

¹Русанова А.А.
ollyvost@tpu.ru

¹Гуменников И.В.
gumennikov@tpu.ru

¹Касати Ф., PhD
kasati@tpu.ru

²Горев В.П., д-р эконом. наук
gorev_vladimir@mail.ru

¹Национальный исследовательский Томский
политехнический университет,
Россия, Томск

²Байкальский государственный университет,
Россия, Иркутск

DOI: 10.24153/2079-5912-2019-10-6-56-61

КРАУДСОРСИНГ КАК МЕТОД НАУЧНЫХ ИССЛЕДОВАНИЙ

Аннотация: В силу своей специфики, краудсорсинговая модель поиска, анализа, сбора и интерпретации информации всегда являлась более комплементарной для сферы бизнеса, где требуется сокращение затрат на обработку большого объема данных множеством участников процесса. Но в последние годы данная методология начала распространяться и на научно-исследовательскую область, что, несомненно, имеет свои положительные и отрицательные стороны. Целью данного исследования является критический анализ такого современного метода исследования как краудсорсинг применительно к научным исследованиям. Объектом исследования является метод в области анализа больших объемов данных на примере исследований научной литературы (использованы реферативные базы данных Scopus и Web of Science). Предметом исследования выступает раскрытие основных характеристик краудсорсингового метода в области анализа научной литературы. Осуществленный разбор основ крауд-вычислений как способа, демонстрирует недостаточную проработку базовых принципов, позволяющих проводить верификацию и стабилизировать результаты исследований, полученных с помощью краудсорсинга. Но при этом, системность применения подобного метода в зарубежных экономических, социологических и иных исследованиях демонстрирует заинтересованность научного сообщества в разработке подобного междисциплинарного подхода, позволяющего минимизировать затраты по проектной и иной деятельности. Научная новизна представленной работы заключается в том, что в рамках отечественной научной мысли практически отсутствуют работы, посвященные рассмотрению метода крауд-вычислений в области социально-экономического анализа. Выводом является утверждение, согласно которому вопрос практического применения указанного инструмента, как метода работы с количественными и качественными данными, может быть реализован при условии выработки устойчивой схемы верификации получаемых данных.

Ключевые слова: краудсорсинг, научные исследования, метод, анализ данных, наукометрические базы.

Исследование выполнено за счет гранта Российского научного фонда (проект №19-18-00282)

Обзор научной литературы занимает важное место и лежит в основе любого исследования вследствие того, что он создает определенный корпус «предыдущих» знаний, накопленных научными коллективами и учеными относительно исследуемой объектно-предметной области. К примеру, М. Бэйкер справедливо утверждал, что создание, как и эволюция нового знания, главным образом происходит посредством таких процессов как аккумуляция и накопление [6]. Именно анализ предшествующих источников, посвященных той или иной проблеме, составляет суть научной новизны исследования и минимизирует разногласия и конфликты интересов, которые могут впоследствии возникнуть при публикации результатов научных изысканий. С развитием электронной системы реферативных баз данных (РИНЦ, Scopus, Web of Science, Google Scholar) произошло множество изменений в научной среде, благодаря которым современный научный коллектив уже не обязан тратить большое количество времени на поиск искомой литературы, а имеет возможность воспользоваться электронным поиском открытых источников, используя запрос в формате «TITLE-ABS-KEY» [12].

Анализ данных по искомой области знаний позволяет ученому, во-первых: 1) иметь представление о глубине изученности данного вопроса с позиции других авторов в мировом сообществе; 2) минимизировать плагиат, так как данная тема уже может быть изучена рядом авторов и данные будут не репрезентативны, что позволит минимизировать конфликты интересов при использовании ссылок. В связи с этим можно заключить, что обзор научной литературы является самым важным аспектом проведения любых научных исследований и фундаментирует репрезентативность и всестороннюю изученность вопроса [7], [18].

К примеру, авторам статьи, в процессе работы над действующим грантом РНФ в области сравнительного анализа оценки финансового неравенства в рамках социально-экономического положения людей старшего поколения, пришлось столкнуться с проблемой синтеза большого объема данных: только по запросу «неравенство» в «экономике» реферативной базой Web of Science за 2017 год было выдано более 1500 результатов, а за последние 5 лет, по данной теме, было опубликовано более 6200 научных статей, не меньший показатель выдала база наукометрических данных Scopus. Таким образом, процесс обработки научных результатов растягивался во времени экспоненциально вследствие необходимости анализа первоначальной выборки и систематизации полученных данных с целью их дальнейшего использования.

В связи с этим, в области практических и фундаментальных исследований, следует актуализировать поиск наиболее релевантного метода изучения научной ли-

тературы, которые поможет проанализировать и систематизировать большой объем наукометрических баз с целью, прежде всего, экономии времени научного коллектива [13], [10], а также анализа тех источников, которые могут быть утеряны или не учтены в процессе работы [16]. В данном случае многие зарубежные ученые, в последнее время, стали прибегать к методу крауд-вычислений (в формате краудсорсинга) под которым понимается использование интеллектуальных усилий целого коллектива людей, которые взаимодействуют посредством информационно-коммуникационных технологий (Интернет). В настоящее время краудсорсинг находит свое применение в большом спектре человеческой деятельности (науке [17], [5], бизнесе [19], в городском и муниципальном управлении [11], [15]).

И, несмотря на то, что данный метод активно набирает обороты, базовые принципы краудсорсинга еще находятся на стадии формирования [2]. В работах [4], [14] зарубежных исследователей ранее уже были проанализированы крауд-вычисления применительно к анализу большого количества научной литературы (от 700 источников и более). Но в них не было отражено какими именно способами возможна минимизация ошибок в самом процессе репрезентации результатов. В работах Е. Кривошеева [9], как и А. Нгуена [3] была попытка спроектировать систему, основанную на крауд-вычислениях и машинном обучении для минимизации затрат при использовании двух видов методов: краудсорсинга и анализа литературы с помощью привлеченных экспертов. Но эти работы в большей части относятся к системам искусственного интеллекта и алгоритмизации, что, несомненно, является важным фактором при расчете показателей результативности метода, но не входит в круг наших интересов.

Если мы говорим о том, что краудсорсинг может считаться методом научных исследований, то, прежде всего, следует определить что является «методом». Методом (греч. – «способ познания») именуется набор теоретических и практических способов освоения действительности, это целая система регулирования принципов познавательной деятельности. Метод также может быть определен характером исследуемого объекта, т.к. неизменным образом метод связан с субъектно-объектной средой.

С целью проведения критического анализа такого современного метода исследования научной литературы, как краудсорсинг, следует раскрыть основные составляющие данного способа. Используя краудсорсинг, исследователь имеет недорогой доступ к географически и демографически разнородной группе участников, с помощью которых возможно собирать информацию в сжатые сроки. Задачи, которые ранее исполнялись небольшой группой исследователей (анализ литературы по панельным исследованиям,

нарративного интервью, группового интервьюирования, работа по сбору статистической информации и т.д.) теперь могут быть распараллелены между людьми благодаря интернет-платформам по методике краудсорсинга [8]. Краудсорсинговые платформы разрабатываются в основе своей под зарубежными транснациональными брендами: Amazon Mechanical Turk (далее по тексту - AMT) и CrowdFlower, Universal Human Relevance System (UHRS) от Microsoft или EWOK от Google, также на российском рынке набирает обороты российская компания, стремящаяся занять свою нишу в области крауд-вычислений: «Яндекс.Толока».

Суть данных платформ заключается в том, что используя некоторую систему критериальных показателей, которые задаются исследователем/компанией, проставляются коэффициенты выполненных/не выполненных/ошибочно индексированных элементов, за которые потом пользователи, откликнувшиеся на данное исследование/задачу, получают зарплату (как правило, не большую, в США-порядка нескольких долларов за выполненную часть задания). К примеру, исследовательской группе необходимо провести качественный анализ источников литературы по проблеме экономического неравенства, размещенных в зарубежных библиографических базах (Scopus, Web of Science). Используя запросы в данных базах, пользователю выдается количество статей с кратким содержанием статьи и ключевыми словами. Далее, на базе краудсорсинговой платформы, (к примеру, АТМ) формируется запрос на выполнение заданий по отбору статей согласно указанным «Заказчиком» параметрам. Запрос должен содержать максимально-понятное описание задания для пользователей краудсорсинговых платформ, так как не все пользователи АМТ имеют высшее образование, но при этом, все имеют «равный статус», не разделяясь по урону образования, опыту в аналитической работе и прочее.

Вследствие всего вышеперечисленного следует отметить ряд недостатков данного метода научного исследования, который зачастую упускается из вида современными исследователями:

1. В силу того, что далеко не все участники крауд-платформ являются специалистами, возникает необходимость в применении схем верификации для управления качеством получаемых результатов (к примеру, мажоритарное голосование, эталонные задания или применение частичной экспертной оценки). Т.е. если и выборочно, но результаты, выдаваемые пользователями таких платформ, должны проходить некоторую валидацию.

2. Ограниченность «географии использования» платформ (AMT, CrowdFlower). Самые популярные платформы работают исключительно с лицами, являющимися

гражданами США. Поэтому в распоряжении ученой группы может стоять только анализ литературы по источникам Scopus, WoS, и при этом, работы, опубликованные в РИНЦ, не будут включены в выборку.

3. Российская платформа «Яндекс.Толока» занимается развитием собственной площадки для сбора и обработки больших массивов данных с привлечением сторонних лиц для выполнения типовых заданий (обработки изображений, рутинной работы). Эта платформа действует на российском рынке и основным ее плюсом является возможность некоторого «отбора» исполнителей, также можно описать свойства («навыков») исполнителей с помощью API в формате JSON («текстовый формат обмена данными, основанный на JavaScript» - прим.авт.). Но, при этом, расчет и пополнение «Яндекс.Кошелек» для расчетов с исполнителями ведется только в валюте Долларов США.

Таким образом, можно заключить, что крауд-вычисления могут способствовать расширению методологии исследователя в будущем, но только при условии доработки существующих моделей, выработки верификационных алгоритмов и схем, расширению географии пользователей платформ и привязки расчетов к отечественной валюте. Стоит подчеркнуть, что в российской научной среде, в настоящий момент, практически отсутствуют публикации, посвященный сбору и анализу, а также выявлению методов обеспечения качества в крауд-вычислениях, за исключением работ А.В. Пономарева [1], что отражает перспективу дальнейшего исследования метода краудсорсинга.

СПИСОК ИСТОЧНИКОВ

1. Пономарев А. В. Методы обеспечения качества в системах крауд-вычислений: аналитический обзор. Тр. СПИИРАН, 54, 2017. С. 65-78.
2. Пономарев А.В. Онтология для описания приложений, использующих элементы крауд-вычислений // Кибернетика и программирование. – 2018. – № 3. – С. 25 - 37. DOI: 10.25136/2306-4196.2018.3.26556 URL: https://nbpublish.com/library_read_article.php?id=26556
3. An T. Nguyen, Byron C Wallace, and Matthew Lease. 2015. Combining Crowd and Expert Labels using Decision Theoretic Active Learning. Proceedings of the 3rd AAI Conference on Human Computation (HCOMP) (2015), pp.120–129.
4. Andrew W. Brown and David B. Allison. 2014. Using Crowdsourcing to Evaluate Published Scientific Literature: Methods and Example. Plos One 9, 7 (2014) // <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100647> (дата обращения: 16.10.2019).
5. Baev V., Sablok G., Minkov I. Next generation sequencing crowd sourcing at BIOCAMP: What promises it holds for us in future? // J. Comput. Sci. Elsevier B.V., 2014. Vol. 5, № 3. pp. 325–326.
6. Baker M. J. (2000) Writing a Literature Review // The Marketing Review; vol. 1, no. 2, pp. 219-247.
7. Earl Steinberg, Sheldon Greenfield, Dianne Miller Wolman, Michelle Mancher, Robin Graham, et al. (2011) Clinical practice guidelines we can trust. National Academia Press.
8. Edith Law, Krzysztof Z. Gajos, Andrea Wiggins, Mary L. Gray, and Alex Williams. (2017) Crowdsourcing as a tool for research: Implications of uncertainty. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, pp. 1544–1561.
9. Krivosheev E., Boualem Benatallah Bahareh Harandizadeh, and Fabio Casati (2018) Crowd-based Multi-predicate Screening of Papers in Literature Reviews (poster). In Proceedings of WWW2018. International World Wide Web Conferences Steering Committee.
10. Krivosheev E., Valentina Caforio, Boualem Benatallah, and Fabio Casati (2017) Crowdsourcing Paper Screening in Systematic Literature Reviews. In Procs of Hcomp2017. AAI.
11. Fraternali P. et al. Putting humans in the loop: Social computing for Water Resources Management // Environ. Model. Softw. Elsevier Ltd, 2012. Vol. 37. pp. 68–77.
12. Haidich A. (2010) Meta-analysis in medical research. Hippokratia 14, Suppl 1, pp. 29-37.
13. JPT Higgins and S. Green. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. Available from www.handbook.cochrane.org
14. Michael L. Mortensen, Gaelen P. Adam, Thomas A. Trikalinos, Tim Kraska, and Byron C. Wallace. 2016. An exploration of crowdsourcing citation screening for systematic reviews. Research Synthesis Methods (2016). RSM-02-2016-0006.R4.
15. Nunes A. a., Galvão T., Cunha J.F.E. Urban Public Transport Service Co-creation: Leveraging Passenger's Knowledge to Enhance Travel Experience // Procedia-Soc. Behav. Sci. Elsevier B.V., 2014. Vol. 111. pp. 577–585.
16. Perrine Créquit, Ludovic Trinquart, Amélie Yavchitz, and Philippe Ravaud (2016) Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. BMC Medicine 14, 1 (2016), 8. <https://doi.org/10.1186/s12916-016-0555-0>.
17. Wechsler D. Crowdsourcing as a method of transdisciplinary research-Tapping the full potential of participants // Futures. Elsevier Ltd, 2014. Vol. 60. pp. 14–22.
18. Yalin Sun, Pengxiang Cheng, Shengwei Wang, Hao Luy, Matthew Lease, Iain Marshall, and Byron C. Wallace (2016) Crowdsourcing Information Extraction for Biomedical Systematic Reviews. In 4th AAI Conference on Human Computation and Crowdsourcing (HCOMP): Works-In-Progress Track: <http://arxiv.org/abs/1609.01017> 3 pages. arXiv:1609.01017.
19. Ye H. (Jonathan), Kankanhalli A. Investigating the antecedents of organizational task crowdsourcing // Inf. Manag. Elsevier B.V., 2015. Vol. 52, № 1. pp. 98–110.

Rusanova A.A.

National Research Tomsk Polytechnic University,
Russia, Tomsk
ollyvost@tpu.ru

Gumennikov I.V.

National Research Tomsk Polytechnic University,
Russia, Tomsk
gumennikov@tpu.ru

Casati F., PhD

National Research Tomsk Polytechnic University,
Russia, Tomsk
kasati@tpu.ru

Gorev V.P.

Doctorate in Economics
Baikal State University,
Russia, Irkutsk
gorev_vladimir@mail.ru

CROWDSORING AS A METHOD OF SCIENTIFIC RESEARCH

The study was supported by the Russian Science Foundation
(project №19-18-00282)

Due to its specificity, the crowdsourcing model of searching, analyzing, collecting and interpreting information has always been more convenient for the business sector. But recently, this methodology has begun to extend into the research field, which undoubtedly has its positive and negative sides. The purpose of this study is a critical analysis of such modern research method as crowdsourcing in relation to scientific research. The object of research is a method in the field of analyzing large amounts of data using the example of scientific literature research (databases Scopus and Web of Science were used). The subject of the study is the disclosure of the main characteristics of the crowdsourcing method in the field of analysis of scientific literature. The analysis of the basics of crowd-computing as a method demonstrates an insufficient study of the basic principles that allow verification and stabilization of the results of studies obtained using crowdsourcing. But at the same time, the systematic application of this method in foreign economic, sociological and other studies demonstrates the interest of the scientific community in developing such an interdisciplinary approach that allows minimizing the costs of project and other activities. The scientific novelty of the presented work lies in the fact that within the framework of Russian scientific thought there are practically no works devoted to the consideration of the method of crowd-computing in the field of socio-economic analysis. The conclusion is the statement, according to which the question of the practical application of this tool, as a method of working with quantitative and qualitative data, can be implemented by provided and stable verification scheme for the received data.

Keywords: crowdsourcing, research, method, data analysis, scientometric databases.

Literature:

1. Ponomarev A. V. Metody obespecheniya kachestva v sistemakh kraud-vychisleniy: analiticheskiy obzor. Tr. SPIIRAN, 54, 2017. 65-78 s.
2. Ponomarev A.V. Ontologiya dlya opisaniya prilozheniy, ispol'zuyushchikh elementy kraud-vychisleniy // Kibernetika i programmirovaniye. – 2018. – № 3. – S. 25 - 37. DOI: 10.25136/2306-4196.2018.3.26556 URL: https://nbpublish.com/library_read_article.php?id=26556
3. An T. Nguyen, Byron C Wallace, and Matthew Lease. 2015. Combining Crowd and Expert Labels using Decision Theoretic Active Learning. Proceedings of the 3rd AAAI Conference on Human Computation (HCOMP) (2015), pp. 120–129.
4. Andrew W. Brown and David B. Allison. 2014. Using Crowdsourcing to Evaluate Published Scientific Literature: Methods and Example. Plos One 9, 7 (2014) // <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100647> (data obrashcheniya: 16.10.2019).
5. Baev V., Sablok G., Minkov I. Next generation sequencing crowd sourcing at BIOCOMP: What promises it holds for us in future? // J. Comput. Sci. Elsevier B.V., 2014. Vol. 5, № 3. pp. 325–326.
6. Baker M. J. (2000) Writing a Literature Review // The Marketing Review; vol. 1, no. 2, pp. 219-247.
7. Earl Steinberg, Sheldon Greenfield, Dianne Miller Wolman, Michelle Mancher, Robin Graham, et al. (2011) Clinical practice guidelines we can trust. National Academia Press.
8. Edith Law, Krzysztof Z. Gajos, Andrea Wiggins, Mary L. Gray, and Alex Williams. (2017) Crowdsourcing as a tool for research: Implications of uncertainty. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, pp. 1544–1561.

9. Krivosheev E., Boualem Benatallah Bahareh Harandizadeh, and Fabio Casati (2018) Crowd-based Multi-predicate Screening of Papers in Literature Reviews (poster). In Proceedings of WWW2018. International World Wide Web Conferences Steering Committee.
10. Krivosheev E., Valentina Caforio, Boualem Benatallah, and Fabio Casati (2017) Crowdsourcing Paper Screening in Systematic Literature Reviews. In Procs of Hcomp2017. AAAI.
11. Fraternali P. et al. Putting humans in the loop: Social computing for Water Resources Management // Environ. Model. Softw. Elsevier Ltd, 2012. Vol. 37. pp. 68–77.
12. Haidich A. (2010) Meta-analysis in medical research. Hippokratia 14, Suppl 1, pp. 29–37.
13. JPT Higgins and S. Green. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. Available from www.handbook.cochrane.org
14. Michael L. Mortensen, Gaelen P. Adam, Thomas A. Trikalinos, Tim Kraska, and Byron C. Wallace. 2016. An exploration of crowdsourcing citation screening for systematic reviews. Research Synthesis Methods (2016). RSM-02-2016-0006.R4.
15. Nunes A. a., Galvão T., Cunha J.F.E. Urban Public Transport Service Co-creation: Leveraging Passenger's Knowledge to Enhance Travel Experience // Procedia-Soc. Behav. Sci. Elsevier B.V., 2014. Vol. 111. pp. 577–585.
16. Perrine Créquit, Ludovic Trinquart, Amélie Yavchitz, and Philippe Ravaud (2016) Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. BMC Medicine 14, 1 (2016), 8. <https://doi.org/10.1186/s12916-016-0555-0>.
17. Wechsler D. Crowdsourcing as a method of transdisciplinary research-Tapping the full potential of participants // Futures. Elsevier Ltd, 2014. Vol. 60. pp. 14–22.
18. Yalin Sun, Pengxiang Cheng, Shengwei Wang, Hao Luy, Matthew Lease, Iain Marshall, and Byron C. Wallace (2016) Crowdsourcing Information Extraction for Biomedical Systematic Reviews. In 4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP): Works-In-Progress Track: <http://arxiv.org/abs/1609.01017> 3 pages. arXiv:1609.01017.
19. Ye H. (Jonathan), Kankanhalli A. Investigating the antecedents of organizational task crowdsourcing // Inf. Manag. Elsevier B.V., 2015. Vol. 52, № 1. pp. 98–110.