются слишком большим потоком заявок, качественно обработать которые конкурсная комиссия не в состоянии. Или же, на-

против, по казавшейся перспективной теме заявки не подаются, поскольку научный задел по этой тематике в России отсутствует, или тема конкурса сформулирована настолько узко, что подать заявку мог бы лишь один научный коллектив.

Другая разновидность той же проблемы – активное финансирование одного направления исследований и невнимание к тесно связанным с ним, что препятствует внедрению результатов на прикладном уровне. Именно это происходит с сектором нанотехнологий в России, финансирование которого в последние годы осуществлялось в сопоставимых с западными странами объемах, но решительного улучшения ситуации с точки зрения экономической отдачи не последовало. Последнего можно было бы избежать, если бы предварительно был проведен анализ междисциплинарного взаимодействия и были бы выявлены нуждающиеся в столь же пристальном внимании смежные с нанотехнологиями отрасли.

Инструменты наукометрического анализа.

Научным заделом страны принято считать совокупность потенциальных инноваций, которые при определенных условиях могут обеспечить создание новых технологий, моделей и др. для решения принципиально новых задач. Под научно-технологическим заделом в свою очередь принято понимать уже имеющиеся результаты интеллектуальной деятельности в сфере науки и техники [1]. Распространенный способ оценки научно-технологических заделов – анализ патентной активности по разным

* Работа выполнена при финансовой поддержке РФФИ (грант № 14-29-05090)

отраслям [3]. Однако для анализа научных заделов недостаточно одного анализа патентов, необходим учет научных публикаций. Количество научных документов исчисляется десятками миллионов; тщательный их анализ может позволить выявить новые научные направления, которые появляются и набирают популярность, в то время как другие идут на спад. В то же время экспертам сложно ориентироваться в большом массиве данных без использования специализированных информационных систем. Количество и разнообразие таких систем в последние годы выросло, как и их влияние на научный сектор – они подталкивают инновационный процесс, поскольку позволяют быстрее реагировать на меняющуюся реальность [9].

Существуют отечественные и зарубежные информационно-аналитические системы, на которые эксперты и лица, принимающие решения, могут опираться при анализе научных заделов. Эти системы можно разделить на базы цитирования, электронные научные библиотеки, базы патентов и научные поисковые системы.

Базы цитирования работают не с полными текстами документов, а с библиографическими данными: аннотациями, ключевыми словами и списками литературы. Это позволяет оценивать значимость научных работ и вклад отдельных ученых, строить профили организаций и сравнивать их между собой. К базам цитирования относятся: зарубежные – Web of Science¹, Scopus², отечественным аналогом является РИНЦ на основе eLIBRARY³. Scopus отличается от WoS большим международным охватом и количеством индексируемых журналов, а также расширенными функциями поиска⁴. Преимущество WoS в том, что в нем проиндексированы журналы с 1900 г. РИНЦ все еще находится на этапе становления: в нем наиболее полно представлены лишь публикации последнего десятилетия, а учет цитирования не достаточно тщательный, множество представленных в нем записей внесено без списка литературы или же библиографические ссылки просто не связаны друг с другом и не могут быть учтены автоматически (см. табл. 1).

Электронные научные библиотеки в отличие от баз цитирования, как правило, специализируются на конкретных областях исследования: Arxiv⁵ содержит материалы по физике, математике, компьютерным наукам, биологии и статистике; PubMed⁶ –

по медико-биологическим наукам; IEEE Xplore⁷ – по компьютерным наукам, электротехнике и электронике; NGC⁸ – по руководствам доказательной медицины. Однако существуют и другие различия. Так, PubMed и IEEE Xplore осуществляют поиск похожих документов на основе анализа списка литературы. Arxiv отличается плоской классификацией (а не иерархической), что делает научный поиск более удобным [12]; также система позволяет составлять поисковый запрос с помощью тезауруса и просматривать опубликованные документы по той или иной теме. Главным достоинством Arxiv является то, что он содержит исходные версии работ (препринты), которые были доступны в сети до того, как попали в журналы с ограниченным доступом [11].

Базы патентов и патентные поисковые машины характеризуются своей привязкой к региону, которую в последние годы интегральные проекты пытаются преодолеть. Так, OPS (Open Patent Service)⁹ – ресурс Европейского Патентного Ведомства¹⁰, насчитывающий более 70 млн. документов со всего мира, который поддерживает поисковые запросы с использованием логических операторов и машинный перевод документов. USPTO¹¹ – бесплатная база по полным текстам всех американских патентов с 1790 г. FPO (FreePatentsOnline)¹² – автоматически пополняющаяся база американских патентов со следующими функциями: построение карты патентов, патентной активности в разных областях науки, автоматической классификацией патентов и пр. Orbit¹³ – система анализа патентов с богатым функционалом, таким как создание пользовательских коллекций, автоматическое построение реферата текста, аналитических диаграмм и пр. [8]. К поисковым машинам по патентам относятся Google Patents¹⁴ и FPO, которые агрегируют документы из разных баз, таких как Патентное ведомство соединенных штатов (the United States Patent and Trademark Office (USPTO)), Европейское патентное ведомство (the European Patent Office (EPO)), и мировая организация по защите интеллектуальной собственности (the World Intellectual Property Organization (WIPO)). Thomson Innovation¹⁵ предоставляет полнотекстовый поиск и многоаспектный анализ по документам из Derwent World

¹ <http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science.html>

² <http://www.elsevier.com/solutions/scopus>

³ <http://elibrary.ru>

⁴ <http://libguides.hsl.washington.edu/c.php?g=99232&p=642081>

⁵ <http://arxiv.org>

⁶ <https://www.ncbi.nlm.nih.gov/pubmed>

⁷ <http://ieeexplore.ieee.org>

⁸ <http://www.guideline.gov>

⁹ <http://www.epo.org/searching/free/ops.html>

¹⁰ <http://www.epo.org>

¹¹ <http://www.uspto.gov/patents-application-process/search-patents>

¹² <http://www.freepatentsonline.com>

¹³ http://www.questel.com/customersupport/userdoc/orbit_documentation/Patent_module_user_guide_Xpert_Search_Questel.pdf

¹⁴ <https://www.google.com/?tbnm=pts>

¹⁵ <http://info.thomsoninnovation.com/en/features/analyze>

Таблица 1.

Сравнительные характеристики информационно-аналитических систем¹

ЗАРУБЕЖНЫЕ	Полнотекстовый поиск	Поиск по метаданным	Система классификации	Поиск похожих документов	Поиск заимствований	Автоматическое построение реферата текста	Автоматическое выделение ключевых слов	Создание пользовательских коллекций	Аналитические функции	Поддержка анализа на русском языке	Поддержка анализа на английском языке
Web of Science		+	П	-	-	-	-	+	+	-	+
Scopus		+	П	+	-	-	-	+	+	-	+
Arxiv	+	+	П	-	-	-	-	-	-	-	+
PubMed	+	+	MeSH	+	-	-	-	-	-	-	+
IEEE Xplore	+	+	П	+	-	-	-	+	-	-	+
NGC	+	+	MeSH	-	-	-	-	-	-	-	+
OPS	+	+	IPC	-	-	-	-	-	-	-	+
USPTO	+	+	IPC+ своя	-	-	-	-	-	-	-	+
Orbit	+	+	IPC, ECLA, ICO, PCL	+	-	-	+	+	+	-	+
Google Patent	+	+	IPC	+	-	-	+	-	+	-	+
FPO	+	+	IPC	-	-	-	-	-	+	-	+
CiteSeerX	+	+	П	+	-	+	-	-	-	-	+
Google Scholar	-	+		+	-	-	-	+	+	+	+
Microsoft Academic Search	-	+	П	+	-	-	-	-	-	-	+
Thomson Innovation	+	+		+	-	-	+	+	+	-	+
ОТЕЧЕСТВЕННЫЕ											
РИНЦ	-	+	ГРНТИ	-	-	-	-	+	+	+	+
FIPS	+	+	IPC	-	-	-	-	-	-	+	-
Exactus Expert	+	+	П	+	+	+	+	+	+	+	+

¹ символ П в столбце «Система классификации» означает, что в описываемой системе применяется своя плоская классификация.

Patents Index¹. ФИПС² – база российских патентов, позволяющая проводить многоаспектный патентный поиск, поиск патентов-аналогов, тематический поиск по ключевым словам и международным классификациям.

При всем своем разнообразии патентные базы и поисковые машины по ним имеют много общего, что обусловлено Парижской конвенцией (1883 г.) и другими международными соглашениями о предоставлении патентной информации и регистрации прав на интеллектуальную собственность резидентами и нерезидентами стран участниц конвенции. В основном различия между патентными базами и поисковыми машинами сводятся к различному наполнению, языку патентов и различного рода аналитическими отчетами. В последние годы наметился тренд отхода от архаических поисковых машин по патентам, использующих булевый поиск по ключевым словам и кодам патентной классификации, в сторону использования методов анализа полнотекстовых документов, использующих последние достижения в области информационного поиска, опирающиеся, в том числе, и на методы компьютерной лингвистики.

Патентные базы являются важным источником информации по актуальному состоянию научного задела страны, а также порой позволяют прогнозировать развитие инноваций. Как показывают исследования [4, 7], анализ баз патентов помогает быстрее выявлять новые тренды в научных исследованиях и перспективных направлениях. Однако нельзя забывать о том, что в большинстве случаев наблюдается сильная взаимосвязь между количеством патентов и публикаций и для прогнозирования требуется дополнительная информация, например, количество коллективов, работающих по конкретной тематике [9]. Потому исследователи нередко предлагают сочетание анализа наукометрических и патентных баз [10].

Научные поисковые машины отличаются от перечисленных выше систем тем, что они могут работать с полным текстом документа, а не только с названием, аннотацией, ключевыми словами или кодами. Google Scholar³ и CiteSeerX⁴ индексируют научные работы, определяют цитирования в автоматическом режиме, а также автоматически извлекают метаданные, находят содержательно похожие документы, строят статистику цитирований и аннотации. Google Scholar отличается более широкой базой документов и более качественным извлече-

нием метаданных, тем самым очень полезен для анализа авторитетности научных организаций и сотрудников. Microsoft Academic Search⁵ реализует поиск по документам из разных областей науки.

Российский аналог научных поисковых машин – Exactus Expert⁶. Это информационно-аналитическая система, главным достоинством которой является интеграция в рамках единой платформы широкого набора функций [13]. Среди возможностей системы следует отметить оценку публикационной активности коллективов исследователей, научный поиск, который удобен в тех случаях, когда пользователь затрудняется сформулировать запрос или недостаточно знаком с темой поиска, поиск некорректных заимствований и ряд других особенностей (см. табл. 1).

Выбор инструмента для анализа российских научно-технологических заделов

Для того чтобы определить место России в мировом научно-технологическом пространстве, необходимо привлечь широкий круг экспертов, однако им, в свою очередь, необходимы инструменты для объективной оценки ситуации. Могут ли российские эксперты опереться только на зарубежные системы при анализе заделов отечественной науки? На наш взгляд, ответ на этот вопрос однозначный: нет, на данном этапе объективный анализ развития российской науки невозможен через призму зарубежных инструментов по ряду причин, о которых пойдет речь ниже.

Напомним, что зарубежные системы (Scopus, WoS) вычисляют свои параметры на основе цитирований^{7,8}. Однако лишь 5-7% всех отечественных публикаций попадают в WoS или Scopus [2], в то время как большинство работ не индексируются, то есть оценить их значимость на основе этих баз невозможно.

В последние годы предпринимаются попытки решить эту проблему путем принуждения российских ученых публиковаться в индексируемых Scopus и WoS журналах. Вполне ожидаемо, что такой подход не дает положительного результата, а существующая в научно-исследовательских институтах и вузах система поощрения публикаций в WoS и Scopus привела к тому, что российские ученые и преподаватели готовы опубликоваться в любом доступном журнале, который входит в заветные списки. К сожалению, самыми доступными

⁵ <http://academic.research.microsoft.com>

⁶ <http://expert.exactus.ru>

⁷ http://elsevierscience.ru/files/pdf/SciVerse_Scopus_User_Guide_RUS.pdf;

⁸ http://www.spa.msu.ru/uploads/files/nautchnaja_dejatelnost/metodichka_vasiljeva.pdf

¹ <http://thomsonreuters.com/en/products-services/intellectual-property/patent-research-and-analysis/derwent-world-patents-index.html>

² http://www1.fips.ru/wps/wcm/connect/content_en/en/pat_lib

³ <http://scholar.google.ru>

⁴ <http://citeseerx.ist.psu.edu/index>

оказываются далеко не самые лучшие журналы. Передко это журналы с низким импакт-фактором или труды конференций, которые практически не читаются и не цитируются ни западными, ни российскими учеными. Таким образом, имея качественную публикацию для хорошего журнала, не индексируемого WoS и Scopus, авторы сознательно размещают ее в журнале с низким рейтингом. Инвестиции значительных усилий в публикации оказываются бесплодными с научной точки зрения, хотя и позволяют заработать бонусы в глазах администрации вузов и институтов. Передко публикация в рейтинговых журналах, индексируемых в WoS и Scopus, а затем и доступ к этой публикации, возможны только на платной основе. В условиях хронического недофинансирования научных институтов и вузов это становится существенным препятствием в продвижении российских научных публикаций в Scopus или WoS.

Кроме того, структура российской науки по Scopus или WoS сильно отличается от структуры в РИНЦ. Если в западных базах больше всего отечественных публикаций по физике, астрономии, науках о материалах, химии, то в РИНЦ лидируют экономика, медицина, юридические науки и педагогика. Это сильно осложняет анализ, так как приходится иметь дело с кривыми зеркалами, которые по-разному отражают российскую науку.

Таким образом, применение только западных инструментов для анализа заделов России невозможно: оценка проводится на нерепрезентативной выборке малого размера, на наши публикации практически не ссылаются, ввиду чего затруднительно использование методов анализа на основе цитирований, структура нашей науки в зарубежных базах сильно отличается от реального положения дел в России, западные системы плохо поддерживают русский язык и т.д. Единственным выходом в сложившейся ситуации является развитие собственных инструментов, таких как Exactus Expert [13] или РИНЦ.

Пример анализа научно-технологического задела

В данном разделе приведены результаты анализа научно-технологического задела России, выполненного при помощи системы Exactus Expert, по научному направлению «фотоника». На этом примере авторы статьи хотят показать, как интеграция на одной платформе существующих российских ресурсов, а именно патентов, авторефератов, научных публикаций и других материалов, находящихся в открытом доступе, может обеспечить качественный научно-технический анализ в масштабах России в целом.

Для анализа научного направления «фотоника» авторы использовали следующие функции системы Exactus Expert: семантический поиск, поиск по метаданным, поиск похожих документов, кластеризация, поиск научных коллективов, тематический анализ предметной области. В качестве источников данных выступали коллекции российских журналов, авторефератов докторских и кандидатских диссертаций, российских патентов и полезных моделей по данным за промежуток 2007–2013 гг. Анализ выполнялся группой аналитиков, которые не являются специалистами или экспертами в предметной области «фотоника».

На рисунке 1 приведены результаты тематического анализа, под которым в системе Exactus Expert понимается распределение количества работ (патентов, статей, авторефератов) по годам. Временной интервал выбран неслучайно – именно в промежутке 2007–2013 гг. имеется достаточно полный набор данных для анализа.

На графике наблюдается стабильный рост числа авторефератов по «фотонике». Для публикаций и патентов также можно отметить позитивный тренд. В 2012 году происходит всплеск российских патентов на изобретения. На основе сопоставления динамики авторефератов, публикаций и патентов

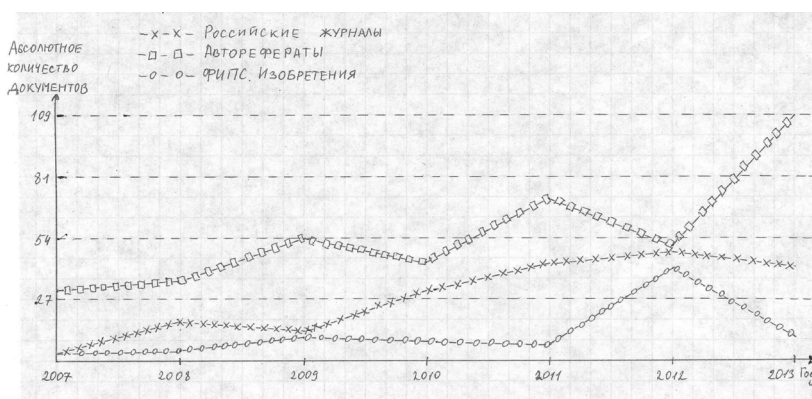


Рис. 1. Тематический анализ по научному направлению «фотоника»

можно сделать предположение, что «фотоника» – перспективное направление: им сегодня в России занимается большое количество аспирантов и докторантов, а в последнее время стало появляться больше прикладных результатов, о чем говорит рост патентной активности.

В качестве обратного примера можно привести результат тематического анализа по запросу «квантовый компьютер» (рис.2). Наибольший интерес российского научного сообщества к квантовым компьютерам приходится на 2007-2010 гг., после чего количество авторефератов и публикаций снижается. При этом за всплеском защит в 2008-2010 годах не последовал стремительный рост патентной активности. Вероятно, сейчас данное направление находится в стадии переосмысления, после чего может начаться как рост, так и дальнейшее угасание направления.

Но вернемся к более общему направлению «фотоника», в результате кластеризации на основе тематического анализа найденных источников было выделено 9 устойчивых кластеров, которые можно охарактеризовать как научные направления (оптоволоконная техника, лазерная медицина, голографические материалы, фотозлектроника, нанофотоника, оптоинформатика и др.). Таким образом, можно сделать вывод, что в России существует минимум 9 направлений исследований внутри направления «фотоника».

Система Eхastus Expert выделила 25 групп исследователей, которые публикуют работы по поднаправлениям «фотоники». На следующем шаге для облегчения работы эксперта система Eхastus Expert может произвести количественный анализ следующих параметров:

- публикационная активность отдельных групп и авторов;
- качество и количество их публикаций, плагиат и автоплагиат;
- принадлежность к междисциплинарным исследованиям.

На основе таких данных можно сделать предположения о том, насколько стабильно направление и отдельные коллективы, другими словами, имеется ли задел по этому направлению, продолжается ли рост количества публикаций и патентов. Так, анализ направления «фотоника» показал, что все найденные коллективы имеют публикации в последние годы, а в целом по публикациям коллективов можно отметить положительный тренд, то есть можно предположить, что наблюдается развитие направления.

Анализ также показал, что внутри направления «фотоника» в последние годы начали складываться междисциплинарные поднаправления, такие как «компьютерная фотоника» и «оптоинформатика». Эти направления были получены исходя из поиска похожих документов и на основании публикационной активности. Авторам статьи удалось найти 9 коллективов, работающих на стыке компьютерных и физико-математических наук, активно публикующихся в последнее время.

Таким образом, на основе проведенного анализа можно предположить, что научные коллективы в этой области будут развиваться, появятся новые междисциплинарные направления исследований. Это подтверждается также тем, что за последние пять лет в России стали открываться научно-образовательные центры, специализирующиеся на исследованиях в области фотоники, организовываться конференции и выставки^{1,2,3}.

Заключение

Проведенное исследование показало, что применение только западных инструментов для оценки научно-технологических заделов России невозможно – слишком мала выборка научных публикаций,

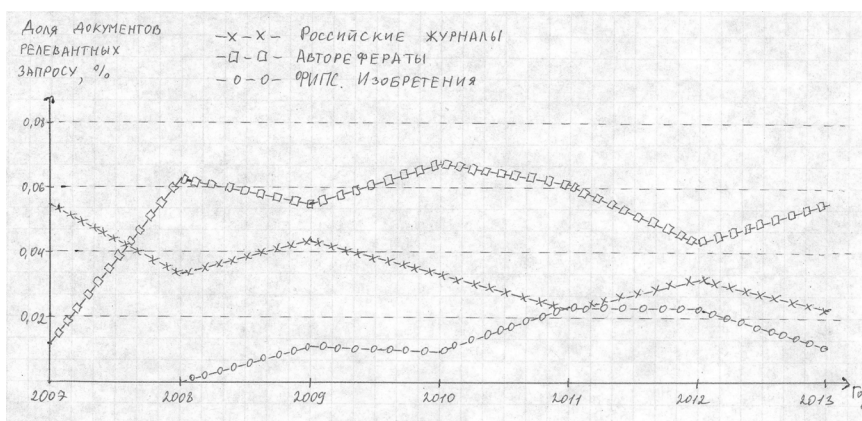


Рис. 2. Тематический анализ по запросу «квантовый компьютер»

¹ http://laser-portal.ru/content_420

² <http://fioconf.mephi.ru/archive/>

³ <http://www.photonics-expo.ru>

на которой проводится анализ, на российских авторов плохо ссылаются, структура российской науки в Scopus и WoS отличается от РИНЦ, ввиду вышесказанного невозможно делать достоверные и обоснованные выводы. Единственным выходом в сложившейся ситуации является развитие собственных инструментов наукометрического анализа, которые хотя и требуют дальнейшего уточнения и апробации, но даже на уровне их текущего развития могут быть использованы для анализа научно-технологических заделов России. Их использование в сочетании с экспертной оценкой позволяет оценить сильные и слабые стороны российской науки, выявить междисциплинарное взаимодействие, что в перспективе позволит повысить качество принимаемых управленческих решений в научно-технологической сфере.

Литература

1. Домнич Е. Л. Научно-технологический задел Дальневосточного федерального округа и стран Северо-Восточной Азии // *Пространственная экономика*. – 2012. – № 3. – С. 79-110.
2. Куракова Н.Г., Григорьев О.Г., Тихомиров И.А., Девяткин Д.А. Оценка соответствия мировому уровню исследований в условиях самоизоляции российской науки: проблемы и возможные решения // *Журнал “Экономика науки”*. 2015. №1.
3. Куракова Н. Г., Петров А. Н. Национальная технологическая инициатива: оценка перспектив технологического лидерства России // *Экономика науки*. – 2015. – № 2.
4. Месропян В.Р. Методика прогнозирования развития научно-технических циклов // *Инновации*. 2012. № 9 (167).
5. Соколов А. В. Форсайт: взгляд в будущее // *Форсайт*. – 2007. – Т. 1. – №. 1.
6. Тищенко В.И. Основания современных методов прогнозирования и определения приоритетов развития науки // *Выявление приоритетных научных направлений: междисциплинарный подход*. – М: ИМЭМО РАН. 2016. – 181 с
7. Abraham B. P., Moitra S. D. Innovation assessment through patent analysis // *Technovation*. 2001. 21 (4).
8. Baudour F., van de Kuilen A. Evolution of the Patent Information World – Challenges of yesterday, today and tomorrow // *World Patent Information*. 2015. Т. 40.
9. Bengisu M., Nekhili R. Forecasting emerging technologies with the aid of science and technology data bases // *Technological Forecasting and Social Change*. 2006. Т. 73. №7.
10. Daim T. U. et al. Forecasting emerging technologies: Use of bibliometrics and patent analysis // *Technological Forecasting and Social Change*. 2006. Т. 73. №. 8.
11. Falagas M.E. et al. Comparison of PubMed, Scopus, Web of Science, and Google scholar: strengths and weaknesses // *The FASEB journal*. 2008. Т. 22. №. 2.
12. Osipov G. et al. Information Retrieval for R&D support // *Professional Search in the Modern World*. Springer International Publishing. 2014.
13. Osipov G. et al. Exactus Expert—Search and Analytical Engine for Research and Development Support // *Novel Applications of Intelligent Systems*. – Springer International Publishing, 2016. – С. 269-285

Тихомиров Илья Александрович. Зав. лабораторией ИСА ФИЦ ИУ РАН. К.т.н, доцент. Окончил в 2002 г. Рыбинскую государственную авиационную технологическую академию им. П.А. Соловьева. Количество печатных работ: более 60. Область интересов: интеллектуальный поиск и анализ больших массивов текстов, фильтрация контента и анализ поведения пользователя в сети, научно-технологический анализ и прогнозирование. E-mail: tih@isa.ru.

Тоганова Наталья Владимировна. И.о. зав. сектора Института мировой экономики и международных отношений РАН. К.э.н. Окончила в 2005 г. Московский государственный лингвистический университет. Количество печатных работ: 40. Область интересов: экономика и инновационная политика Германии и ЕС, развитие зеленой энергетики. E-mail: toganova@imemo.ru.

Ананьева Маргарита Игоревна. Инженер-исследователь ИСА ФИЦ ИУ РАН. Окончила в 2013 г. Московский государственный лингвистический университет. Количество печатных работ: 3. Область интересов: компьютерная лингвистика, анализ текстовой информации. E-mail: ananyeva@isa.ru.