

Метод поиска экспертов с использованием библиографических данных

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: тематический поиск, библиографические данные, поиск экспертов, информационные системы, наукометрия

Method for expert search using bibliographic data

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 search, bibliographic data, expert search, information systems, scientometrics

1. Введение

С развитием средств коммуникаций и систем удаленного взаимодействия все большую актуальность приобретает задача поиска высококвалифицированных специалистов и экспертов для выполнения работ в удаленном режиме. Такие работы могут являться краткосрочными, периодическими или выполняться на постоянной основе. В настоящее время поиск экспертов наиболее востребован в СМИ. Журналисты и издатели производят поиск консультантов по тематикам публикуемых статей. Владельцы новостных и информационных порталов ищут специалистов для написания статей по заданным тематическим направлениям. В производственной сфере и при разработке программного обеспечения также производится поиск экспертов для проведения разовых консультаций, для работы на постоянной основе в стартапах и развивающихся проектах. Подобный поиск производится обычно или через социальные сети [1, 2], или через биржи задач, на которых заказчики размещают задания, а эксперты выкладывают описания своих компетенций [3,4,5]. Для успешного проведения поиска в таких системах необходимо активное участие в процессе поиска как заказчиков работ, так и самих экспертов.

Задача тематического поиска экспертов в научной сфере существенно отличается от описанных выше задач поиска для СМИ и разработки ПО. Такой поиск востребован при подборе рецензентов для научных статей, выборе экспертов при проведении конкурсов, при поиске высококвалифицированных консультантов для научной работы и выборе студентами научных руководителей. В большинстве случаев при подобном поиске эксперты не готовы принимать активное участие в процессе поиска. Ученый с мировым именем может согласиться прочитать и дать рецензию на статью по просьбе редакции, но он не будет самостоятельно искать направленные в редакцию для публикации статьи, на которые он мог бы дать рецензии. В этой связи, применение биржевых систем для тематического поиска в научной сфере практически невозможно, и требуется использовать иные принципы и механизмы поиска.

2. Обзор существующих систем

Наиболее простым способом поиска специалистов является поиск по описаниям компетенций в профиле сотрудника. Например, в проекте

поиска экспертов среди сотрудников университетов США и Канады [6] экспертам предлагается самостоятельно в личном профиле заполнить список своих компетенций, по которому в дальнейшем поисковая система будет осуществлять поиск по ключевым словам. Такой подход прост в реализации, однако имеет очень низкую полноту поиска. В своих компетенциях сотрудники указывают общие понятия, но не указывают специальных терминов. Например, в системе [6] можно найти 40 зарегистрированных экспертов с компетенцией «big data», но нельзя найти ни одного специалиста по «k-means» или «SVM». Кроме того, подобный метод не позволяет оценить квалификацию экспертов и провести ранжирование результатов поиска.

Более качественные методы поиска экспертов реализуются на основе анализа данных о публикациях. Такие данные содержатся в системах цитирования и наукометрических системах. Реализующие подобный метод поиска проекты можно разделить на три группы.

Проекты первой группы позволяют производить тематический поиск публикаций с последующим ручным анализом информацией об авторах. В качестве примера таких проектов можно привести CiteSeerX [7], Semantic Scholar[8] и сайты издательств [9,10,11].

Проекты второй группы не только позволяют проводить поиск по статьям, но и показывают агрегированные данные по авторам публикаций. Например, проект Digital Library [S12] позволяет просмотреть список всех авторов, найденных по запросу статей, с указанием количества найденных статей для каждого автора. Проект Google Scholar [S13] также позволяет просматривать информацию об авторах найденных статей. В этом проекте для авторов реализована возможность редактировать свой профиль. Проект Microsoft Academic Graph [S14] позволяет проводить полнотекстовый поиск по статьям, и для подборки найденных статей составляет агрегированные списки авторов, тем статей, журналов, конференций и организаций.

К третьей группе проектов относятся проекты с возможностью поиска профилей авторов по ключевым словам. Наиболее известным проектом этой группы является китайский проект AMiner [S15]. Для проведения тематического анализа в нем используются данные о публикациях авторов, полученных патентах и участии в проектах. Поисковый запрос на основе выбранных ключевых слов может быть уточнен с использованием одной из сорока рубрик дополнительного тематического классификатора. Список найденных авторов может ранжироваться как по соответствию тексту запроса, так и по различным показателям авторитетности автора. Для каждого автора в списке результатов поиска дополнительно показывается общее количество зарегистрированных статей автора, количество найденных цитирований и перечень автоматически построенных компетенций. Для детализации и

дополнительного ознакомления с компетенциями автора есть возможность перейти в его автоматически построенный профиль, в котором представлены: распределение выделенных компетенций по годам публикаций; список соавторов; список публикаций, патентов и проектов автора.

Наиболее серьезным недостатком перечисленных выше проектов с точки зрения русскоязычных пользователей является полное отсутствие или крайне незначительный объем используемых для анализа русскоязычных статей. Кроме того, проблемы с поддержкой русского языка влекут за собой сложности с определением имен и фамилий в англоязычных статьях, что приводит к некорректному созданию профилей и неправильной привязке статей к авторам. Например, для статьи [16] системой определены авторы «V. Valery , Z. Maxim, K. Alexander, K. Maxim, S. Dmitry» [17]. Также некорректно работает во многих случаях определение организаций. Например, для статьи [18] из пяти работающих в Московском Государственном Университете им. М.В. Ломоносова соавторов статьи три автора отнесены к организации "Moscow State University", а два к "Michigan State University" [19]. Подобные ошибки определения авторов приводят к значительному ухудшению качества поиска среди русскоязычных авторов.

Следует отметить, что проекты AMiner и Microsoft Academic Graph предоставляют пользователям возможность скачать свои данные для самостоятельного анализа, но только за 2018ый год. Данные последних лет, наиболее востребованные для проведения анализа, являются закрытыми.

Отдельно следует выделить русскоязычный проект РИНЦ [20]. В проекте не предусмотрен поиск авторов по ключевым словам, однако, возможен поиск авторов по фамилии, различным идентификаторам (ORCID и другие), по городу или организации. Дополнительно предусмотрена фильтрация по одной из 70 тематических рубрик. Результаты поиска ранжируются по цитируемости. Основным недостатком поискового модуля в этом проекте является невозможность конкретизации тематической области поиска, поскольку имеющиеся рубрики дают слишком общую формулировку темы (например, «математика», «геология»). Кроме того, в системе отсутствует возможность проведения оценки релевантности поисковому запросу. При поиске по разделу «математика» эксперт с количеством цитирований 30 будет иметь более низкий ранг, чем эксперт с количеством цитирований 40, даже если у первого из них все работы посвящены математике, а у второго только 10% относятся к разделу «математика», а остальные – к другим разделам классификатора.

На основе представленного выше обзора можно сформулировать требования, которые предъявляются к системам тематического поиска

экспертов. Система должна производить поиск в автоматическом режиме без участия в процессе поиска самих экспертов. Для анализа должны использоваться не только статьи, но и другая информация о научной деятельности эксперта. Предназначенная для использования в России система должна уметь работать с русскоязычными материалами, в том числе, распознавать русскоязычные имена и фамилии. В системе должна обеспечиваться возможность выполнения поисковых запросов по общим тематическим разделам рубрикатора и по специальным тематическим направлениям, например, с использованием для поиска заданных пользователем ключевых слов[21].

3. Описание метода

Использование информации из наукометрических систем [22,23] позволяет производить автоматический поиск экспертов и ранжировать результаты поиска. Для тематического поиска авторов могут использоваться следующие размещенные в наукометрической системе данные: библиографические описания статей, тезисов и монографий; доклады на конференциях; диссертации; описания проектов; патенты и свидетельства о регистрации ПО; выступления в СМИ; описание читаемых учебных курсов и другие. По каждому из перечисленных типов научной продукции в наукометрической системе может содержаться полнотекстовое описание (название, аннотация, ключевые слова) и список авторов, задаваемый вручную или с использованием автоматических алгоритмов распознавания[24,25]. Для некоторых типов работ в описании может содержаться дополнительная информация, отражающая значимость работ: цитируемость, принадлежность спискам WoS и ВАК, количество сделанных по проекту публикаций, тип читаемых курсов, объем финансирования проектов и другие.

Задача тематического поиска сотрудников может быть сформулирована следующим образом. Задано множество возможных запросов Q , множество авторов A и множество объектов, разбитых на непересекающиеся подмножества $S = S_1 \cup S_2 \cup \dots \cup S_N$, где N – количество типов научной продукции. Задана функция релевантности текстового описания запросу поиска по ключевым словам и рубрикам $F : S \times Q \rightarrow \mathfrak{R}$, функция значимости работы $L : S \rightarrow \mathfrak{R}$ и функция принадлежности работы автору $P : S \times A \rightarrow (0..1)$. Требуется построить функцию соответствия автора поисковому запросу $L : A \times Q \rightarrow \mathfrak{R}$.

Функция F реализуется с использованием стандартных средств контекстного поиска с учетом морфологического анализа.

Реализация функция L различна для разных типов работ. Для статей ее значение определяется вхождением журнала в списки ВАК, Scopus, РИНЦ, рейтингом журнала и цитируемостью самой статьи. Для проектов

значение функции определяется объемом финансирования, а также количеством и значимостью сделанных по проекту публикаций. Для читаемых курсов учитывается категория курса: потоковые лекции; семинары; спецкурсы по выбору кафедры; спецкурсы по выбору студента. Для остальных типов работ значение функции тождественно равно 1.

Функция P отражает вероятную степень участия автора в создании работы. Как показано в работе [26], при определении функции дискретным образом $P : S \times A \rightarrow \{0,1\}$ (1-для всех соавторов работы и 0 в других случаях) значительно падает точность поиска, поскольку соавторы могут только косвенно участвовать в написании работы и не являться специалистами по рассматриваемой в работе тематике. Поэтому, для задания связи авторов с публикациями и докладами используется формула из работы [25]. Для первого автора публикации значение функции равно $\frac{1}{2} + \frac{1}{2n}$, для остальных авторов $\frac{1}{2n}$, где n - количество соавторов публикации. Для остальных типов объектов функция равна $\frac{1}{n}$.

На первом этапе выполнения запроса осуществляется поиск по объектам всех типов с определением релевантности запроса в соответствии с функцией релевантности текстового описания запросу. На основе полученной выборки для каждого типа объектов по каждому автору вычисляется коэффициент соответствия запросу:

$$R_i(a, q) = \sum_{s \in S_i} F(s, q) P(s, a) L(s).$$

С использованием рассчитанных коэффициентов производится ранжирование авторов для каждого типа работ, и для каждого автора определяется его ранг $ord_i(a, q) \in (0..1)$. Значение 1 соответствует максимальному значению R_i по всем авторам для данного типа работ, 0 – минимальному. Суммарный рейтинг каждого автора вычисляется по формуле

$$M(a, q) = \sum k_i ord_i(a, q),$$

где k_i - коэффициенты значимости каждого типа, рассчитанные на этапе обучения.

Для всех отобранных авторов строится граф соавторства, в котором определяется авторитетность вершин S с использованием алгоритма PageRank [27]. Функция L определяется по формуле

$$L = M \sqrt{S}.$$

Разработанный алгоритм может использовать данные любой наукометрической системы. Однако, при разработке программной реализации алгоритма необходимо учитывать формат данных, используемых в конкретной наукометрической системе. Разработанная

авторами программная реализация алгоритма апробировалась на данных наукометрической информационно-аналитической системы ИСТИНА[22], которая содержит данные по 27 организациям, в том числе, МГУ им. М.В. Ломоносова, РХТУ им. Д.И. Менделеева, ИНЭОС РАН, ИФХЭ РАН, ФГАУ «НМИЦ Нейрохирургии им. ак. Н.Н. Бурденко» Минздрава России, МПГУ, ИФЗ РАН, ИНХС РАН, ИГ РАН и другие.

4. Интерфейс поиска

Для просмотра результатов поиска и апробации алгоритма реализован интерфейс, позволяющий производить поиск по ключевым словам с возможностью фильтрации результатов по одной из 70 общих тематических рубрик. На рис. 1 приведен результат выполнения поискового запроса по ключевым словам «тематический анализ»

Ключевые слова (через запятую)

Область знаний

Подразделение

Отправить Show by Items Search:

V	NAME	DEP
1.139	д.ф.-м.н. проф. Васенин Валерий Александрович	404 Лаборатория автоматизации экспериментальных исследований(Совм)(МГУ); Межфакультетская кафедра математического моделирования и компьютерных исследований(МГУ); Кафедра вычислительной математики(Совм)(МГУ) Заведующий лабораторией; Заведующий кафедрой; Профессор
1.012	к.ф.-м.н. Афонин Сергей Александрович	404 Лаборатория автоматизации экспериментальных исследований(МГУ); Кафедра вычислительной математики(Совм)(МГУ); Межфакультетская кафедра математического моделирования и компьютерных исследований(Совм)(МГУ) Ведущий научный сотрудник; Доцент; Доцент
0.957	к.ф.-м.н. Козицын Александр Сергеевич	404 Лаборатория автоматизации экспериментальных исследований(МГУ); Кафедра вычислительной математики(Совм)(МГУ) Ведущий научный сотрудник; Доцент
0.542	Щачнев Дмитрий Алексеевич	404 Лаборатория автоматизации экспериментальных исследований(Совм)(МГУ) Специалист (зоолог, программист, геолог, инженер и т.п.)
0.417	к.ф.-м.н. Голомазов Денис Дмитриевич	None
0.391	д.ф.-м.н. проф. Мальковский Михаил Георгиевич	Кафедра алгоритмических языков(МГУ) Профессор
0.314	к.т.н. Коршунов Андрей Александрович	404 Лаборатория автоматизации экспериментальных исследований(МГУ) Ведущий научный сотрудник
0.307	Занчурин Максим Анатольевич	404 Лаборатория автоматизации экспериментальных исследований(МГУ); Межфакультетская кафедра математического моделирования и компьютерных исследований(Совм)(МГУ) Младший научный сотрудник; Ассистент

Рис. 1. Результат поиска по запросу «Тематический анализ»

На рис.2 приведен пример выполнения запроса по ключевым словам «Теория упругости»

Ключевые слова (через запятую)	теория упругости
Область знаний	Все
Подразделение	Все

1.167	д.ф.-м.н. проф. Шешенин Сергей Владимирович	Кафедра теории пластичности(МГУ) Профессор
1.028	д.т.н. Лурье Сергей Альбертович	Лаборатория механики прочности и разрушения материалов и конструкций(Совм.)(ИПМех РАН); Кафедра механики композитов(Совм.)(МГУ) Ведущий научный сотрудник; Профессор
0.821	д.ф.-м.н. член-корр. Ломакин Евгений Викторович	Кафедра теории пластичности(МГУ) Заведующий кафедрой
0.721	д.ф.-м.н. доц. Никабадзе Михаил Ушангиевич	Кафедра механики композитов(МГУ) Профессор
0.707	д.ф.-м.н. проф. Тарлаковский Дмитрий Валентинович	202 Лаборатория динамических испытаний(МГУ) Заведующий лабораторией
0.547	д.ф.-м.н. Горбачев Владимир Иванович	205 Лаборатория прочности и ползучести при высоких температурах(Совм.)(МГУ); Кафедра механики композитов(МГУ) Ведущий научный сотрудник; Заведующий кафедрой
0.462	д.ф.-м.н. проф. Васин Рудольф Алексеевич	None
0.458	д.ф.-м.н. доц. Вершинин Анатолий Викторович	Кафедра вычислительной механики(МГУ) Доцент

Рис. 2. Результат поиска по запросу «Теория упругости»

Для детального изучения результатов поисковой выдачи можно перейти по ссылке на профиль автора или на список найденных по запросу научных результатов рис 3.

Show by 10 items Search:

Тип работы	Название работы	Релевантность
Публикация	Автоматизированная система тематического анализа информации	13
Публикация	Методы и средства тематического анализа данных в больших системах на основе ключевых слов и косвенных связей между ними	13
Доклад	Методы и средства тематического анализа данных в больших системах на основе ключевых слов и косвенных связей между ними	13
Публикация	Интеллектуальная Система Тематического Исследования Научно-технической информации (ИСТИНА)	10
Достежение	Развитие и поддержка ИАС «Наука МГУ» («ИСТИНА»)	10
Проект	Развитие и сопровождение системы подготовки принятия решений на основе анализа информации о результатах научно-исследовательской, педагогической и инновационной деятельности ИАС «ИСТИНА»	10
Проект	Математическое и программное обеспечение сложных информационно-вычислительных систем	10

Рис. 3. Список работ сотрудника по заданной запросом тематике

Разработанный интерфейс позволяет получить ранжированный список авторов, имеющих научные результаты в заданной запросом тематической области, и провести более детальный анализ, в том числе, ознакомится с аннотациями и полными текстами работ.

Автоматический поиск экспертов предполагается использовать для подбора рецензентов в журналы. Кроме того, предполагается предоставить

пользователям возможность поиска экспертов по ключевым словам на основной странице поиска системы.

5. Заключение

В настоящий момент сервисы тематического поиска экспертов в русскоязычном сегменте Интернет представлены недостаточно широко. Использование алгоритмов тематического анализа данных наукометрических систем позволяет создавать качественные сервисы для решения задачи поиска экспертов. В представленном в настоящей работе алгоритме для тематического поиска и ранжирования используется комбинирование полнотекстовых и графовых методов анализа. Программная реализация алгоритма апробирована на данных наукометрической системы ИАС ИСТИНА.

Литература

- 1.Группа «Помогите журналисту». — <http://www.facebook.com/groups/pomogitej>
- 2.Группа «Герои и эксперты для СМИ» — <http://www.facebook.com/groups/1792146334361293>
- 3.Биржа копирайтинга — <http://www.textsale.ru>
- 4.Биржа копирайтинга Адвего — <http://Advego.ru>
- 5.Биржа контента eTXT — <http://eTXT.ru>
- 6.Поисковая система Search Engine to Find Experts from Universities — <http://network.expertisefinder.com>
- 7.Поисковая система CiteSerX — <http://citeseerx.ist.psu.edu>
- 8.Поисковая система Semantic Scholar — <http://www.semanticscholar.org>
- 9.Поисковая система ScienceDirect — <http://www.sciencedirect.com>
- 10.Издательство Шпрингер — <http://link.springer.com>
- 11.Поисковая система издательства IEEE — <http://ieeexplore.ieee.org>
- 12.Поисковая система Digital Library — <http://dl.acm.org>
- 13.Система цитирования Google Scholar— <http://scholar.google.com>
- 14.Система поиска авторов Microsoft Academic — <http://academic.microsoft.com>
- 15.Система поиска авторов Aminer — <http://www.aminer.cn>
16. Vasenin V., Zanchurin M., Kozitsyn A. et al. Architectural and technological aspects of the cloud data analysis system development, case of istina system // 5th International Conference on Actual Problems of System and Software Engineering, APSSE 2017. — Vol. 1989 of CEUR Workshop Proceedings (CEUR-WS.org). — CEUR Workshop Proceedings, Vol-1989, 2017. — P. 90-96.

17. Описание статьи «Architectural and technological aspects of the cloud data analysis system development, case of istina system» в системе Microsoft Academic — <http://academic.microsoft.com/paper/2771441762>
18. Васенин В.А., Афонин С.А., Зензинов А.А. и др. Механизмы системы ИСТИНА для интеллектуального анализа состояния и стимулирования хода выполнения проектов в сфере науки и высшего образования// Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23-28 сентября 2019 г., г. Новороссийск) / Под ред. В. В. Воеводин. — ИПМ им. М.В.Келдыша Москва, 2019. — С. 210-221.
19. Описание статьи «Mechanisms of <ISTINA> system for intelligent state analysis and progress stimulation for projects in the sphere of science and higher education» в системе Microsoft Academic — <http://academic.microsoft.com/paper/2986189703>
20. Система цитирования РИНЦ — <http://elibrary.ru>
21. 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 United States, — 2019 — P 151-155,
22. Васенин В. А., Архитектурно-технологические аспекты разработки и сопровождения больших информационно-аналитических систем в сфере науки и образования// Программная инженерия. — 2017. — Т. 8, № 10. — С. 448-455
23. Афонин С.А., Голомазов Д.Д., Козицын А.С., Использование систем семантического анализа для организации поиска научно-технической информации// Программная инженерия, — 2012, — № 2, — С. 29-34.
24. Козицын А.С., Афонин С.А.. Разрешение неоднозначностей при определении авторов публикации с использованием графов соавторства в больших коллекциях библиографических данных. Программная инженерия — 2017 — №8(12) — 556-562,
25. Козицын А.С., Афонин С.А., Алгоритм разрешения неоднозначности имен авторов в ИАС ИСТИНА// Современные информационные технологии и ИТ-образование — 2020 — Т.16, №1 — С.108-117
26. Козицын А.С., Афонин С.А., Шачнев Д.А. Метод оценки тематической близости научных журналов// Программная Инженерия — №6 — 2020 — с. 335-341
27. David F. Gleich, PageRank Beyond the Web. //Proc. Society for Industrial and Applied Mathematics — 2015 — 321-363.

References

1. Gruppa «Pomogite zhurnalistu». — <http://www.facebook.com/groups/pomogitej>
2. Gruppa «Geroi i eksperty dlia SMI» — <http://www.facebook.com/groups/1792146334361293>
3. Birzha kopiraitinga — <http://www.textsale.ru>
4. Birzha kopiraitinga Advego — <http://Advego.ru>
5. Birzha kontenta eTXT — <http://eTXT.ru>
6. Poiskovaia sistema Search Engine to Find Experts from Universities — <http://network.expertisefinder.com>
7. Poiskovaia sistema CiteSerX — <http://citeseerx.ist.psu.edu>
8. Poiskovaia sistema Semantic Scholar — <http://www.semanticscholar.org>
9. Poiskovaia sistema ScienceDirect — <http://www.sciencedirect.com>
10. Izdatelstvo Shpringer — <http://link.springer.com>
11. Poiskovaia sistema izdatelstva IEEE — <http://ieeexplore.ieee.org>
12. Poiskovaia sistema Digital Library — <http://dl.acm.org>
13. Sistema tsitirovaniia Google Scolar— <http://scholar.google.com>
14. Sistema poiska avtorov Microsoft Academic — <http://academic.microsoft.com>
15. Sistema poiska avtorov Aminer — <http://www.aminer.cn>
16. Vasenin V., Zanchurin M., Kozitsyn A. et al. Architectural and technological aspects of the cloud data analysis system development, case of istina system // 5th International Conference on Actual Problems of System and Software Engineering, APSSE 2017. — Vol. 1989 of CEUR Workshop Proceedings (CEUR-WS.org). — CEUR Workshop Proceedings, Vol-1989, 2017. — P. 90-96.
17. Opisanie stati «Architectural and technological aspects of the cloud data analysis system development, case of istina system» v sisteme Microsoft Academic — <http://academic.microsoft.com/paper/2771441762>
18. Vasenin V.A., Afonin S.A., Zenzinov A.A. i dr. Mekhanizmy sistemy ISTINA dlia intellektualnogo analiza sostoianiia i stimulirovaniia khoda vypolneniia proektov v sfere nauki i vysshego obrazovaniia// Nauchnyi servis v seti Internet: trudy XXI Vserossiiskoi nauchnoi konferentsii (23-28 sentiabria 2019 g., g. Novorossiisk) / Pod red. V. V. Voevodin. — IPM im. M.V.Keldysha Moskva, 2019. — S. 210-221.
19. Opisanie stati «Mechanisms of <ISTINA> system for intelligent state analysis and progress stimulation for projects in the sphere of science and higher education» v sisteme Microsoft Academic — <http://academic.microsoft.com/paper/2986189703>
20. Sistema tsitirovaniia RINTs — <http://elibrary.ru>
21. 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 United States, — 2019 — P 151-155,
22. Vasenin V. A., Arkhitekturno-tehnologicheskie aspekty razrabotki i soprovozhdeniia bolshikh informatsionno-analiticheskikh sistem v sfere nauki i obrazovaniia//Programmnaia inzheneriia. — 2017. — T. 8, № 10. — S. 448-455
 23. Afonin S.A., Golomazov D.D., Kozitsyn A.S., Ispolzovanie sistem semanticheskogo analiza dlia organizatsii poiska nauchno-tehnicheskoi informatsii//Programmnaia inzheneriia, — 2012, — № 2, — S. 29-34.
 24. Kozitsyn A.S., Afonin S.A.. Razreshenie neodnoznachnostei pri opredelenii avtorov publikatsii s ispolzovaniem grafov soavtorstva v bolshikh kollektssiiakh bibliograficheskikh dannyykh. Programmnaia inzheneriia — 2017 — №8(12) — 556-562,
 25. Kozitsyn A.S., Afonin S.A., Algoritm razresheniia neodnoznachnosti imen avtorov v IAS ISTINA// Sovremennyye informatsionnyye tekhnologii i IT-obrazovanie — 2020 — T.16, №1 — S.108-117
 26. Kozitsyn A.S., Afonin S.A., Shachnev D.A. Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov// Programmnaia Inzheneriia — N6 — 2020 — s. 335-341
 27. David F. Gleich, PageRank Beyond the Web. //Proc. Society for Industrial and Applied Mathematics — 2015 — 321-363.