

Тематический анализ в наукометрических системах

А.С. Козицын, С.А. Афонин, Д.А. Шачнев

НИИ механики МГУ им. М.В. Ломоносова

Аннотация. Применение современных методов тематического анализа для аналитической обработки больших объемов информации используется в настоящее время практически во всех сферах человеческой деятельности, в том числе, в наукометрии. Многие наукометрические системы и системы цитирования, включая всемирно известные WoS, Scopus, Google Scholar, разрабатывают тематические рубрикаторы для поиска и обработки информации. Важными практическими задачами, которые могут решаться с применением методов тематической классификации, являются: оценка динамики развития тематических направлений в организации, в отдельной стране и в мировой науке в целом; поиск статей по заданной тематике; поиск и оценка авторитетности экспертов; поиск журналов для публикации и другие актуальные задачи. В рамках работ по проекту ИСТИНА, разрабатываемой в МГУ им. Ломоносова, созданы программные реализации алгоритмов для решения некоторых из перечисленных задач, и ведутся научные исследования с целью создания новых эффективных математических моделей и алгоритмов в этой области.

Ключевые слова: тематическая классификация, библиографические данные, граф соавторства, информационные системы

Thematic analysis in scientometric systems

A.S. Kozitsin, S.A. Afonin, D.A. Shacnev

Institute of Mechanics Lomonosov Moscow State University

Abstract. The use of modern methods of thematic analysis for the analytical processing of information is currently used in almost all areas of human activity, including scientometrics. Many scientometric and citation systems, including the world famous WoS, Scopus, Google Scholar, develop thematic categories for searching and processing information. Most important tasks that can be solved using thematic classification methods are: assessment of the dynamics of the development of thematic areas in the organization, country and in world science; search for articles on a given topic; search and assessment of the authority of experts; search for journal for publication and other relevant

tasks. The Lomonosov Moscow State University is currently developing and using the system ISTINA. In this project, algorithms have been created that solve some of the problems listed. Scientific research is underway to create new effective mathematical models and algorithms in this area.

Keywords: thematic classification, bibliographic data, graph of co-authorship, Information Systems

1. Введение

Применение современных методов тематического анализа для аналитической обработки больших объемов информации используется в настоящее время практически во всех сферах человеческой деятельности, в том числе, в наукометрии. Результаты тематического анализа научной информации могут использоваться в целях уточнения наукометрических показателей, принятия управленческих решений, информационного поиска и для определения правил доступа к информации.

Расчет наукометрических показателей используется для оценки значимости статей (цитируемость), авторитетности журналов (импакт-фактор, h5-индекс, h5-медиана), влияния на научное сообщество отдельных авторов (индекс Хирша и g-индекс) оценки деятельности организаций в целом (i-индекс) [1]. Однако, многими авторами отмечается, что характеристики распределения абсолютных числовых значений наукометрических показателей имеют существенную зависимость от анализируемой тематической области [2]. Например, значения индекса цитируемости статей за последние 2 года имеют различную медиану для физики и математики, поскольку математические статьи дольше цитируются, но медленнее «набирают» количество ссылок. Аналогичное несоответствие показателей наблюдается и в журналах в целом. Например, лучшие российские журналы по данным РИНЦ, представленные на странице статистики elibrary.ru/titles_compare.asp, за 2018-ый год по состоянию на 20.04.2020 имеют следующие показатели цитируемости по разным рубрикам: физика 9200; биология 4600; математика 3600; психология; механика 1500; информатика 1100. В этой связи, проводить сравнение по абсолютным значениям наукометрических показателей статей, журналов или авторов из разных тематических направлений некорректно. Необходимо в таких случаях использовать нормализованную среднюю цитируемость [3] или другие аналогичные показатели, учитывающие тематическую область проводимых исследований. Построение таких нормализованных показателей требует проведения тематической классификации больших объемов научных статей и журналов.

При осуществлении управленческой деятельности использование результатов тематического анализа позволяет оценивать состояние различных направлений исследований, проводить их сравнение с мировым

уровнем, выявлять новые тематические направления для определения для определения политики выделения материальных ресурсов для стимулирования научной деятельности. При этом необходимо оценивать не только текущие значения показателей, но и их динамику во времени, а также мировые показатели. Например, уменьшение показателей по определенной тематике исследований на фоне роста этих же показателей в мире может означать отток научных кадров в организации из этой области исследований или устаревание оборудования.

Еще одним важным направлением применения тематического анализа является создание эффективных механизмов для проведения информационного поиска. Объектами поиска могут являться: публикации, журналы, персоны, организации и другие объекты. На основе проведения тематической классификации и кластеризации могут решаться такие актуальные задачи, как поиск опубликованных материалов по заданной тематике, поиск наиболее авторитетных экспертов в определенной предметной области, определение списка журналов для публикации и оценка их значимости, выделение новых тематических направлений в какой-либо области и поиск научных коллективов.

Определение тематических связей между объектами информационной системы [4] также может использоваться для автоматического построения онтологий и определения правил доступа к данным в моделях логического разграничения доступа ABAC (Attribute-Based Access Control) [5], которые в настоящее время в значительной степени потеснили старые модели разграничения доступа: ролевую модель RBAC; мандатную модель MBAC и дискреционную модель DAC.

Многие крупные наукометрические системы и системы цитирования имеют инструментарий, позволяющий проводить тематический анализ данных.

2. Использование тематического анализа в современных системах

Возможности тематического анализа различных наукометрических систем различаются по типу обрабатываемой информации, видам классификаторов, источникам информации, по набору используемых методов классификации и кластеризации.

Проект Web of Science (WoS) дополнительно к индексации по авторским ключевым словам (Author Keywords) и ключевым словам, выделенным автоматически (KeyWords Plus), использует два основных классификатора: одноуровневый классификатор Web of Science Categories для журналов, содержащий 250 категорий и двухуровневый классификатор статей Research Area по 150 областям науки. Кроме того используется классификатор Essential Science Indicators из 22 категорий.

Имеет сервис Manuscript Matcher рекомендации журнала для публикации по тексту рукописи [6].

Проект Google Scholar использует двухуровневый классификатор с 8 элементами первого уровня и 400 элементами второго уровня. Разбиение по темам может использоваться для фильтрации журналов при показе их показателей (h5-индекса и h5-медианы). Для корректировки данных, в соответствии с основной методикой Google, активно использует взаимодействие с пользователем для сбора информации «снизу вверх», позволяя авторам создавать собственные страницы со списком статей, фотографией, описаниями интересов (Google Scholar Citations).

Проект Scopus использует двухуровневый классификатор All Science Journal Classification Codes (ASJC), содержащий 4 записи первого уровня и около 350 записей второго уровня для классификации журналов. На странице www.scival.com можно посмотреть распределение журналов по областям и относительные нормированные характеристики для выбранных разделов классификатора, изменение количества публикаций по тематикам с течением времени и другие показатели.

Проект РИНЦ использует трехуровневый Государственный Рубрикатор НТИ России (ГРНТИ), содержащий около 8 тыс. рубрик (elibrary.ru/rubrics.asp). Тематическую классификацию можно использовать для поиска журналов и статей, а также для фильтрации результатов отбора журналов при выдаче их наукометрических параметров.

Кроме перечисленных выше классификаторов коммерческих систем существует целый ряд общепринятых классификаторов не связанных с какой-то конкретной системой цитирования. На мировом уровне наиболее известным считается трехуровневый классификатор OECD Fields of Science, содержащий более двухсот рубрик, который планировалось использовать, в том числе, в проекте «Карта российской науки». Во многих российских журналах для тематической классификации статей самими авторами используется более подробная Универсальная десятичная классификация (УДК), содержащая более 150 тысяч рубрик. Также для тематической классификации различных научных материалов используются Рубрикатор ВИНТИ, содержащий более 53 тысяч рубрик, и целый ряд других тематических классификаторов: Классификатор Российского научного фонда (РНФ) [7]; Классификатор Российского фонда фундаментальных исследований (РФФИ) [8]; Международная патентная классификация (МПК) [9], Общероссийский классификатор стандартов (ОКС) [10], Mathematics Subject Classification (MSC) [11]; Journal of Economic Literature Classification (JEL) [12] и другие (scs.viniti.ru/MapService/treeList.aspx). При наличии такого многообразия классификаторов закономерным является появление различных проектов по их согласованию. Например, проект по сопоставлению классификаторов Scopus и OECD [13], а также проект ВИНТИ [14].

Все перечисленные выше проекты ставили своей целью подсчет показателей цитирования научных публикаций и проведение их тематической классификацию по областям науки. Следующим шагом развития стало появление на их основе систем оценки научной деятельности организаций в целом.

Испанский проект SCImago Journal & Country Rank Гранадского университета (или «Атлас науки») оценивает на основе данных Scopus агрегированные данные о научной деятельности в Испании, Португалии и странах Южной Америки. На сайте проекта www.scimagojr.com приводятся показатели не только по научным журналам, но и по странам в целом. Индекс SJR, разработанный авторами проекта является альтернативой импакт-фактору.

Проект Faculty Scholarly Productivity Index (FSPI) оценивает на основе данных Scopus метрические показатели университетов США. Помимо количества публикаций и показателей цитирования в этом проекте для расчета ранга университета используются данные о полученных наградах и премиях, а также об объемах федерального финансирования исследований. На основе агрегированных данных производится ранжирование более чем 350 университетов.

Проект Times Higher Education (THE) ставит своей задачей оценку университетов всего мира [15]. Разработанный в рамках проекта индекс World University Rankings строится на основе данных о цитируемости WoS, которые составляют 32.5% рейтинга [16].

Проект QS World University Rankings [17] оценивает по показателям исследовательской и преподавательской деятельности, репутации у работодателей, а также количеству иностранных студентов и преподавателей.

Проект Academic Ranking of World University (ARWU), который часто называют «Шанхайским рейтингом» (www.shanghairanking.com), учитывает получение выпускниками университета Нобелевских премий, количество опубликованных статей в журналах «Nature» и «Science» и показателей цитируемости.

Следует отметить, что проведение подобных сравнений без учета языка преподавания, так же, как и оценка журналов без учета их тематической области, дает не вполне точные результаты [18]. Например, сравнение университетов всего мира по уровню цитируемости только в англоязычных источниках показывает, что в университетах США, Англии и Канады процент преподавателей и студентов, свободно владеющих английским языком, значительно выше, чем в России. Аналогично, сравнение доли иностранных студентов и преподавателей в университетах с английским и с японским языком преподавания показывает не столько уровень образования в учебном заведении, сколько количество иностранцев, свободно владеющих данным языком.

Учет области проведения исследования при анализе наукометрических данных, в том числе языка, тематической области и других подобных характеристик, является необходимым требованием для построения более объективных наукометрических показателей.

3. Тематический анализ с использованием текстовой информации и классификаторов

В процессе разработки и развития наукометрической системы ИСТИНА всегда уделялось особое внимание развитию методов интеллектуального анализа информации, в том числе, методам тематического анализа. Объем данных, обрабатываемой в настоящее время этой системой значительно уступает мировым системам цитирования, поскольку охватывает всего 28 организаций, 900 тысяч публикаций, 70 тысяч монографий и 13 тысяч патентов. Однако, количество типов используемых данных значительно выше. Кроме публикаций и патентов в системе присутствует полная информация о данных по научным проектам (НИР, НИОКР, гранты), докладах на конференциях, диссертациях и дипломах, об участии сотрудников в деятельности различных советов и редколлегиях, получаемых ими премиях и наградах, читаемых учебных курсах и других данных [19]. Кроме того, информация в системе проходит двойную проверку. Основным принципом работы системы является движение информации «снизу-вверх». На первом этапе пользователь, как наиболее заинтересованное лицо, регистрирует в системе все свои работы, которые отображаются на его персональной странице. На втором этапе ответственные сотрудники подразделений подтверждают достоверность данных. Подобный метод сбора информации с использованием создания персональных страниц в настоящий момент использует в проекте Google Scholar Citations корпорация Google, являющаяся лидером на рынке обработки текстовых данных. Но в силу объективных причин в этой системе невозможно организовать второй этап верификации.

Одним из наиболее простых способов проведения тематического анализа является использование классификаторов с ручным сопоставлением объектов и тематических классов, в том числе использование тематической классификации карточек журналов. Такой подход используется в Scopus, WoS и РИНЦ. В наукометрической системе ИСТИНА на начальном этапе для анализа активности сотрудников и организаций на различных тематических направлениях также были реализованы методы анализа с использованием рубрикации журналов рубрикаторы. С использованием интерактивного интерфейса на странице статистики организации [20] представлены данные о распределении числа статей, цитируемости WoS, числа авторов и других агрегированных характеристик по рубрикам Scopus и ГРНТИ. Данные могут предоставляться как по отдельным подразделениям, так и по организации

в целом с возможностью фильтрации по году, по преодолению порогового значения и по принадлежности к группе журналов: Scopus, Top25, ВАК, сборники и другие.

Информация о публикациях

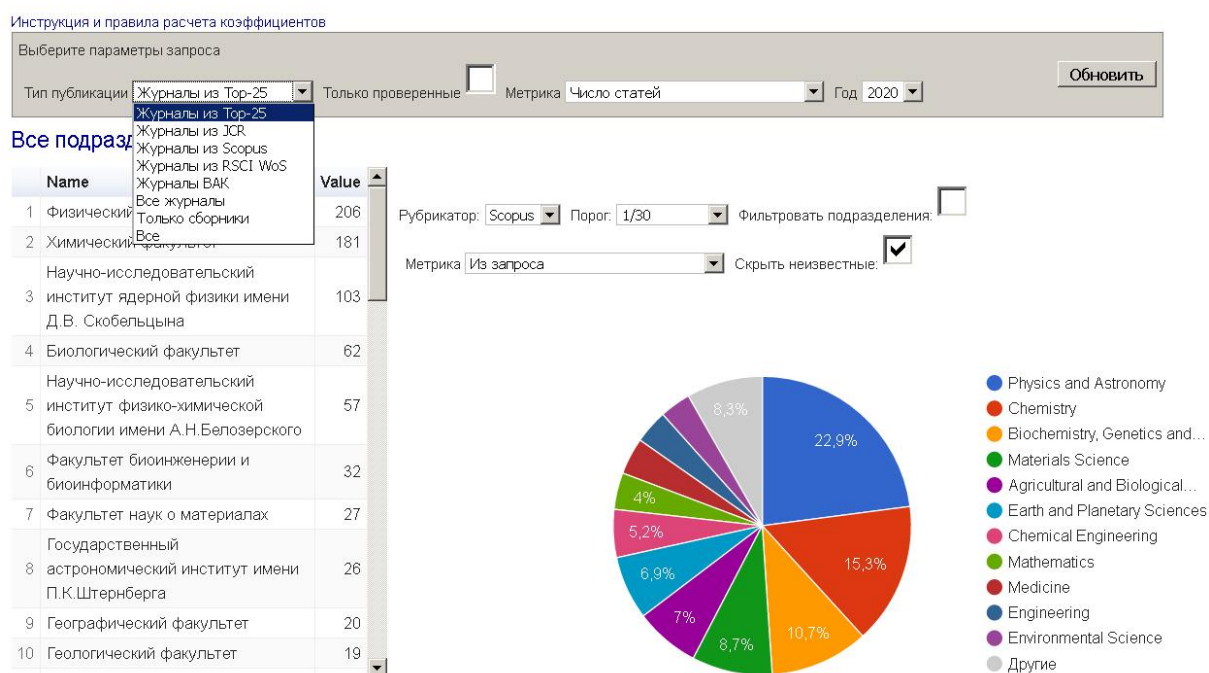


Рис. 1. Интерфейс для анализа распределения данных по рубрикам

Подобный интерфейс предоставляет пользователю возможность оценить степень публикационной активности сотрудников в различных тематических областях. Однако, такой подход не позволяет анализировать информацию с достаточной степенью детализации.

Вторым возможным подходом является определение тематики и поиск информации по ключевым словам, аннотациям или полным текстам статей. Ключевые слова могут задаваться авторами работы при ее регистрации в системе или вычисляться в процессе индексации из аннотации или полных текстов статей, например, Author Keywords и KeyWords Plus в WoS. Такой подход позволяет в большей степени конкретизировать тематику поиска, которая необходима для таких задач, как выделение новых тематических направлений или поиск информации по конкретной информационной потребности пользователя. Следует отметить, что применение подобного тематического анализа не ограничивается только поиском информации. Например, в работе [21] предлагается использовать тематический анализ для оценки качества журнала. Основная гипотеза состоит в том, что в «хороших» научных журналах статьи должны быть посвящены фиксированному набору тематик и эти тематики должна меняться со временем. Таким образом, после обучения тематического алгоритма на наборе журналов можно

провести тематическую и временную классификацию статей из этих журналов. Качество журнала будет пропорционально точности классификации, с которой содержащиеся в нем статьи были отнесены к этому журналу и к временному промежутку публикации.

Основная сложность использования для тематического анализа ключевых слов состоит в ограниченности его набора. При описании статей авторы обычно указывают менее 20 слов. На первом этапе авторы выделяют из статьи основные понятия, и их выбор является субъективным, на втором этапе для каждого понятия указывается только одно его отображения на множество ключевых слов без учета возможных синонимов. Таким образом, статьи схожей тематики могут иметь непересекающийся набор ключевых слов, и точность определения их тематического сходства значительно снижается. Вместе с тем, подобный подход к поиску схожих по тематике статей реализован в некоторых системах цитирования.

Проект WoS предоставляет пользователям сервис Manuscript Matcher подбора журнала для публикации по тексту статьи. Для работы сервис требует предварительной регистрации пользователя. После загрузки заголовка предполагаемой к публикации статьи и аннотации сервис определяет ключевые слова и ищет соответствие с ключевыми словами журналов. Результат показывается в виде списка журналов с указанием описания журнала, а также список общих ключевых слов с указанием меры сходства с загруженной статьей. Сервис может быть полезен для авторов, которые используют узкоспециализированные термины, например, в химии, биологии или астрономии. Для более общих тем сопоставление терминов дает не очень точный результат. Например, для статьи «Determining the thematic proximity of scientific journals and conferences using Big Data technologies» лучшими журналами в результатах поиска являются «Scientometrics» и «Journal of the association for information science and technology», однако, в top5 попадают журналы «Journal of medical systems» и «Journal of digital imaging» по терминам «create software tools» и «full-text information».

Проект РИНЦ предлагает пользователям сервис поиска похожих статей. Пользователь может выбрать одну из уже проиндексированных в системе статей и запросить поиск похожих статей по тематике. Но результаты такого поиска обладают еще меньшей точностью, чем результаты работы предыдущего сервиса. Например, для статьи «Архитектура, методы и средства базовой составляющей системы управления научной информацией "ИСТИНА-НАУКА МГУ"» определяется 14 тысяч близких по тематике статей, и в списке top10 нет ни одной статьи, которая была бы связана с рассматриваемой в статье системой или каким-либо аналогом, и только одна статья затрагивает вопросы наукометрии. В результатах поиска по тематическому сходству

top3 составляют: «Information technology of software architecture structural synthesis of information system», «Анализ применения asp.net при разработке информационной системы analysis of the asp.net development information system», «Общий обзор agris (agricultural research information system)».

Одним из возможных способов улучшения полноты поиска по ключевым словам и разрешение вопросов омонимии является расширение набора ключевых слов на основе построения связей между ключевыми словами [22], а также использование переводов терминов. В проекте ИСТИНА для автоматизации процесса перевода используются материалы Википедии, а также бесплатные сервисы компании Abbyu. Поиск по ключевым словам используется как первый этап тематического анализа в разрабатываемых алгоритмах поиска экспертов и подбора журналов, которые апробируются в настоящее время на данных системы ИСТИНА. Результаты проводимых исследований в этом направлении нельзя еще использовать для реализации промышленного ПО, однако, уже сейчас можно утверждать, что использование только тематического анализа на основе ключевых слов, аннотаций и текстов не позволяет получить удовлетворительный результат классификации. В этой связи, в разрабатываемых алгоритмах используется комбинирование методов полнотекстового анализа и методов анализа теории графов, которые используют явные или неявные связи между классифицируемыми объектами.

4. Тематический анализ с использованием связей объектов

Использование связей между объектами (или граф объектов) позволяет дополнить или уточнить данные анализа в случае недостатка информации. Объекты в графе могут быть одного типа, например, статьи и ссылки между статьями, или иметь разный тип, например, сотрудники и их проекты. Целью анализа графа может быть расширение области поиска или уточнение значимости объектов в существующей области поиска.

Одним из примеров дополнения данных в графе с объектами разного типа является задача поиска экспертов по заданной тематике [23]. Для поиска экспертов в графе авторства определяются наиболее связанные с экспертами объекты (статьи, монографии, проекты, отчеты и другие), определяется степень связи, выделяются ключевые слова объектов, на основе расширенного набора ключевых слов и весов связей графа строится информационный портрет пользователя и оценивается его близость с исходным поисковым запросом. Для апробации алгоритма использовались данные наукометрической системы ИСТИНА. Использование подобных алгоритмов в системах подсчета цитирования затруднено, поскольку граф связей объектов в них содержит только два типа вершин: авторы и публикации. В полноценных наукометрических системах

информационный портрет пользователя составляется из большого количества типов объектов, что улучшает качество результатов.

Примером решения задачи уточнения данных на основе связей между аналогичными объектами является алгоритм определения авторства статьей [24], который реализован в системе ИСТИНА. Предполагается, что авторские коллективы обладают определенной устойчивостью, и вероятность публикации двумя авторами нескольких совместных статей гораздо выше, чем написание статьи, в которых одного из автора заменяет полный однофамилец. В соответствии с этой гипотезой, для разрешения неоднозначности при определении авторов статьи среди всех возможных однофамильцев строится граф соавторства и выбирается наиболее связанная компонента.

Используя граф соавторства можно также решать задачу определение тематической близости журналов без использования данных полнотекстового анализа. Основной гипотезой при реализации этого метода является предположение, что значительная часть авторов публикует статьи в своей предметной области, и, следовательно, несколько журналов, в которых публикуется одинаковый набор авторов, близки по тематике. Исходя из этой гипотезы, тематическая близость двух журналов рассчитывается как взвешенная сумма авторов, имеющих публикации в обоих журналах. Формальное описание алгоритма приводится в работе [25]. Основным отличием данного алгоритма от аналогичных алгоритмов, использующих полнотекстовый анализ или анализ по ключевым словам, является нечувствительность к языку журналов и, как следствие, возможность поиска связей журналов на разных языках. Программная реализация алгоритма используется в системе ИСТИНА для предоставления пользователям удобного интерфейса тематического поиска журналов.

Еще одной важной практической задачей, которая может решаться с использованием описания связей между объектами в наукометрической системе, является задача определения авторитетности экспертов при осуществлении их поиска по тематическому описанию. Для ориентированных графов классическим алгоритмом оценки авторитетности вершин в графе является алгоритм PageRank, который использовался в системе Google для ранжирования результатов поиска. Алгоритм основан на предположении, что входящее ребро в графе подтверждает авторитетность вершины. Причем значимость этого подтверждения тем выше, чем выше авторитетность исходящей вершины. В наукометрических системах алгоритм может эффективно использоваться для анализа графа цитируемости. Для анализа неориентированного графа соавторства и других подобных графов в наукометрических системах возможно использование целого ряда других характеристик: степень связности (количество ребер для каждой вершины); степень близости

(среднее кратчайшее расстояние до других вершин графа); степень посредничества (количество кратчайших путей между всеми парами вершин, проходящих через заданную вершину); степень влиятельности (степень связности, в которой вклад каждого ребра зависит от степени влиятельности соседней вершины, например, PageRank); кросс-кликовая центральность (число клик, которым принадлежит узел) и другие. Предварительные эксперименты, проведенные на данных системы ИСТИНА, показывают, что такой подход может быть достаточно эффективен для использования при ранжировании результата поиска экспертов, автоматического определения устойчивых научных коллективов и других подобных задач.

5. Заключение

Использование алгоритмов тематического анализ для решения целого ряда задач обработки информации в наукометрических системах позволяет создавать удобные сервисы для поиска и обработки информации. Комбинирование полнотекстовых и графовых методов анализа позволяет увеличить точность и полноту представляемых результатов. В настоящий момент в системах научного цитирования такие сервисы представлены недостаточно широко. Научные изыскания на этом направлении, проводимые с использованием данных проекта ИСТИНА, могут позволить получить новые механизмы поиска и обработки наукометрической информации.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект N18-07-01055.

Литература

1. Руководство по наукометрии: индикаторы развития науки и технологии Акоев М.А., Маркусова В.А., Москалева О.В., Писляков В.В.. Екатеринбург: Издательство Уральского университета, 2014, 248 с.
2. Орлов А.И., Наукометрия и управление научной деятельностью//Управление большими системами. Специальный выпуск 44: Наукометрия и экспертиза в управлении наукой. Институт проблем управления им. В. А. Трапезникова РАН. 2013. С.538--568.
3. Бричковский В.В., Наукометрический анализ в информационном обеспечении инновационной деятельности// В мире науки. 2017. №~8(174). С.64--67.
4. Афонин С.А., Козицын А.С., Шачнев Д.А., Программные механизмы агрегации данных, основанные на онтологическом представлении структуры реляционной базы наукометрических данных// Программная инженерия. 2016. Т.7, №9. С. 408-413.

5. Afonin S., Ontology models for access control systems// Proc. of the 3rd International Conference Russian-Pacific Conference on Computer Technology and Applications (RPC). 2018. P. 1-6.
6. Сервис подбора журнала WoS. — <http://mjl.clarivate.com/home>
7. Классификатор РНФ. — <http://www.rscf.ru/node>
8. Классификатор РФФИ. — http://www.rfbr.ru/rffi/ru/contest_documents
9. Классификатор МПК. — <http://www.fips.ru>
10. Классификатор ОКС. — <http://classinform.ru/oks.html>
11. Классификатор MSC. — <http://www.ams.org/msc/>
12. Классификатор JEL. — http://www.aeaweb.org/journal/jel_class_system.html.
13. Проект по сопоставлению классификаторов Scopus и OECD. — <http://report03.metrics.ekt.gr/en/appendixIII>.
14. Проект по сопоставлению классификаторов ВИНТИ. — <http://scs.viniti.ru/MapService/mapform.aspx>.
15. Проект Times Higher Education. — <http://www.timeshighereducation.com>.
16. Индекс World University Rankings. — <http://gtmarket.ru/ratings/the-world-university-rankings/info>.
17. Проект QS World University Rankings. — <http://www.topuniversities.com>
18. Кинчарова А.В., Методология мировых рейтингов университетов: анализ и критика //Университетское управление: практика и анализ, N2, 2014
19. Данные проекта ИСТИНА. — <http://istina.msu.ru/statistics/activity/>.
20. Статистика организации в проекте ИСТИНА. — <http://istina.msu.ru/statistics/organization/214524/dynamic>.
21. Краснов Ф.В., Сравнительный анализ коллекций научных журналов// Труды СПИИРАН. 2019. Т.18. С.767-793.
22. Афонин С.А., Лунев К.В. Выявление тематических направлений в коллекции наборов ключевых слов // Программная инженерия. - 2015. - № 2. - С. 29-39.
23. Vasenin Valery, Lunev Kirill, Afonin Sergey, and Shachnev Dmitry. Methods for intelligent data analysis based on keywords and implicit relations: The case of "istina" data analysis system. In Proc. of the International Conference Actual Problems of Systems and Software Engineering (APSSE 2019), IEEE Conference Proceedings, pages 151-155, United States, 2019. United States
24. Козицын А.С., Афонин С.А.. Разрешение неоднозначностей при определении авторов публикации с использованием графов соавторства в больших коллекциях библиографических данных. Программная инженерия, 8(12):556-562, 2017
25. Козицын А.С., Афонин С.А.. Нахождение скрытых зависимостей между объектами на основе анализа больших массивов библиографических данных.//In Proc. of the International Conference

References

1. Rukovodstvo po naukometrii: indikatory razvitiia nauki i tekhnologii Akoev M.A., Markusova V.A., Moskaleva O.V., Pisliakov V.V.. Ekaterinburg: Izdatelstvo Uralskogo universiteta, 2014, 248 c.
2. Orlov A.I., Naukometriia i upravlenie nauchnoi deiatelnosti//Upravlenie bolshimi sistemami. Spetsialnyi vypusk 44: Naukometriia i ekspertiza v upravlenii nauko. Institut problem upravleniia im. V. A. Trapeznikova RAN. 2013. S.538--568.
3. Brichkovskii V.V., Naukometricheskii analiz v informatsionnom obespechenii innovatsionnoi deiatelnosti// V mire nauki. 2017. №~8(174). S.64--67.
4. Afonin S.A., Kozitsyn A.S., Shachnev D.A., Programmnye mekhanizmy agregatsii dannykh, osnovannye na ontologicheskome predstavlenii struktury reliatsionnoi bazy naukometricheskikh dannykh// Programmnaia inzheneriia. 2016. T.7, №9. S. 408-413.
5. Afonin S., Ontology models for access control systems// Proc. of the 3rd International Conference Russian-Pacific Conference on Computer Technology and Applications (RPC). 2018. P. 1-6.
6. Servis podbora zhurnala WoS. - <http://mjl.clarivate.com/home>
7. Klassifikator RNF. - <http://www.rscf.ru/node>
8. Klassifikator RFFI. - http://www.rfbr.ru/rffi/ru/contest_documents
9. Klassifikator MPK. - <http://www.fips.ru>
10. Klassifikator OKS. - <http://classinform.ru/oks.html>
11. Klassifikator MSC. - <http://www.ams.org/msc/>
12. Klassifikator JEL. - http://www.aeaweb.org/journal/jel_class_system.html.
13. Proekt po sopostavleniiu klassifikatorov Scopus i OECD. - <http://report03.metrics.ekt.gr/en/appendixIII>.
14. Proekt po sopostavleniiu klassifikatorov VINITI. - <http://scs.viniti.ru/MapService/mapform.aspx>.
15. Proekt Times Higher Education. - <http://www.timeshighereducation.com>.
16. Indeks World University Rankings. - <http://gtmarket.ru/ratings/the-world-university-rankings/info>.
17. Proekt QS World University Rankings. - <http://www.topuniversities.com>
18. Kincharova A.V., Metodologii mirovykh reitingov universitetov: analiz i kritika //Universitetskoe upravlenie: praktika i analiz, N2, 2014
19. Dannye proekta ISTINA. - <http://istina.msu.ru/statistics/activity/>.
20. Statistika organizatsii v proekte ISTINA. - <http://istina.msu.ru/statistics/organization/214524/dynamic>.

21. Krasnov F.V., Sravnitelnyi analiz kollektzii nauchnykh zhurnalov// Trudy SPIIRAN. 2019. T.18. S.767-793.
22. Afonin S. A., Lunev K. V. Vyiavlenie tematicheskikh napravlenii v kollektzii naborov kliuchevykh slov // Programmaia inzheneriia. - 2015. - № 2. - S. 29-39.
23. Vasenin Valery, Lunev Kirill, Afonin Sergey, and Shachnev Dmitry. Methods for intelligent data analysis based on keywords and implicit relations: The case of "istina" data analysis system. In Proc. of the International Conference Actual Problems of Systems and Software Engineering (APSSE 2019), IEEE Conference Proceedings, pages 151-155, United States, 2019. United States
24. A.S. Kozitsyn and S.A. Afonin. Razreshenie neodnoznachnosti pri opredelenii avtorov publikatsii s ispolzovanie grafov soavtorstva v bolshikh kollektziiakh bibliograficheskikh dannykh. Programmaia inzheneriia, 8(12):556-562, 2017
25. A.S. Kozitsyn, S.A. Afonin. Nakhozhdenie skrytykh zavisimosti mezhdu obiektami na osnove analiza bolshikh massivov bibliograficheskikh dannykh.//In Proc. of the International Conference Actual Problems of Systems and Software Engineering (APSSE 2019), IEEE Conference Proceedings, pages 320-328, 2019