

Методы тематического поиска конференций по наукометрическим данным

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

НИИ механики МГУ им. М.В. Ломоносова

Аннотация. Тематический поиск информации используется в различных сферах деятельности. Применение средств тематического анализа для поиска конференций позволяет увеличивать полноту поиска и охват конференций, способствует расширению круга научного общения молодых ученых и формированию более тесных научных связей. Разработанные авторами алгоритмы позволяют производить поиск с использованием графа соавторства и опорного множества авторов. Такое множество может быть получено с использованием методов тематического поиска экспертов или на основе заданных образцов. Разработанные алгоритмы нечувствительны к языку и учитывают авторитетность конференций в научном сообществе. Апробация проводилась на данных наукометрической системы ИАС ИСТИНА.

Ключевые слова: тематический поиск, конференции, наукометрические системы, граф соавторства, информационные системы, поиск экспертов.

Methods for thematic search of conferences based on scientometric data

A.S. Kozitsyn, S.A. Afonin, D.A. Shachnev

Institute of Mechanics Lomonosov Moscow State University

Abstract. Thematic information search is used in various fields of activity. The use of thematic analysis tools to search for conferences allows you to increase the completeness of the search and coverage of conferences, helps to expand the circle of scientific communication of young scientists and the formation of closer scientific connections. The search algorithms developed by the authors use the co-authorship graph and the reference set of authors. The set can be obtained using methods of thematic search of experts or based on given samples. The developed algorithms are language insensitive and take into account the authority of conferences in the scientific community. Approbation was carried out on the data of the scientometric system IAS ISTINA.

Keywords: thematic search, conferences, scientometric systems, co-authorship graph, information systems, expert search

1. Введение

Тематический поиск информации используется в различных сферах деятельности. Например, тематический поиск экспертов востребован в редакциях при поиске рецензентов, на наукоемких производствах при поиске консультантов или потенциальных исполнителей проектов, в СМИ при поиске авторов для написания статей по заданной тематике. Тематический поиск журналов позволяет на основе аккумулированного опыта научного сообщества автоматически подбирать журналы, соответствующие заданным научным интересам автора, и помогает молодым сотрудникам увеличить свою публикационную активность и повысить показатели цитируемости. Применение средств тематического анализа для поиска конференций позволяет увеличивать полноту поиска и охват конференций, способствует расширению круга научного общения молодых ученых и формированию более тесных научных связей. Следует отметить, что осуществление тематического поиска, в том числе, поиска конференций, требует использования специализированных поисковых информационных систем.

В настоящее время существует большое количество проектов, реализующих в той или иной степени хранение, рубрикацию и поиск конференций.

В Интернет представлена значительная группа проектов, целью которых является продвижение услуг платных публикаций. Такие проекты, как правило, имеют небольшой охват конференций и не предоставляют сервисов качественного поиска. Например, в проекте konferencii.ru зарегистрировано в феврале 2022 года всего 7 конференций в области "математика", причем все конференции "широкого профиля" (например, "Научный форум: технические и физико-математические науки"). Проект "Научные конференции России" (www.kon-ferenc.ru) также представляет конференции "широкого профиля" для быстрой публикации статей (например, "Диссеминация инновационного опыта как фактор модернизации науки и образования"). Проект "научные-конференции.рф" (na-konferencii.ru) содержит около 30 конференций с датой начала в феврале 2022 и не имеет тематического поиска. Проект "Социальная научная сеть" (www.science-community.org) содержит 12 конференций в области математики. Основным недостатком всех проектов этого типа является наличие большого процента фиктивных конференций, которые активно рекламируются для предоставления авторам возможности опубликовать свои работы на платной основе в максимально сжатые сроки (до трех дней). Такой "шум" не только затрудняет поиск, но и отпугивает

организаторов авторитетных конференций, которые не хотят регистрировать свои мероприятия в подобном контексте.

Проект obshestvo.org специализируется на молодежных мероприятиях и содержит описание реальных конференций и олимпиад, но количество зарегистрированных мероприятий также незначительно. За все время в рубрике математика зарегистрировано 126 мероприятий.

Проект konferen.ru позиционируется как календарь научных конференций России и содержит около тысячи планируемых мероприятий: олимпиад; школ; конференций и других. Поиск на сайте возможен по словам из названия и по одной из 39 рубрик.

Проект поиска мероприятий [WorldExpo \(worldexpo.pro\)](http://WorldExpo(worldexpo.pro)) специализируется на представлении коммерческих выставок, форумов, семинаров и, в меньшей степени, конференций.

В Интернет представлены также зарубежные системы поиска конференций, например, проект www.conferencealerts.com предоставляет возможность поиска по рубрике и стране, проект www.allconferences.com, охватывает около 50 тысяч конференций и предоставляет возможность фильтрации по стране, городу, названию и рубрике.

Системы обзора конференций организаций (например, conf.msu.ru в МГУ им. М.В.Ломоносова, www.nstu.ru/science/scientific_events/search НГТУ, www.ruthenia.ru Тарусского университета) специализируются в большей степени на мероприятиях своей организации и имеют небольшой охват в масштабах России.

Существует отдельная категория систем, предназначенных для организации конференций. Большая часть зарегистрированных в этих системах конференций имеет реально действующий программный комитет, рецензируемые доклады, и, как следствие, хороший научный уровень. Например, система [EasyChair \(easychair.org\)](http://EasyChair(easychair.org)), содержит большое количество конференций, но не имеет системы поиска. Российский аналог "Конференции России" (ruconf.ru) имеет систему контекстного поиска и поиска по классификатору, но количество загруженных в нее конференций незначительно (машиностроение -1, медицина -2 и так далее).

На основании приведенного обзора можно сделать вывод, что в России не существует хороших систем тематического поиска конференций, которые бы позволяли производить фильтрацию по конкретной тематике, и учитывали бы при фильтрации авторитетность конференции.

2. Способы тематического поиска

В зависимости от содержащихся в информационных системах данных и от информационной потребности пользователя могут использоваться разные способы осуществления поиска. Самым простым видом поиска является поиск по ключевым словам с использованием

полных текстов докладов конференций. Основным недостатком такого метода поиска является невозможность его проведения при отсутствии таких полнотекстовых данных.

Альтернативным способом поиска является поиск с использованием опорного множества авторов. Опорное множество $H = \{(a, v)\}$, состоящее из пар «автор»-«вес автора», может вычисляться с использованием поиска экспертов по ключевым словам, по выбранному пользователем журналу или конференции.

При использовании поиска экспертов пользователь должен указать множество ключевых слов, наилучшим образом описывающих его информационную потребность. По указанным ключевым словам поисковая система производит подбор экспертов с использованием алгоритмов [1], [2] или других, и составляет взвешенное опорное множество H , которое будет использоваться для дальнейшего поиска.

Во втором случае пользователь указывает образец для поиска в виде журнала или конференции, которые соответствуют выбранной им тематике. Поисковая система на основе этого образца вычисляет взвешенное опорное множество на основе информации об авторах статей или докладов из выбранного источника. Для этого для каждой публикации d каждому соавтору a ставится в соответствие определенный вес по формуле

$$w(a, d) = \frac{1}{2} \delta_{in} + \frac{1}{2k},$$

где n -порядковый номер соавтора в библиографическом списке публикации, а k -количество соавторов публикации a .

Множество весов авторов a вычисляется как максимальное значение его веса по всем публикациям журнала или конференции по формуле

$$v = \max_d (w(a, d)).$$

В опорное множество H включаются авторы, для которых $v > 0$.

Для всех авторов из построенного опорного множества рассматривается множество всех докладов, авторами которых они являются. Степень тематической близости конференции c запросу пользователя определяется на основании количества докладов, входящих в построенное множество, степени участия автора в докладе [3] и весу автора в опорном множестве.

$$F(c) = \sum_{(a,v) \in H} \min(v, \max_{d \in c} (w(a, d)))$$

Таким образом, степень соответствия конференции запросу определяется ее популярностью у других авторов, интересующихся соответствующей тематикой.

3. Программная реализация

Апробация алгоритма производилась на данных наукометрической системы ИСТИНА[4]. Программная реализация трех описанных выше механизмов поиска выполнена на встроенном языке СУБД системы, для разработки интерфейса использовался механизм SQLReport[5].

При поиске по ключевым словам пользователю предлагается указать список ключевых слов, выбрать метод поиска, ограничить по желанию область поиска по рубрикатору ГРНТИ и подразделениям организации. Результатом поиска является список конференций с указанием их соответствия запросу (Рис.1).

Список конференций по теме с поиском экспертов	
Show by <input type="text" value="10"/> items	Search: <input type="text"/>
V	NAME
0.65	XXVIII Международная научная конференция студентов, аспирантов и молодых ученых "Ломоносов 2021"
0.381	XXI Менделеевский съезд по общей и прикладной химии
0.294	Ломоносовские чтения - 2017
0.285	XX Менделеевский съезд по общей и прикладной химии
0.225	Ломоносовские чтения - 2016
0.192	VII Всероссийская Каргинская конференция "Полимеры-2017"
0.155	Научная конференция грантодержателей РНФ «Фундаментальные химические исследования XXI-го века»
0.146	XIX Менделеевский съезд по общей и прикладной химии
0.128	VI Всероссийская Каргинская конференция "Полимеры-2014"

Рис. 1. Поиск с использованием механизмов подбора экспертов

При поиске конференций с использованием эталонного журнала или конференции пользователь сначала должен выбрать один журнал или одну конференцию, которые представляют для него интерес, и перейти по ссылке для поиска похожих по тематике конференций.

Show by 10 items		Search: <input type="text"/>			
N	Конференция	Вес	Количество докладов	Похожие конференции	Похожие журналы
1	Ломоносовские чтения 2021. Секция вычислительная математика и кибернетика, 20-29 апреля 2021(2021)	3,94	136	конференции	журналы
2	ЛОМОНОСОВСКИЕ ЧТЕНИЯ Научная конференция Секция механики(2021)	3,25	155	конференции	журналы
3	XXVIII Международная научная конференция студентов, аспирантов и молодых ученых "Ломоносов 2021"(2021)	2,85	979	конференции	журналы
4	XII Международная научная конференция «Интеллектуальные системы и компьютерные науки»(2021)	1,87	86	конференции	журналы
5	27-я МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ по компьютерной лингвистике и интеллектуальным технологиям Диалог(2021)	1,67	1	конференции	журналы
6	Математические методы распознавания образов (ММРО-20)(2021)	1,58	9	конференции	журналы
7	XXII Международная конференция по вычислительной механике и современным прикладным программным системам (ВМСППС'2021)(2021)	1,58	11	конференции	журналы
8	CAICS 2020: NATIONAL CONGRESS ON COGNITIVE RESEARCH, ARTIFICIAL INTELLIGENCE AND NEUROINFORMATICS(2020)	1,54	19	конференции	журналы
9	VII Актуальные проблемы системной и программной инженерии (АПСПИ 2021)(2021)	1,45	5	конференции	журналы
3rd International Science and Technology Conference «Modern					

Рис. 2. Поиск конференций по журналу

Результаты поиска представляются в виде списка найденных конференций с указанием степени близости к эталонному объекту, а также с возможностью дальнейшего перемещения по графу близости конференций (Рис.2).

Для создания единого информационного пространства поиска граф близости конференций объединен с графом близости журналов [6]. В текущей реализации системы ИСТИНА доступ к поиску и навигации по графу близости журналов и конференций осуществляется из карточки журнала по ссылке «Похожие по тематике журналы». В дальнейшем планируется создать отдельный модуль работы с конференциями и добавить возможность поиска в карточку конференции.

4. Заключение

Разработанные авторами алгоритмы тематического поиска конференций позволяют на основе наукометрических данных производить подбор интересующих конечного пользователя конференций с учетом их авторитетности в научном сообществе. Разработанные алгоритмы нечувствительны к языку и не требуют наличия полнотекстовых описаний докладов и материалов конференций.

Литература

1. Shachnev D. A. Searching for activity results and experts in a given subject area, taking results significance into account // Программная инженерия. 2021. Том. 12, №. 5. P. 260-266.
2. Kozitsin A., Afonin S., Shachnev D. Algorithm for searching experts in scientometric systems // CEUR Workshop Proceedings. Proceedings of the 23rd Conference on Scientific Services & Internet (SSI 2021). Vol. 3066. CEUR, 2022. - P. 59-68.
3. Козицын А.С., Афонин С.А., Шачнев Д.А. Алгоритм поиска по ключевым словам специалистов в заданной предметной области// Современные информационные технологии и ИТ-образование. 2021. Т. 17. № 1. С. 124-133.
4. Васенин В.А., Занчурич М.А., Козицын А.С., Кривчиков М.А., Шачнев Д.А., Архитектурно-технологические аспекты разработки и сопровождения больших информационно-аналитических систем в сфере науки и образования// Программная инженерия. 2017. Том 8, № 10. С. 448-455.
5. Afonin S., Kozitsyn A., Astapov I. Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. 2014. P. 529-534.
6. Козицын А.С., Афонин С.А., Шачнев Д.А. Метод оценки тематической близости научных журналов// Программная инженерия. 2020. № 6. С. 335-341.

References

1. Shachnev D. A. Searching for activity results and experts in a given subject area, taking results significance into account // Programmnaia inzheneriia. 2021. Tom. 12, No. 5. P. 260-266.
2. Kozitsin A., Afonin S., Shachnev D. Algorithm for searching experts in scientometric systems // CEUR Workshop Proceedings. Proceedings of the 23rd Conference on Scientific Services & Internet (SSI 2021). Vol. 3066. CEUR, 2022. - P. 59-68.
3. Kozitsyn A.S., Afonin S.A., Shachnev D.A. Algoritm poiska po kliuchevym slovam spetsialistov v zadannoi predmetnoi oblasti// Sovremennye informatsionnye tekhnologii i IT-obrazovanie. 2021. T. 17. № 1. S. 124-133.
4. Vasenin V.A., Zanchurin M.A., Kozitsyn A.S., Krivchikov M.A., Shachnev D.A., Arkhitekturno-tekhnologicheskie aspekty razrabotki i soprovozhdeniia bolshikh informatsionno-analiticheskikh sistem v sfere nauki i obrazovaniia// Programmnaia inzheneriia. 2017. Tom 8, № 10. S. 448-455.

5. Afonin S., Kozitsyn A., Astapov I. Sqlreports: Yet another relational database reporting system // Proceedings of the 9th International Conference on Software Engineering and Applications. 2014. P. 529-534.
6. Kozitsyn A.S., Afonin S.A., Shachnev D.A. Metod otsenki tematicheskoi blizosti nauchnykh zhurnalov//Programmnaia inzheneriia. 2020. № 6. S. 335-341.