

РАЗРАБОТКА МОДИФИЦИРОВАННОГО МЕТОДА WINNOWERING ДЛЯ АГРЕГИРОВАНИЯ ДАННЫХ БИБЛИОГРАФИЧЕСКОЙ ИНФОРМАЦИИ ИЗ СИСТЕМ ЦИТИРОВАНИЯ В УСЛОВИЯХ НЕПОЛНОЙ ИНФОРМАЦИИ

И.П. Болодурина, Ю.П. Иванова (Луговскова), Л.М. Анциферова, В.Д. Блинов
Оренбургский государственный университет, г. Оренбург, Россия

В настоящее время переход к представлению библиографической информации о научных работах к электронному виду послужил причиной повышенного интереса к наукометрическим исследованиям. При этом существующие наукометрические методы подвергаются критике со стороны ученых, так как неполная библиографическая база и инструменты ее оценивания не позволяют наиболее точно оценить вклад научного труда. Проблема качества наукометрических оценок, как правило, основывается на исследовании данных некоторой системы цитирования, которая не включает полные сведения обо всех публикациях авторов, содержащихся в других системах цитирования. **Цель исследования.** Данное исследование направлено на разработку адаптивного подхода для формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из систем цитирования РИНЦ, «Академия Google» и Scopus. **Методы.** Определение агрегированного списка публикаций для анализа наукометрических показателей проведено методом Winnowing, алгоритмом Левенштейна, методом шинглов и Джаро – Винклера. В рамках экспериментального исследования проведена оценка эффективности применения рассмотренных методов для агрегирования информации систем цитирования на основе анализа точности, полноты и F-меры. **Результаты.** Эксперименты на тестовых данных списка публикаций авторов Оренбургского государственного университета из систем цитирования РИНЦ, «Академия Google» и Scopus показали, что наиболее точные списки публикаций по критерию F-меры сформировал метод Winnowing. Для повышения производительности данного алгоритма проведена двухэтапная оптимизация процесса агрегирования, которая позволила улучшить время работы алгоритма при формировании списка библиографических описаний. **Заключение.** Предложенный подход для формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из систем цитирования РИНЦ, «Академия Google» и Scopus позволяет повысить производительность при формировании списка публикаций авторов и показывает хорошую эффективность при определении наукометрических характеристик авторов.

Ключевые слова: система цитирования, наукометрические методы, агрегирование библиографической информации, модификация метода Winnowing, метод Левенштейна, метод шинглов.

Введение

В настоящее время исследования в области наукометрии стали наиболее актуальной областью для дискуссий среди научного сообщества. Связано это с тем, что существующие системы цитирования научных работ представили библиографическую информацию в электронном виде и стали разрабатывать новые инструменты для удобной работы с ней [1]. При этом существующие наукометрические методы [2, 3], как правило, основываются на неполной библиографической базе и не позволяют наиболее точно оценить вклад научного труда. Однако необходимость оценить хотя бы косвенно качественную сторону публикаций авторов остается и, следовательно, возникает потребность в разработке новых, более объективных алгоритмов формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из систем цитирования РИНЦ, «Академия Google» и Scopus.

Основными недостатками существующих наукометрических методов оценки качества научных публикаций являются:

– проблема отсутствия полной библиографической базы, связанная с публикацией научных работ только в тех системах цитирования, на основании которых они индексируются [4, 5];

– низкие показатели производительности существующих методов агрегирования библиографической информации и отсутствие наиболее точных инструментов оценки [6];

– вольная интерпретация результатов анализа наукометрических показателей [7] при составлении рейтингов на основании какой-либо библиографической базы.

Для решения поставленной проблемы воспользуемся математическим аппаратом нечеткой логики для формирования агрегированного списка библиографических описаний с использованием различных функций хэширования, а также современными технологиями проектирования баз данных и параллельного программирования.

1. Обзор исследований

Исследованиями и разработкой наукометрических методов с целью формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из различных систем цитирования занимаются по всему миру.

В исследовании [8] авторы рассмотрели возможности применения данных двух систем цитирования библиографической информации Scopus и Web of Knowledge при оценивании вклада научного труда автора или некоторой организации. Результаты исследования показали, что наиболее полную на данный момент информацию предоставляет ресурс «Российского индекса научного цитирования», однако не всегда полученные списки библиографических описаний показывают исчерпывающую информацию.

Однако проблема манипуляций с библиометрическими показателями связана с неоправданно широким использованием библиометрических данных различными контролирующими институтами [9, 10]. В связи с этим точность данных, которые получает РИНЦ, напрямую зависит от организации, а также его ответственного представителя.

Авторы статьи [11] провели большой обзор основных наукометрических показателей, позволяющих учитывать количество публикаций и цитирований. Разработан новый подход к учету дополнительной информации по соавторам, по уровню журнала, продолжительности научной карьеры и т. п. Выявлена проблема скрытых и неформальных ссылок, а также ошибок в списке источников. В частности, в работе [12] исследователя А.С. Холодова предложено исключить из рассмотрения импакт-факторы журналов, описывающие ранжирование типов публикаций и не влияющие на значимость научного труда автора.

В публикации [13] представлены данные вычислений и анализа индексов Хирша для преподавателей в области методологии поведенческих наук в Испании. Результаты показывают, что индекс Хирша существенно увеличивается за счет самоцитирования и среднее значение у профессоров существенно не превышает среднее значение показателя у доцентов. Эффективность вычисления индекса Хирша подтверждена в работе [14], где проанализирована возможность учитывать авторство цитируемых статей подробно.

В исследовании [15] проанализированы автоматизированные системы агрегирования цитирований российских ученых, направленные на оптимизацию списка научных работ, их цитирования, а также вычисления наукометрических показателей. Авторы предложили подход, позволяющий не только оптимизировать список библиографических описаний, но и агрегировать результаты из нескольких систем цитирования.

Таким образом, обзор исследований показал, что использование различных подходов к вычислению публикационной активности и агрегированию библиографических данных играет важную роль в оценке качества научного потенциала.

Данное исследование направлено на разработку адаптивного подхода для формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из систем цитирования РИНЦ, «Академия Google» и Scopus.

2. Постановка задачи поиска нечетких публикаций на основе данных систем цитирования

Рассмотрим задачу поиска дубликатов библиографических описаний для анализа сходства научных публикаций в системах цитирования РИНЦ, Scopus и «Академия Google». Стоит отметить, что при анализе схожих публикаций необходимо учитывать:

- 1) порядок слов, транспозиции символов, а также орфографические ошибки;
- 2) наличие одинаковых ФИО у разных авторов;
- 3) согласованность ФИО автора работы;
- 4) отсутствие полного текста научной публикации.

Пусть необходимо определить, является ли библиографическая запись \bar{X}_i и \bar{X}_j дубликатом или нечетким дубликатом. Это означает, что необходимо вычислить меру близости $\rho(\bar{X}_i, \bar{X}_j) \in [0; 1]$, характеризующую степень сходства библиографических описаний запись \bar{X}_i и \bar{X}_j .

Будем считать, что две публикации \bar{X}_i и \bar{X}_j являются *полными дубликатами*, если мера близости $\rho(\bar{X}_i, \bar{X}_j) = 1$. При этом две публикации \bar{X}_i и \bar{X}_j являются *нечеткими дубликатами*, если мера близости $\rho(\bar{X}_i, \bar{X}_j) > \theta$, где θ – пороговое значение для оценки сходства. Как правило, значение θ подбирается экспериментальным путем и является главным недостатком алгоритмов нечеткого вывода.

Задача построения такой меры близости $\rho(\bar{X}_i, \bar{X}_j)$ и определения соответствующего порогового значения θ является задачей поиска дубликатов библиографических описаний для анализа сходства научных публикаций в различных системах цитирования, которую можно решать с помощью синтетических и лексических методов. В рамках данного исследования проанализировано применение специализированных метрик Левенштейна и Джаро – Винклера, метода шинглов и модификации WInnowing.

3. Методы выявления нечетких дубликатов

Метод шинглов

Основная идея *метода шинглов* заключается в том, чтобы представить библиографическое описание в виде последовательностей длиной k (шинглов), которые состоят из соседствующих слов. При этом два описания \bar{X}_i и \bar{X}_j являются нечеткими дубликатами, если множества их шинглов в большей части пересекаются между собой.

Метод Джаро – Винклера

Метод Джаро – Винклера представляет собой меру схожести строк для измерения расстояния между двумя последовательностями символов. Можно считать, что расстояние Джаро между двумя словами (лексемами) вычисляется как минимальное количество односимвольных операций, необходимых для преобразования одного слова в другое.

Метод Левенштейна

Метрика, вычисляющая минимальное количество замен, вставок и удалений, необходимых для преобразования одной строки в другую, называют *метрикой Левенштейна*. При этом анализируемые строки считаются дубликатами, если метрика Левенштейна равна нулю.

Метод WInnowing

Данный алгоритм является модификацией метода шинглов, однако представленные последовательности шинглов хэшируются и разбиваются на «окна» размером $(t - k + 1)$, где t – шумовой порог (минимальная длина подстроки, при которой общие подстроки не игнорируются). После этого из каждого окна выбираются минимальные значения соответствующих хэш-функций, которые сравниваются на основе коэффициента ассоциативности Джаккара.

Критерии оценки качества и эффективности алгоритмов выявления нечетких дубликатов

Для оценки качества и эффективности описанных выше алгоритмов выявления нечетких дубликатов для формирования агрегированного списка библиографических описаний выделим основные критерии качественной оценки в табл. 1.

Критерии оценки выявления нечетких дубликатов

Таблица 1

Evaluation criteria for detecting fuzzy duplicates

Table 1

Наименование критерия	Формула расчёта
Точность	$prec = \frac{\text{истинно положительных}}{\text{положительных}}$
Полнота	$rec = \frac{\text{истинно положительных}}{\text{истинных}}$
F-мера	$F = \frac{2 \cdot prec \cdot rec}{prec + rec}$

Наиболее достоверным критерием оценки качества построенного агрегированного списка библиографических описаний является «F-мера», так как данный критерий одновременно учитывает влияние точности и полноты.

Отметим, что точность вычисляется как отношение количества верно добавленных публикаций в общий список ко всему количеству выбранных публикаций. При этом полнота вычисляется как отношение количества верно добавленных публикаций в общий агрегированный список библиографических описаний к общему количеству истинных публикаций, которые должны были присутствовать в агрегированном списке публикаций.

4. Вычислительные эксперименты

Вычислительные эксперименты, выполненные в работе, проводились на тестовой выборке списка публикаций одного из авторов Оренбургского государственного университета, имеющего достаточное количество публикаций в системах цитирования – 439 (213 – в РИНЦ, 220 – в «Академия Google», 6 – в Scopus).

Результаты работы представленных выше алгоритмов для построения агрегированного списка библиографических описаний с рассчитанными критериями эффективности представлены в табл. 2.

Результаты работы методов выявления нечетких дубликатов

Таблица 2

The results of the methods for detecting fuzzy duplicates

Table 2

Названия метода	Точность	Полнота	F-мера
Метод шинглов	0,918	0,948	0,932
Метод Джаро – Винклера	0,901	0,936	0,923
Метод Левенштейна	0,897	0,979	0,936
Метод Winnowing	0,920	0,957	0,938

Среди рассмотренных методов выявления нечетких дубликатов наиболее эффективным по критерию F-меры является метод Winnowing. В рамках данного исследования представлена его модификация с целью повышения производительности алгоритма и понижения времени работы. В связи с тем, что существующие алгоритмы нечеткого поиска дубликатов требуют значительных вычислительных ресурсов, проведем двухэтапную оптимизацию:

- 1) отказ от канонизации исходного текста библиографического описания;
- 2) проведение распараллеливания алгоритма Winnowing.

Результаты исследований показали, что применение канонизации не позволяет существенно улучшить показатели точности и полноты и, следовательно, позволяет отказаться от этого этапа без ущерба качества реализации. С другой стороны, на данный момент существует множество инструментов повышения скорости работы алгоритмов, таких как специализированная библиотека Numba, которая позволяет автоматически оптимизировать и распараллелить программный код.

Краткие сообщения

Сравнительный анализ времени выполнения классического метода Wnnowing с оптимизированным вариантом на разных выборках публикаций авторов Оренбургского государственного университета представлен на рис. 1.

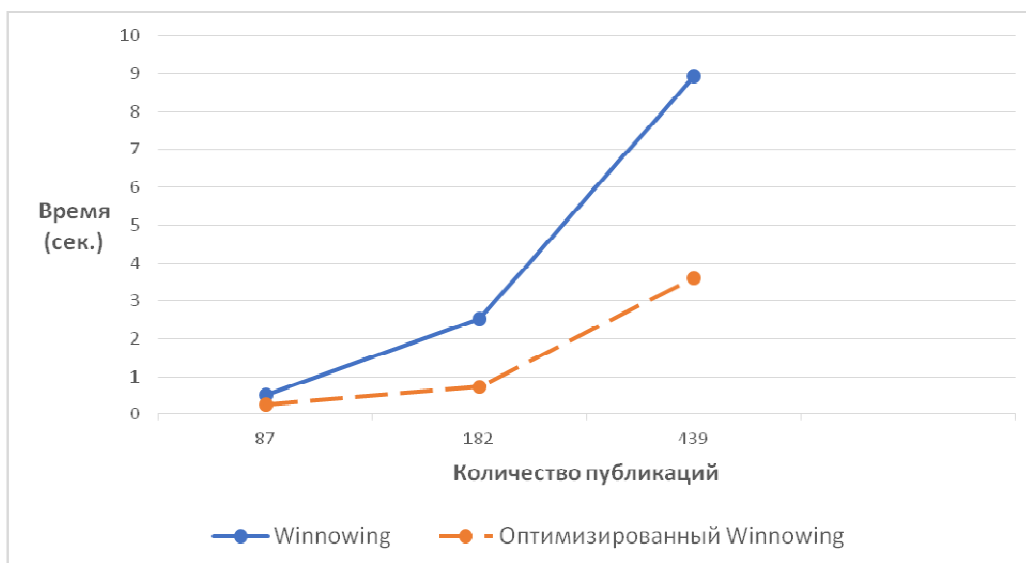


Рис. 1. Сравнительный анализ производительности алгоритмов
Fig. 1. Comparative analysis of algorithms performance

Таким образом, в результате исследования алгоритмов построения агрегированного списка библиографических описаний оптимизированный метод Wnnowing показал приемлемое значение показателя F -меры и наиболее высокую производительность.

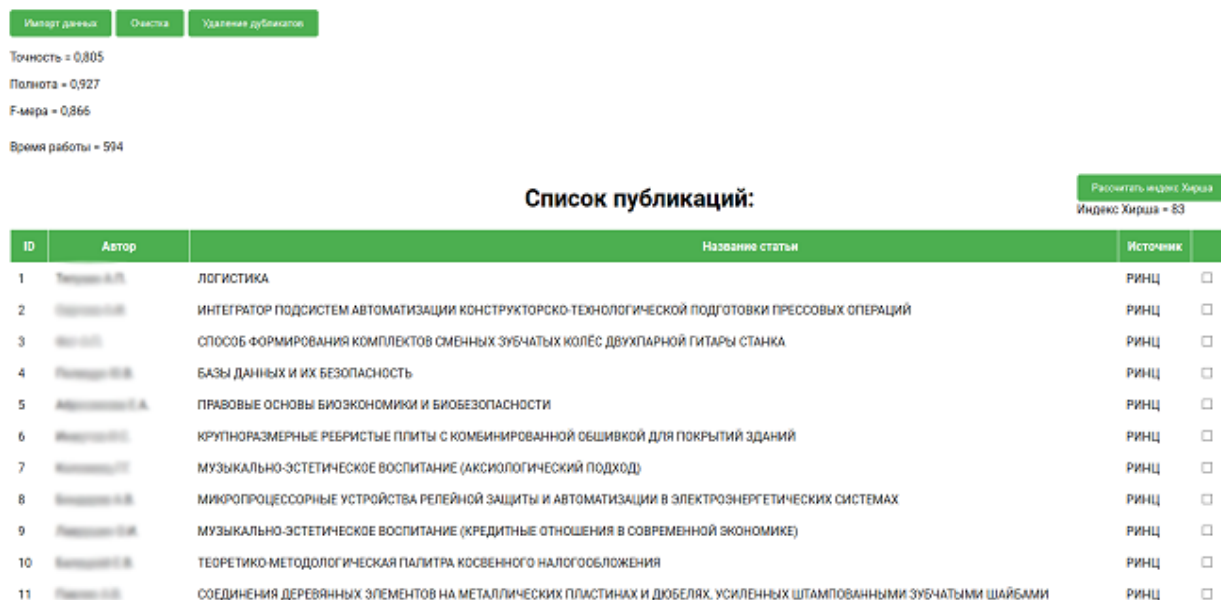


Рис. 2. Автоматизированная система построения агрегированного списка публикаций
Fig. 2. Automated system for building an aggregated list of publications

В результате выполнения представленного алгоритма формируется агрегированный список неповторяющихся научных трудов автора, причём для каждой публикации вычисляется количество цитирований в различных системах. Разработана автоматизированная система (рис. 2), реализующая возможность расчёта основных наукометрических показателей организации на основе библиографических описаний из систем цитирования РИНЦ, Scopus и «Академия Google».

Заключение

В рамках данного исследования разработан адаптивный подход для формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из систем цитирования РИНЦ, «Академия Google» и Scopus.

Определение агрегированного списка публикаций для анализа наукометрических показателей проведено методом Winnowing, алгоритмом Левенштейна, методом шинглов и Джаро – Винклера. В рамках экспериментального исследования проведена оценка эффективности применения рассмотренных методов для агрегирования информации систем цитирования на основе анализа точности, полноты и F-меры.

Эксперименты на тестовых данных списка публикаций авторов Оренбургского государственного университета из систем цитирования РИНЦ, «Академия Google» и Scopus показали, что наиболее точные списки публикаций по критерию F-меры сформировал метод Winnowing. Для повышения производительности данного алгоритма проведена двухэтапная оптимизация процесса агрегирования, которая позволила улучшить время работы алгоритма при формировании списка библиографических описаний

Предложенный подход для формирования агрегированных данных библиографической информации научной организации в условиях неполной информации из систем цитирования РИНЦ, «Академия Google» и Scopus позволяет повысить производительность при формировании списка публикаций авторов и показывает хорошую эффективность при определении наукометрических характеристик авторов.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-01065, а также гранта Президента Российской Федерации для государственной поддержки ведущих научных школ Российской Федерации (НШ-2502.2020.9).

Литература

1. Yang, K. *Citation Analysis: A Comparison of Google Scholar, Scopus, and Web of Science* / K. Yang, L.I. Meho // *Proceedings of the American Society for Information Science and Technology*. – 2006. – Vol. 43, no. 1. – P. 1–15.

2. *Collection statistics for fast duplicate document detection* / A. Chowdhury, O. Frieder, D. Grosman, C. McCabe // *ACM Trans. Inform. Syst.* – 2002. – Vol. 20, no. 2. – P. 171–191.

3. Kotsemir, M.N. *Publication Activity of Russian Researches in Leading International Scientific Journals* / M.N. Kotsemir // *Acta naturae*. – 2012. – Vol. 4, no. 2 (13). – P. 15–35.

4. Мазов, Н.А. Программы для наукометрических и библиометрических исследований: краткий обзор и сравнительный анализ / Н.А. Мазов, В.Н. Гуреев // *Труды 15-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции»*. – 2013. – С. 23–28.

5. Косинов, Д.И. Использование статистической информации при выявлении схожих документов / Д.И. Косинов // *Сборник «Интернет-математика»*. – 2007. – С. 84–90.

6. Baneyx, A. «Publish or Perish» as citation metrics used to analyze scientific output in the humanities: international case studies in economics, geography, social sciences, philosophy, and history / A. Baneyx // *Archivum Immunologiae et Therapiae Experimentalis*. – 2008. – Vol. 56, no. 6. – P. 363–371.

7. Болдырев, П.А. Обзор программных средств в области анализа публикационной активности учёных / П.А. Болдырев, И.Б. Крылов // *XXIII студенческая международная заочная научно-практическая конференция «Молодёжный научный форум: технические и математические науки»*. – 2015. – <http://nauchforum.ru/ru/node/6914> (дата обращения: 01.10.2020).

8. Квелидзе-Кузнецова, Н.Н. Библиометрические показатели как оценочный критерий деятельности преподавателя университета / Н.Н. Квелидзе-Кузнецова, С.А. Морозова // *Universum: Вестник Герценовского университета*. – 2009. – № 4. – С. 38–45.

9. Москалева, О.В. Потери публикаций России: почему и как избежать? / О.В. Москалева // *4-я Международная научно-практическая конференция «Научное издание международного уровня – 2015: современные тенденции в мировой практике редактирования, издания и оценки научных публикаций»*. – 2015. – С. 87–91.

Краткие сообщения

10. Полянин, А.Д. Об индексе Хирша и других наукометрических показателях / А.Д. Полянин // *Научное сообщество*. – 2013. – № 8–9. – С. 20–22.

11. Штовба, С.Д. Обзор наукометрических показателей для оценки публикационной активности учёного / С.Д. Штовба, Е.В. Штовба // *Управление большими системами. Специальный выпуск 44: «Наукометрия и экспертиза в управлении наукой»*. – 2013. – С. 262–278.

12. Холодов, А.С. Об индексах цитирования научных работ / А.С. Холодов // *Вестник Российской академии наук*. – 2015. – Т. 85, № 4. – С. 310–320.

13. Garcia-Perez, M.A. The Hirsch *h* index in a nonmainstream area: methodology of the behavioral sciences in Spain / M.A. Garcia-Perez // *The Spanish Journal of Psychology*. – 2009. – Vol. 12, no. 2. – P. 833–849.

14. Egghe, L. Mathematical theory of the *h*- and *g*-index in case of fractional counting of authorship / L. Egghe // *Journal of the American Society for Information Science and Technology*. – 2008. – Vol. 59, no. 10. – P. 1608–1616.

15. Krylov, I.B. Several characteristics of existing automated systems according to survey of russian scientists publishing activity / I.B. Krylov, P.A. Boldyrev // *Theoretical & Applied Science*. – 2015. – no. 5 (25). – P. 6–9.

Болодурин Ирина Павловна, д-р техн. наук, профессор, заведующий кафедрой прикладной математики, Оренбургский государственный университет, г. Оренбург; prmat@mail.osu.ru.

Иванова (Луговскова) Юлия Петровна, канд. физ.-мат. наук, доцент кафедры прикладной математики, Оренбургский государственный университет, г. Оренбург; ulia_lugovskova@inbox.ru.

Анциферова Лариса Михайловна, канд. пед. наук, доцент кафедры прикладной математики, Оренбургский государственный университет, г. Оренбург; antsiferova_68@mail.ru.

Блинов Владислав Дмитриевич, студент, Оренбургский государственный университет, г. Оренбург; blnvldslv@gmail.com.

Поступила в редакцию 5 октября 2020 г.

DOI: 10.14529/ctcr200413

DEVELOPMENT OF A MODIFIED WINNOWING METHOD FOR AGGREGATING BIBLIOGRAPHIC INFORMATION DATA FROM CITATION SYSTEMS UNDER THE CONDITIONS OF INCOMPLETE INFORMATION

I.P. Bolodurina, prmat@mail.osu.ru,

Yu.P. Ivanova (Lugovskova), ulia_lugovskova@inbox.ru,

L.M. Antsiferova, antsiferova_68@mail.ru,

V.D. Blinov, blnvldslv@gmail.com

Orenburg State University, Orenburg, Russian Federation

Currently, transition to the electronic presentation of bibliographic information about scientific works has caused an increased interest in scientometric research. At the same time, the existing scientometric methods are criticized by scientists, since the incomplete bibliographic base and tools for its assessment do not allow the most accurate assessment of the contribution of scientific work. The problem of the quality of scientometric assessments, as a rule, is based on the study of the data of a certain citation system, which does not include complete information about all publications of the authors contained in other citation systems. **Aim.** This study is aimed at developing an adaptive

approach for the formation of aggregated data of bibliographic information of a scientific organization in conditions of incomplete information from the citation systems of the RSCI, “Google Academy” and Scopus. **Methods.** The definition of the aggregated list of publications for the analysis of scientometric indicators was carried out by the Winnowing method, the Levenshtein algorithm, the shingle method and the Jaro–Winkler method. In the framework of the experimental study, the effectiveness of the application of the considered methods for aggregating information from citation systems was assessed based on the analysis of accuracy, completeness and F-measure. **Results.** Experiments on test data from the list of publications by authors of the Orenburg State University from the citation systems RSCI, Google Academy and Scopus showed that the Winnowing method formed the most accurate lists of publications by the F-measure criterion. To improve the performance of this algorithm, a two-stage optimization of the aggregation process was carried out, which made it possible to improve the running time of the algorithm when generating a list of bibliographic descriptions. **Conclusion.** The proposed approach for the formation of aggregated data of bibliographic information of a scientific organization in conditions of incomplete information from the citation systems of the Russian Science Citation Index, Google Academy and Scopus allows increasing productivity in the formation of a list of authors' publications and shows good efficiency in determining the scientometric characteristics of authors.

Keywords: citation system; scientometric methods; aggregation of bibliographic information; modification of the Winnowing method; Levenshtein method, shingle method.

References

1. Yang K., Meho L.I. Citation Analysis: A Comparison of Google Scholar, Scopus, and Web of Science. *Proceedings of the American Society for Information Science and Technology*, 2006, vol. 43, no. 1, pp. 1–15.
2. Chowdhury A., Frieder O., Grossman D., McCabe C. Collection Statistics for Fast Duplicate Document Detection. *ACM Trans. Inform. Syst.*, 2002, vol. 20, no. 2, pp. 171–191.
3. Kotsemir M.N. Publication Activity of Russian Researches in Leading International Scientific Journals. *Acta Naturae*, 2012, vol. 4, no. 2 (13), pp. 15–35.
4. Mazov N.A., Gureev V.N. [Programs for Scientometric and Bibliometric Research: An Overview and Comparative Analysis]. *Trudy 15-j Vserossijskoj nauchnoj konferencii “Jelektronnye biblioteki: perspektivnye metody i tehnologii, jelektronnye kollekcii”* [Proceedings of the 15th All-Russian Scientific Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections”], 2013, pp. 23–28. (in Russ.)
5. Kosinov D.I. [Use of Statistical Information to Identify Similar Documents]. *Sbornik “Internet-matematika”* [Collection “Internet Mathematics”], 2007, pp. 84–90. (in Russ.)
6. Baneyx A. “Publish or Perish” as Citation Metrics Used to Analyze Scientific Output in the Humanities: International Case Studies in Economics, Geography, Social Sciences, Philosophy, and History. *Archivum Immunologiae et Therapiae Experimentalis*, 2008, vol. 56, no. 6, pp. 363–371.
7. Boldyrev P.A., Krylov I.B. *Obzor programmnykh sredstv v oblasti analiza publikatsionnoj aktivnosti uchenykh* [Review of Software Tools in the Field of Analysis of the Publication Activity of Scientists]. Available at: <http://nauchforum.ru/ru/node/6914> (accessed 01.10.2020).
8. Kvelidze-Kuznetsova N.N., Morozova S.A. [Bibliometric Indicators as an Evaluation Criterion of the Teacher's Activity University]. *Universum: Bulletin of Herzen University*, 2009, no. 3, pp. 38–45. (in Russ.)
9. Moskaleva O.V. [Losses of Russian Publications: Why and How to Avoid it?]. *4-ya Mezhdunarodnaya nauchno-prakticheskaya konferentsiya “Nauchnoe izdanie mezhdunarodnogo urovnya – 2015: sovremennye tendentsii v mirovoy praktike redaktirovaniya, izdaniya i otsenki nauchnykh publikatsiy” (26–29 maya 2015 g., Sankt-Peterburg)* [4th International Scientific and Practical Conference “Scientific publication of the international level – 2015: current trends in world practice of editing, publishing and evaluating scientific publications” (May 26–29, 2015, St. Petersburg)], 2015, pp. 87–91. (in Russ.)
10. Polyanin A.D. [About the Hirsch Index and Other Scientometric Indicators] *Nauchnoe soobshchestvo* [Science community], 2013, no. 8–9, pp. 20–22. (in Russ.)
11. Shtovba S.D. [Review of Scientometric Indicators for Assessing the Publication Activity of a Scientist] *Upravlenie bol'shimi sistemami. Spetsial'nyy vypusk 44: “Naukometriya i ekspertiza v uprav-*

Краткие сообщения

lenii naukoy” [Management of Large Systems. Special Issue 44: “Scientometrics and Expertise in Science Management”], 2013, pp. 262–278. (in Russ.)

12. Kholodov A.S. [About Citation Indexes of Scientific Works]. *Vestnik Rossiyskoy akademii nauk* [Bulletin of the Russian Academy of Sciences], 2015, no. 4, pp. 310–320. (in Russ.)

13. Garcia-Perez M.A. The Hirsch h-Index in a Nonmainstream Area: Methodology of the Behavioral Sciences in Spain. *The Spanish Journal of Psychology*, 2009, vol. 12, no. 2, pp. 833–849.

14. Egghe L. Mathematical Theory of the h- and g-Index in Case of Fractional Counting of Authorship. *Journal of the American Society for Information Science and Technology*, 2008, vol. 59, no. 10, pp. 1608–1616.

15. Krylov I.B., Boldyrev P.A. Several Characteristics of Existing Automated Systems According to Survey of Russian Scientists Publishing Activity. *Theoretical & Applied Science*, 2015, no. 5 (25), pp. 6–9.

Received 5 October 2020

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Разработка модифицированного метода Winnowing для агрегирования данных библиографической информации из систем цитирования в условиях неполной информации / И.П. Болодурина, Ю.П. Иванова (Луговскова), Л.М. Анциферова, В.Д. Блинов // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2020. – Т. 20, № 4. – С. 126–134. DOI: 10.14529/ctcr200413

FOR CITATION

Bolodurina I.P., Ivanova (Lugovskova) Yu.P., Antsiferova L.M., Blinov V.D. Development of a Modified Winnowing Method for Aggregating Bibliographic Information Data from Citation Systems under the Conditions of Incomplete Information. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2020, vol. 20, no. 4, pp. 126–134. (in Russ.) DOI: 10.14529/ctcr200413