

А. С. Козычн (Москва, МГУ им. М. В. Ломоносова). **Способ уточнения результатов ранжирования при тематическом поиске специалистов.**

УДК 004.912

Резюме: При осуществлении тематического поиска специалистов по библиографическим данным степень соответствия запросу определяется на основе их публикаций без учета авторитетности. В работе рассматривается метод уточнения результатов ранжирования результатов тематического поиска специалистов на основе графа соавторства.

Ключевые слова: граф соавторства, тематический анализ, библиография, наукометрия.

В настоящее время в мире существует большое количество систем, позволяющих производить поиск специалистов. Большинство таких систем аналогичны биржам наукоемких заказов, на которых сами специалисты должны принимать участие в процессе поиска заказов. Такие системы эффективны в сфере бизнеса и СМИ. В научной сфере более востребованы поисковые системы, которые позволяют находить ведущих специалистов по требуемой тематике без необходимости их активного участия в процессе поиска. Источниками необходимой для поиска информации [1] при таком подходе являются системы цитирования и наукометрические системы [2].

Для проведения первичного поиска может использоваться простой контекстный поиск, пространства понятий [3], поиск по онтологиям [4] и по переводам слов [5].

Результатом работы таких алгоритмов является построение отображения $L : A \times Q \rightarrow \mathbb{R}$, где A — множества авторов, Q — множество поисковых запросов. При таком подходе L зависит только от наличия публикаций у каждого автора, и не зависит от наличия связей между авторами.

В настоящей работе представлен метод уточнения результатов ранжирования на основе анализа графа соавторства [6].

Формальная постановка задачи. Дано множество авторов A , граф соавторства $G(A, E)$ и функция поиска по запросу $L : A \times Q \rightarrow \mathbb{R}$. Требуется построить отображение $H : L \times G \rightarrow \mathbb{R}$, уточняющее результаты ранжирования с учетом авторитетности авторов.

Для построения оценки авторитетности используется итерационный алгоритм. На нулевом шаге на основе графа соавторства строится подграф $G^*(V, E)$, где $V = a \in A : L(a, q) > 0$, $E = (v_i, v_j, t_{ij})$, t_{ij} — количество совместных публикаций авторов a_i и a_j . Значение отображения H на нулевом шаге для каждого автора $a_i \in V$ задается равным 1 $h_i^{(0)} = 1$.

Далее на каждом шаге происходит пересчет

$$\hat{h}_i^{(n)} = \sum_j \frac{h_j^{(n-1)} \cdot t_{ij}}{k_j},$$

$$h_i^{(n)} = \frac{\hat{h}_i^{(n)}}{\sum_j \hat{h}_j^{(n)}},$$

где $k_i = \sum_j t_{ij}$.

Процесс пересчета останавливается при условии $|\max_i (h_i^{(n)} - h_i^{(n-1)})| < \delta$, где δ — константа алгоритма.

Скорректированная значимость сотрудника при поиске по ключевым словам определяется как $L \cdot \sqrt{H}$.

Апробация программной реализации алгоритма проводилась на данных системы ИСТИНА [7].

СПИСОК ЛИТЕРАТУРЫ

1. *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. // *Afonin S. A., Golomazov D. D., Kozitsyn A. S.* Ispolzovanie sistemsemanticheskogo analiza dlya organizacii poiska nauchno-tehnicheskoy informacii. — Programmnaya inzheneriya, 2012, № 2, p. 29–34.
2. *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, т. 8, № 10, с. 448–455. // *Vasenin V. A., Zanchurin M. A., Kozitsyn A. S., Krivchikov M. A., Shachnev D. A.* Arhitekturno-tehnologicheskie aspekty razrabotki i soprovozhdeniya bolshih informacionno-analiticheskikh sistem v sfere nauki i obrazovaniya. — Programmnaya inzheneriya, 2017, v. 8, № 10, p. 448–455.
3. *Шундеев А. С.* Об изменении размерности векторного представления текстовых данных. — Программная инженерия, 2019, т. 10, № 6, с. 265–273. // *Shundeev A. S.* Ob izmenenii razmernosti vektornogo predstavleniya tekstovyyh dannyyh. — Programmnaya inzheneriya, 2019, v. 10, № 6, p. 265–273.
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. // *Afonin S. A., Kozitsyn A. S., Shachnev D. A.* Software mechanisms for scientometrical data aggregation based on ontological representation of the relational database structure. — Software Engineering, 2016, v. 7, № 9, p. 408–413. (In Russian.)
5. *Козицын А. С., Афонин С. А., Зензинов А. А.* Алгоритм определения переводов статей с использованием статистических данных. — Электронные библиотеки, 2018, т. 21, № 6, с. 494–505. *Kozitsyn A. S., Afonin S. A., Zenzinov A. A.* Algoritm opredeleniya perevodov statej s ispolzovaniem statisticheskikh dannyyh. — Elektronnyye biblioteki, 2018, v. 21, № 6, p. 494–505.
6. *Козицын А. С., Афонин С. А.* Разрешение неоднозначностей при определении авторов публикации с использованием графов соавторства в больших коллекциях библиографических данных. — Программная инженерия, 2017, т. 8, № 12, с. 556–562. // *Kozitsyn A. S., Afonin S. A.* Razreshenie neodnoznachnostej pri opredelenii avtorov publikacii s ispolzovaniem grafov soavtorstva v bolshih kollekcijyah bibliograficheskikh dannyyh. — Programmnaya inzheneriya, 2017, v. 8, № 12, p. 556–562.
7. *Васенин В. А., Афонин С. А., Козицын А. С., Голомазов Д. Д., Бахтин А. В., Ганкин Г. М.* Интеллектуальная система тематического исследования научно-технической информации (ИСТИНА). — Обозрение прикл. и промышл. матем., 2012, т. 19, № 2, с. 239. // *Vasenin V. A., Afonin S. A., Kozitsyn A. S., Golomazov D. D., Bahtin A. V., Gankin G. M.* Intellektualnaya sistema tematicheskogo issledovaniya nauchno-tehnicheskoy informacii (ISTINA). — Obozrenie prikladnoj i promyshlennoj matematiki, 2012, v. 19, № 2, p. 239–239.

UDC 004.912

Kozitsyn A. S. (Moscow, Lomonosov Moscow State University). **Method for specifying the ranking results in a thematic specialist search.**

Abstract: The relevance of a search query in a thematic search for specialists in bibliographic data is determined by article keywords only. The paper considers a method for specifying the ranking results based on the co-authorship graph.

Keywords: thematic analysis, scientometrics, bibliography, co-authorship graph.