

УДК 004.021, 004.42

МЕТОД ПОИСКА ЭКСПЕРТОВ ПО ДАННЫМ НАУКОМЕТРИЧЕСКИХ СИСТЕМ

А. С. Козицын¹ [0000-0002-8065-9061], С. А. Афонин² [0000-0003-3058-9269]

^{1,2} НИИ механики МГУ им. М.В. Ломоносова, Мичуринский пр., 1, Москва, 119192

¹alexanderkz@mail.ru, ²serg@msu.ru

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, отдельной стране и мировой науке в целом; поиск статей по заданной тематике; поиск и оценка авторитетности экспертов; поиск журналов для публикации и другие актуальные задачи. Авторами 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.

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 пуб-

лируемых 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 через социальные сети или биржи задач. Активный поиск экспертов через социальные сети производится при помощи публикации объявлений и поиска среди знакомых или в специализированных сообществах, например, сообществ [1, 2], содержащих по 12 тысяч участников. Для СМИ также существуют биржи статей, на которых можно найти исполнителя для написания статьи по заданной тематике [3–5]. Поиск специалистов на таких биржах производится путем размещения заказов. Заказчики размещают задания, а эксперты выкладывают описания своих компетенций. Специалисты должны быть зарегистрированы на бирже и регулярно просматривать заказы по своей тематике. Такой метод поиска эффективен для поиска специалистов не очень высокой квалификации, которые специализируются на написании текстов для СМИ и интернет-ресурсов с целью заработка. Квалификация специалистов оценивается в большинстве случаев по количеству проданных заказов. Для успешного проведения поиска в таких системах необходимо активное участие в процессе поиска как заказчиков работ, так и самих экспертов.

Существуют специализированные сайты, которые позволяют экспертам обмениваться контактами. Примером служит проект deadline.media, предназначенный для поиска специалистов для СМИ. Каждый зарегистрированный пользователь самостоятельно отмечает темы, в которых он является экспертом, и его контакты будут высвечиваться при поиске по рубрикам. Рубрикатор одноуровневый, содержит около 80 рубрик, например, «наука, инновации и технологии», «происшествия», «культура и искусство». База контактов содержит около 14 тысяч пользователей разной степени актуальности. Также можно упомянуть следующие англоязычные сервисы: www.profnetconnect.com для поиска PR-специалистов; свернутый на настоящий момент проект поиска экспертов в произвольных областях www.allexperts.com; сервис для анализа материалов экспертами-добровольцами

перед публикацией pressfeed.ru. Заинтересованность экспертов в подобной работе может быть не только финансовая. Например, эксперт может быть заинтересован в дополнительном упоминании своего имени в СМИ.

Аналогичные биржи создаются для поиска эксперта при создании проектов, например, сайт для поиска экспертов expertme.ru. Организатор проекта указывает описание проекта и требования к эксперту, а эксперты при поиске указывают свои компетенции и просматривают предложения заказчиков. Оценка квалификации эксперта производится только на основании предоставленного им описания и последующего личного собеседования.

Общими недостатком представленных выше ручных и автоматизированных методов являются необходимость активного участия в поиске обеих сторон и отсутствие критериев оценки квалификации эксперта. Оценка производится либо на основе субъективной оценки небольшого количества других людей, либо на основе активности выполнения работ на поисковом сайте, что никак не связано с реальной квалификацией эксперта.

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, 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.

1. 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] осуществляется поиск по

ключевым словам среди сотрудников университетов США и Канады, зарегистрированных в системе. Контекстный поиск производится только по описаниям компетенций, которые представили сами сотрудники в описании своего профиля. Какой-либо оценки соответствия заявленной тематики и реальной научной деятельности не производится. 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.

Существует значительное количество социальных сетей исследователей, позволяющих загружать свои публикации, обмениваться ими и осуществлять поиск. Проект Academia.edu позволяет загружать в профиль свои статьи и производить поиск по чужим статьям (полноценный полнотекстовый поиск доступен только на платной основе). Проект ResearchGate дополнительно организует совместную работу над статьями. Проект www.bibsonomy.org позволяет загружать и индексировать русскоязычные статьи, однако, в очень небольшом объеме. За 10 лет в нем содержится только 14 русскоязычных статей, содержащих слово «поиск».

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

Проекты первой группы позволяют производить тематический поиск публикаций с последующим ручным анализом информации об авторах. Например, проект CiteSeerX [7] университета штата Пенсильвания содержит около 1 млн документов и позволяет искать статьи по ключевым словам и фамилиям авторов. В профиле каждого автора предоставляется подборка его статей. Проект Seman-

ticscholar [8] содержит 40 млн статей и позволяет осуществлять поиск по ключевым словам. Сайты издательств предоставляют возможность поиска по ключевым словам и библиографическим данным для платного доступа к текстам статей [9–11]. Недостатком таких проектов является отсутствие оценки релевантности автора поисковому запросу. Пользователю необходимо найти статьи, соответствующие запросу, и вручную перебрать всех соавторов для оценки их компетенции.

Проекты второй группы не только позволяют проводить поиск по статьям, но и показывают агрегированные данные по авторам публикаций. Проект Digital Library [12] позволяет производить поиск статей по ключевым словам и показывает при просмотре результатов список авторов всех найденных статей с указанием для каждого автора количества статей. Имеет удобный интерфейс поиска, однако база поиска составляет всего 3 млн статей.

Проект Google Scholar [13] имеет наиболее полный индекс по статьям, в том числе на русском языке. При поиске по ключевым словам показывает количество цитирований статей по своим данным, значительно превышающее количество цитирований по другим системам. Следуя глобальной концепции Google по обучению системы силами самих пользователей, сервис предоставляет пользователю возможность создавать свой профиль и добавлять в профиль статьи из результатов выполнения поискового запроса или с заполнением формы описания статьи вручную. Таким образом, пользователь может добавить в свой профиль и русскоязычные статьи, но информация не верифицируется, и для большинства русскоязычных авторов профили отсутствуют или не заполнены. Оценка соответствия автора тематике запроса также производится вручную перебором всех соавторов найденных работ.

Проект Microsoft Academic Graph (academic.microsoft.com) [14] позволяет проводить поиск экспертов на основе анализа научных публикаций. На основе ключевых слов поискового запроса пользователю предоставляется подборка статей, релевантных запросу. Для сформированной подборки пользователю показываются распределения: по списку наиболее популярных тем для этих статей; по типам публикаций; журналам; научным организациям; конференциям; разделам классификаторов и авторам. Каждый элемент этих списков содержит показатель агрегированной релевантности запросу пользователя и может быть использован

для дополнительной фильтрации. Таким образом, можно выбрать автора или организацию, которые наиболее близки запросу пользователя, и просмотреть список их публикаций. Проект англоязычный, не обрабатывает запросы на русском языке и не анализирует русскоязычные статьи. Коррекция авторами данных о своих работах не предусмотрена.

К третьей группе относятся проекты с возможностью поиска профилей авторов по ключевым словам. Наиболее известным проектом этой группы является китайский проект AMiner [15]. Для проведения тематического анализа в нем используются данные о публикациях авторов, полученных патентах и участии в проектах. Поискový 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. В отдельных вкладках для просмотра доступен перечень статей, проектов и патентов эксперта. Представлены соседние вершины из графа соавторства (Ego Network). Дополнительно представлено распределение количества статей с заданными ключевыми словами по годам. Например, можно видеть, что популярность “K-means” и “Elliptic curves” в открытых научных статьях была довольно высока в период с 2001 по 2011 годы, а потом начала спадать. А популярность “SVM”, наоборот, начала расти после 2008 года.

Наиболее серьезным недостатком перечисленных выше проектов с точки зрения русскоязычных пользователей является полное отсутствие или крайне незначительный объем русскоязычных статей, используемых для анализа. Этот факт особенно важен для экспертов в гуманитарных областях, поскольку результаты научной деятельности многих известных российских ученых гуманитарного

направления мало представлены в англоязычных журналах, индексируемых Web of Science.

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 соавторов, работающих в Московском государственном университете им. М.В. Ломоносова, статьи трех авторов отнесены к организации “Moscow State University”, а двух – к “Michigan State University” [19]. Подобные ошибки определения авторов приводят к значительному ухудшению качества поиска среди русскоязычных авторов. Еще одним недостатком является отсутствие удобной фильтрации по территориальному признаку или языку. Во многих случаях пользователю требуется найти не лучшего эксперта в мире по выбранной тематике, а достаточно хорошего эксперта, с которым он сможет проконсультироваться.

Следует отметить, что проекты AMiner и Microsoft Academic Graph предоставляют пользователям возможность скачать свои данные для самостоятельного анализа: библиографические описания статей (авторы, название, журнал, страницы и другие); ключевые слова; ссылки на pdf-версию и аннотации (при наличии), а также описания авторов. Данные предоставляются только за 2018-й 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), 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 существующих систем и оценки их сильных и слабых сторон можно сформулировать требования, которые предъявляются к системам тематического поиска экспертов в научной сфере. Система должна производить поиск в автоматическом режиме без участия в процессе поиска самих экспертов. Для обеспечения полноты поиска и более объективной оценки квалификации автора 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 специальным тематическим направлениям, например, с использованием для поиска заданных пользователем ключевых слов [21]. Результаты поиска должны автоматически ранжироваться по авторитетности найденных специалистов в заданной тематической области. Исходные данные для осуществления такого поиска можно получать из наукометрических систем.

2. 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 данные, размещенные в наукометрической системе: 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. Заданы 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. Заданы 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. Для улучшения полноты поиска может также использоваться автоматическое расширение набора ключевых слов близкими терминами и переводами [26].

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 [27], 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 тематике, рассматриваемой в работе. 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.

3. 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»

The screenshot shows a search interface with the following elements:

- Search filters:**
 - Ключевые слова (через запятую):
 - Область знаний:
 - Подразделение:
- Search results table:**

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	Занчурич 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».

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

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

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»

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

Автоматический поиск экспертов предполагается использовать для подбора рецензентов в журналы. Кроме того, планируется предоставить пользователям возможность поиска экспертов по ключевым словам на основной странице поиска системы.

ЗАКЛЮЧЕНИЕ

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

СПИСОК ЛИТЕРАТУРЫ

1. Группа «Помогите журналисту».
URL: <http://www.facebook.com/groups/pomogitej>
2. Группа «Герои и эксперты для СМИ».
URL: <http://www.facebook.com/groups/1792146334361293>
3. Биржа копирайтинга. URL: <http://www.textsale.ru>
4. Биржа копирайтинга Адвего. URL: <http://Advego.ru>
5. Биржа контента eTXT. URL: <http://eTXT.ru>
6. Поисковая система Search Engine to Find Experts from Universities.
URL: <http://network.expertisefinder.com>
7. Поисковая система CiteSerX. URL: <http://citeseerx.ist.psu.edu>
8. Поисковая система Semantic Scholar. URL: <http://www.semanticscholar.org>
9. Поисковая система ScienceDirect. URL: <http://www.sciencedirect.com>
10. Издательство Шпрингер. URL: <http://link.springer.com>
11. Поисковая система издательства IEEE. URL: <http://ieeexplore.ieee.org>
12. Поисковая система Digital Library. URL: <http://dl.acm.org>
13. Система цитирования Google Scholar. URL: <http://scholar.google.com>
14. Система поиска авторов Microsoft Academic.
URL: <http://academic.microsoft.com>
15. Система поиска авторов Aminer. URL: <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. 2017. Vol. 1989. 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. URL: <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. Воеводина. 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.

URL: <http://academic.microsoft.com/paper/2986189703>

20. Система цитирования РИНЦ. URL: <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). С. 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. *Shachnev D.A.* Searching for activity results and experts in a given subject area, taking results significance into account // Программная инженерия. 2021. Vol. 12, no. 5. P. 260–266.

27. *Козицын А.С., Афонин С.А., Шачнев Д.А.* Метод оценки тематической близости научных журналов // Программная инженерия. 2020. №6. С. 335–341.

28. *David F. Gleich.* PageRank Beyond the Web // Proc. Society for Industrial and Applied Mathematics. 2015. P. 321–363.

METHOD FOR EXPERT SEARCH USING SCIENTOMETRIC SYSTEM DATA

A. S. Kozitsin¹ [0000-0002-8065-9061], **S. A. Afonin**² [0000-0003-3058-9269]

^{1,2} Institute of Mechanics Lomonosov Moscow State University

¹alexanderkz@mail.ru, ²serg@msu.ru

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*

REFERENCES

1. Gruppa «Pomogite zhurnalistu».
URL: <http://www.facebook.com/groups/pomogitej>
2. Gruppa «Geroi i eksperty dlia SMI».
URL: <http://www.facebook.com/groups/1792146334361293>
3. Birzha kopiraitinga. URL: <http://www.textsale.ru>
4. Birzha kopiraitinga Advego. URL: <http://Advego.ru>
5. Birzha kontenta eTXT. URL: <http://eTXT.ru>
6. Poiskovaia sistema Search Engine to Find Experts from Universities.
URL: <http://network.expertisefinder.com>
7. Poiskovaia sistema CiteSerX. URL: <http://citeseerx.ist.psu.edu>
8. Poiskovaia sistema Semantic Scholar.
URL: <http://www.semanticscholar.org>
9. Poiskovaia sistema ScienceDirect. URL: <http://www.sciencedirect.com>
10. Izdatelstvo Springer. URL: <http://link.springer.com>
11. Poiskovaia sistema izdatelstva IEEE. URL: <http://ieeexplore.ieee.org>
12. Poiskovaia sistema Digital Library. URL: <http://dl.acm.org>
13. Sistema tsitirovaniia Google Scholar. URL: <http://scholar.google.com>
14. Sistema poiska avtorov Microsoft Academic.
URL: <http://academic.microsoft.com>
15. Sistema poiska avtorov Aminer. URL: <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. 2017. Vol. 1989. 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.
URL: <http://academic.microsoft.com/paper/2771441762>
18. *Vasenin V.A., Afonin S.A., Zenzinov A.A. i dr.* Mekhanizmy sistemy ISTINA dlia intellektualnogo analiza sostoianiiia i stimulirovaniia khoda vypolneniia proektov v

sfere nauki i vysshego obrazovaniia // Nauchnyi servis v seti Internet: trudy XXI Vse-rossiiskoi 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.

URL: <http://academic.microsoft.com/paper/2986189703>

20. Sistema tsitirovaniia RINTs. URL: <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-tekhnologicheskie aspekty razrabotki i so-provozhdeniia bolshikh informatsionno-analiticheskikh sistem v sfere nauki i obra-zovaniia // Programmnaia inzheneriia. 2017. T. 8, № 10. S. 448–455.

23. *Afonin S.A., Golomazov D.D., Kozitsyn A.S.* Ispolzovanie sistem seman-ticheskogo analiza dlia organizatsii poiska nauchno-tekhnicheskoi informatsii // Pro-grammnaia inzheneriia. 2012. № 2. S. 29–34.

24. *Kozitsyn A.S., Afonin S.A.* Razreshenie neodnoznachnosti pri opredelenii avtorov publikatsii s ispolzovaniem grafov soavtorstva v bolshikh kolleksiakh biblio-graficheskikh dannykh // Programmnaia inzheneriia. 2017. №8(12). S. 556–562.

25. *Kozitsyn A.S., Afonin S.A.* Algoritm razresheniia neodnoznachnosti imen avtorov v IAS ISTINA // Sovremennye informatsionnye tekhnologii i IT-obrazovanie. 2020. T. 16, №1. S. 108–117.

26. *Shachnev D.A.* Searching for activity results and experts in a given subject area, taking results significance into account // Programmnaia Inzheneriia. 2021. Vol. 12, no. 5. P. 260–266.

27. *Kozitsyn A.S., Afonin S.A., Shachnev D.A.* Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov // Programmnaia Inzheneriia. 2020. №6. S. 335–341.

28. *David F. Gleich.* PageRank Beyond the Web // Proc. Society for Industrial and Applied Mathematics. 2015. P. 321–363.

СВЕДЕНИЯ ОБ АВТОРАХ



КОЗИЦЫН Александр Сергеевич – ведущий научный сотрудник, к. ф.-м. н., окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области информационного поиска и баз данных.

Alexander Sergeevich KOZITSYN – Leading Researcher, Ph. D., graduated from M.V. Lomonosov Moscow State University. Specialist in the field of information retrieval and database.

email: alexanderkz@mail.ru,
ORCID: 0000-0002-8065-9061



АФОНИН Сергей Александрович – ведущий научный сотрудник, к. ф.-м. н., окончил мехмат МГУ им. М.В. Ломоносова. Специалист в области регулярных языков и информационных систем.

Sergey Alexandrovich AFONIN – Leading Researcher, Ph. D., graduated from M.V. Lomonosov Moscow State University. Specialist in the field of regular languages and information systems.

email: serg@msu.ru,
ORCID:0000-0003-3058-9269

Материал поступил в редакцию 21 октября 2021 года