

А. С. Козыцын (Москва, МГУ им. М.В.Ломоносова). **Способ уточнения авторитетности авторов по библиографическим данным.**

УДК 004.912

DOI https://doi.org/10.52513/08698325_2023_30_1_1

Резюме: Точность методов ранжирования результатов информационного поиска по научной тематике во многом зависит от качества алгоритмов определения авторитетности авторов. В работе рассматривается метод уточнения определения авторитетности авторов на основе библиографических данных наукометрических систем.

Ключевые слова: авторитетность, библиография, информационный поиск, наукометрия, ранжирование.

При анализе состояния современной науки используются различные наукометрические показатели. Для оценки авторов наиболее известными показателями являются количество статей, количество цитирований и индекс Хирша. Однако, попытка проведения измерений в любой системе вызывает ответную реакцию системы, искажающую результаты измерений. Например, при измерении градусником температуры воды ее температура меняется в результате процесса теплообмена с градусником. Аналогично, при внедрении наукометрических показателей оценки деятельности авторов научное сообщество отвечает увеличением среднего количества соавторов в статье, созданием коллабораций, договорными взаимными ссылками и другими методами. По данным наукометрической системы «ИСТИНА» [1] среднее количество соавторов после введения системы оценки сотрудников с 2010-го по 2022-й гг. увеличилось на 35% (см. табл. 1).

Таблица 1. Увеличение среднего количества соавторов

2010	2011	2012	2013	2014	2015	2016
2,74	2,88	3,02	3,05	3,03	3,13	3,21
2017	2018	2019	2020	2021	2022	
3,18	3,24	3,34	3,53	3,65	3,7	

Поскольку качество интеллектуального анализа данных, в том числе тематического анализа [2], существенно зависит от качества исходных данных, необходимо использовать алгоритмы очистки исходных данных от подобного «шума».

Одним из возможных показателей для такой очистки может являться позиция автора в библиографических данных публикации. Согласно стандартам библиографии авторы должны располагаться в лексикографическом порядке. Однако на практике первым часто указывается автор, который внес основной вклад в работу над статьей. Для оценки доли таких статей можно использовать статистические данные о распределении авторов.

Обозначим a_k количество статей с k соавторами, r_k — количество статей с первым автором, согласующимся с лексикографическим порядком соавторов. Для случая, когда библиографические данные всех статей сделаны в соответствии с лексикографическим порядком, будет выполняться равенство $r_k = a_k$. Для случая, когда библиографические данные всех статей сделаны в соответствии с вкладом автора, будет выполняться равенство

$$r_k = \frac{a_k}{k}.$$

Тогда для оценки доли статей, которые используют лексикографический порядок, можно использовать величину

$$L = \frac{r_k - \frac{a_k - r_k}{k-1}}{a_k}$$

В работе [3] показано, что доля таких статей для четырех соавторов и более, рассчитанная по библиографическим данным наукометрической системы «ИСТИНА», составляет менее 9%.

В этой связи, при оценке авторитетности автора необходимо учитывать его позицию в библиографических данных статей. Этот показатель должен быть равен 0 для авторов, которые никогда не бывают первыми авторами, 1 для авторов, которые являются первыми авторам в $1/k$ от общего числа их статей с k соавторами, и расти с увеличением доли таких статей. Таким образом, при оценке автора a можно использовать показатель $P(a)$ следующего вида:

$$P(a) = \frac{\sum_{d \in D_a} C(d) |\{d \in D_a \mid \text{ord}(a, d) = 1\}|}{|D_a|^2},$$

где $C(d)$ — количество соавторов статьи, D_a — множество статей автора a , $\text{ord}(a, d)$ — порядковый номер автора a в библиографическом списке статьи d .

Апробация программной реализации алгоритма проводилась на данных системы «ИСТИНА» [1].

Таблица 2. Распределение показателя по должностям

Должность	<0,25	0,25–0,5	0,5–0,75	0,75–1,0	1,0–1,5	1,5–2,0	2,0–3,0	>3,0
Ведущий научный сотрудник	18	9	7	8	26	13	13	6
Ведущий специалист	40	3	4	4	15	9	13	12
Доцент	12	3	4	10	48	13	8	2
Инженер 1-ой категории	28	3	2	4	22	9	20	12
Младший научный сотрудник	16	4	4	7	22	14	19	14
Научный сотрудник	20	6	7	9	22	11	16	9
Преподаватель	7	1	1	12	66	7	4	2
Профессор	12	6	4	7	41	17	10	3
Старший научный сотрудник	19	7	8	8	24	15	13	6
Старший преподаватель	15	2	3	11	55	8	5	1

В табл. 2 для различных должностей представлен результат расчета показателя $P(a)$. В ячейках таблицы указан процент сотрудников, значение показателя для которых попадает в соответствующий диапазон. Как видно из таблицы, наибольшая доля сотрудников имеет показатель в диапазоне $P(a) \in [1, 1, 5]$ и $P(a) \in [0, 0, 25]$. Если для вспомогательного персонала (инженеры и ведущие сотрудники) нахождение значительной доли сотрудников в интервале $[0, 0, 25]$ является обоснованным, то для научных сотрудников это может свидетельствовать о наличии потенциальных проблем в работе.

СПИСОК ЛИТЕРАТУРЫ

1. *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, в. 2, с. 239–240. // *Vasenin V. A., Afonin S. A., Kozitsyn A. S., Golomazov D. D., Bahtin A. V., Gankin G. M.* Intelligent system for research area study of scientific and technical information (ISTINA). — OPPM Surveys Appl. Industr. Math., 2012, v. 19, is. 2, p. 239. (In Russian.)

2. Козицын А. С., Афонин С. А., Шачнев Д. А. Использование методов тематического анализа в наукометрических системах — Электронные библиотеки, 2021, т. 24, No 2, с. 315–338. // *Kozitsyn A. S., Afonin S. A., Shachnev D. A. The use of thematic analysis methods in scientometric systems.* — Russian Digital Libraries J., 2021, v. 24, № 2, p. 315–338. (In Russian.)
3. Козицын А. С., Афонин С. А., Шачнев Д. А. Метод оценки тематической близости научных журналов — Программная инженерия, 2020, т. 11, № 6, с. 335–341. // *Kozitsyn A. S., Afonin S. A., Shachnev D. A. Method for assessing the thematic proximity of scientific magazines.* — Software Engineering, 2020, v. 11, № 6, p. 335–341. (In Russian.)

UDC 004.912

DOI <https://doi.org/10.52513/08698325.2023.30.1.1>

Kozitsin A. S. (Moscow, Lomonosov Moscow State University). **Method for calculating the rank of authors from bibliographic data.**

Abstract: The accuracy of methods for ranking the results of information retrieval on scientific topics largely depends on the quality of the algorithms for determining the rank of authors. The paper considers a method for calculating the rank of authors based on bibliographic data of scientometric systems.

Keywords: authority, bibliography, information retrieval, ranking, scientometrics.