

# **РАЗРАБОТКА АГРЕГИРУЮЩЕЙ СИСТЕМЫ АНАЛИЗА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ УЧЁНЫХ НА ОСНОВЕ МЕЖДУНАРОДНЫХ И РОССИЙСКОЙ СИСТЕМ ЦИТИРОВАНИЯ В УСЛОВИЯХ ОГРАНИЧЕННОГО ДОСТУПА**

**Болдырев П.А., Крылов И.Б.**

**ФГБОУ ВО «Оренбургский государственный университет», г. Оренбург**

Наука и органы управления наукой и ее финансирование сейчас остро нуждаются в надежной, доступной для измерения и сравнения информации о научных результатах, о результативности вложений в науку и в конкретные научные исследования [1]. Сегодня для оценивания результативности научной деятельности совместно с экспертными заключениями все чаще используются наукометрические показатели [2,3]. Возросший интерес к наукометрическим показателям вызван в первую очередь возможностью автоматизации процесса оценивания деятельности ученых с использованием открытых реферативных баз данных Web of Science, Scopus, РИНЦ. Дешевизна и быстрота проверки, а также отсутствие человеческого фактора обуславливают популярность наукометрических показателей в экспресс-оценивании научной деятельности ученых. Пороговые значения по наукометрическим показателям представляют собой некий фильтр, который отсеивает слабых кандидатов, тем самым сокращает затраты на проведение дорогостоящего и трудоемкого экспертного оценивания качества научных результатов, а также позволяет выявить перспективные направления развития науки [4,5]. Особую актуальность данный вопрос приобретает в условиях ограниченного доступа к системам цитирования, обусловленного высокой ценой доступа к информации<sup>1</sup>.

Целью разработки является проектирование информационно-аналитической IT-платформы, агрегирующей данные о научной деятельности учёных на основе использования библиометрических и наукометрических показателей международных и российской реферативных баз данных в условиях ограниченного доступа: Web of Science, Scopus, РИНЦ. Формируемая база данных позволит оценить результативность научной деятельности, провести анализ объединенных данных из нескольких систем цитирования на основе критериев оценки качества в условиях неполной определённости. В данном случае под условиями неполной определённости понимается отсутствие строго типизированного и полного набора данных в различных системах цитирования в условиях ограниченного доступа. Это поможет поддержать или более точно идентифицировать любые расхождения при анализе наукометрических показателей и дать независимую оценку деятельности ученого. Предлагаемую систему также возможно использовать для оценки фундаментальных исследований, результаты которых

---

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 16-47-560973 и 16-47-560390

непосредственно не связаны с экономическим эффектом, через отзыв сообщества на публикации с результатами исследования. Формально этот отзыв выражают индексом цитирования.

Сегодня в наукометрии используется три главных показателя: индекс цитируемости ученого, индекс Хирша, импакт-фактор. Сетевая модель позволит интегрировать публикационные и цитатные показатели по всей вертикали социального института науки: от научного сотрудника-автора, структурного подразделения и учреждения, где работает круг авторов, до министерств и ведомств. Такого рода статистические сведения, в свою очередь, помогут проводить объективную оценку деятельности ученых [6].

На сегодняшний день существует большое количество международных и национальных источников информации, предоставляющие информацию об исследователях, данные об участии в научных исследованиях и публикационной активности учёных – так называемые системы цитирования. Национальные системы научного цитирования успешно функционируют в большинстве развитых и развивающихся стран, таких как Россия (РИНЦ [7]), Индия, Китай, Иран, Испания, Япония и другие. Помимо национальных систем цитирования можно выделить международные, среди которых наиболее крупными по количеству изданий и универсальными, является база данных Scopus [8] издательской корпорацией Elsevier и Web of Science [9] компании Thomson Reuters.

Тем не менее, не смотря на существование как международных систем цитирования, так и национальных, представляющих широкие возможности для анализа публикационной активности учёных, зачастую их возможностей в отдельности оказывается недостаточным для получения объективной и полной картины о публикационной активности российских учёных.

Системы цитирования [2] и [3] не включают большинство публикаций на русском языке. Система цитирования [1] не отражает полной картины публикационной активности авторов, так как не имеет доступа к большому количеству зарубежных публикаций, имеет недочёты при учёте цитирований, а также не включает большинство работ до 2000 года.

Для устранения недостатков, не позволяющих получить объективную и полную картину о публикационной активности российских учёных, предпринимались попытки по созданию программных средств, которые предоставляют дополнительные возможности по работе с системами цитирования. К числу авторов, занимающихся данным вопросом можно отнести Гуреева В.Н., Колесникову Т.А., Мазова Н.А., Матвееву Е.В., Силину А.Ю., Yang K., Meho L.I., Couto F.M., A. Baneyx, P. Jacsó и других.

В процессе аналитического обзора научно-технической литературы в области создания программных средств, которые предоставляют дополнительные возможности по работе с системами цитирования, было выявлено 2 основных признака, позволяющих провести классификацию разработанных на данный момент программных средств: источник исходных данных; характер проводимого анализа. Классификация существующих на

данный момент программных средств на основе выделенных факторов представлена на рисунке 1.

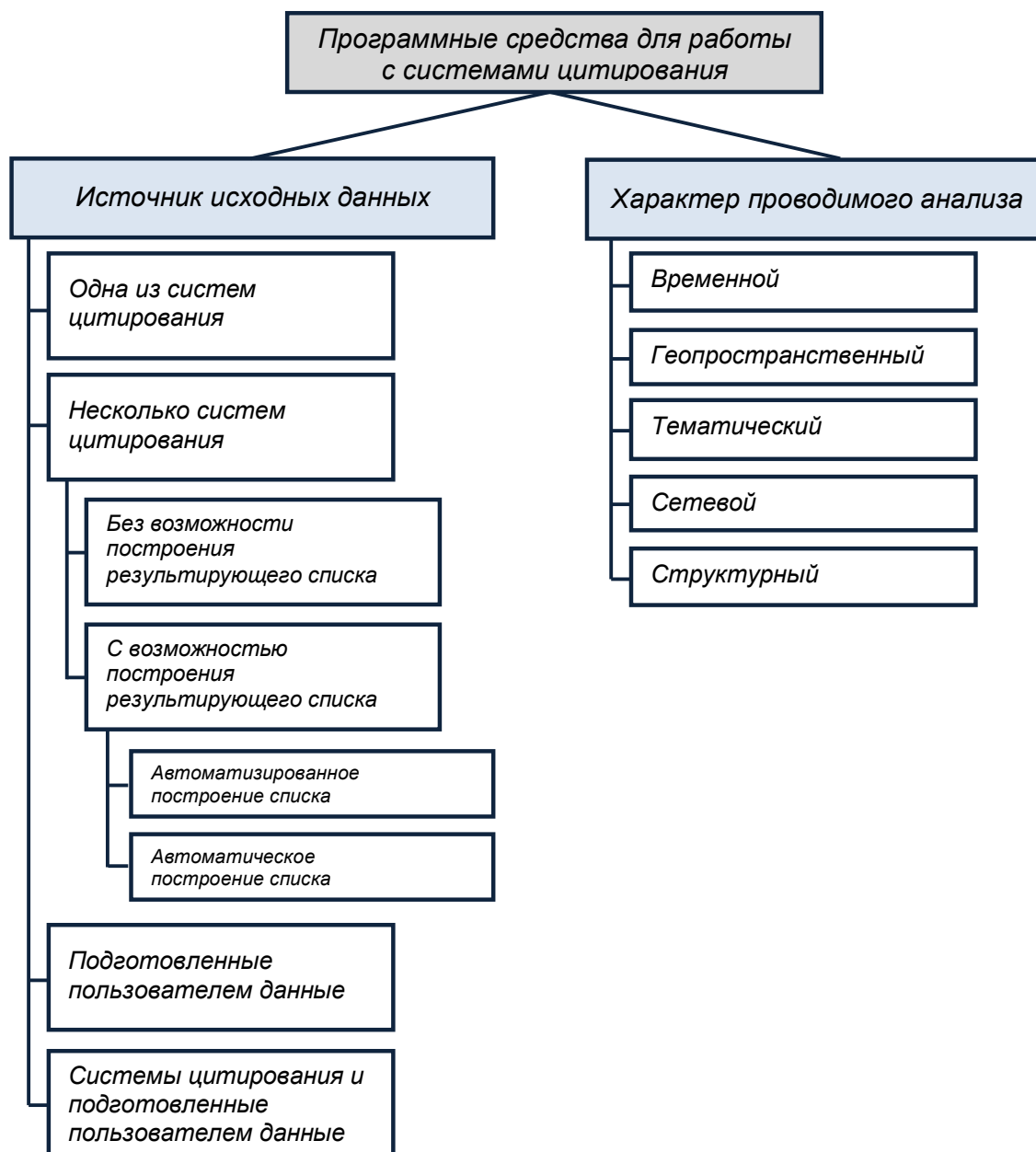


Рисунок 1 – Классификация существующих программных средств на основе выделенных факторов

Выделенные группы по признаку «Источник исходных данных», их особенности и недостатки, а также названия программных средств представлены в таблице 1.

Таблица 1 – Описание выделенных групп по признаку «Источник исходных данных»

| № | Источник исходных данных                                                                            | Программные средства                                 | Недостатки                                                                                                                                                                                                                                                                                                                                                           |
|---|-----------------------------------------------------------------------------------------------------|------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Одна из систем цитирования                                                                          | Publish or Perish [10], CleanPop [11], HistCite [12] | Не отражают полную картину публикационной активности                                                                                                                                                                                                                                                                                                                 |
| 2 | Несколько систем цитирования (без возможности построения результирующего списка)                    | CiteSpace [13]                                       | Отсутствует возможность построения результирующего списка, требуется ручная обработка всего списка публикаций и цитирований                                                                                                                                                                                                                                          |
| 3 | Несколько систем цитирования (с возможностью автоматизированного построения результирующего списка) | Leydesdorff Software                                 | Результирующий список строится путём исключения полностью совпадающих публикаций и цитирований, что приводит к существенным временным затратам на проверку полученных данных                                                                                                                                                                                         |
| 4 | Несколько систем цитирования (с возможностью автоматического построения результирующего списка)     | CiteSearch [14]                                      | Отсутствует возможность дополнения исходных данных, полученных из систем цитирования данными из других источников, подготовленными вручную. При анализе используются международные системы цитирования, не используется РИНЦ. Отсутствует возможность автоматического построения результирующего списка публикаций и цитирований в условиях неполной определённости. |

|   |                                                           |                         |                                                                                                                                                                                                                                 |
|---|-----------------------------------------------------------|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5 | Подготовленные пользователем данные                       | bibExcel [15]           | Используются данные, подготовленные пользователем из собственных источников или экспортированных из систем цитирования, что зачастую может стать очень трудоёмкой или даже невыполнимой задачей в условия ограниченного доступа |
| 6 | Системы цитирования и подготовленные пользователем данные | Science of Science Tool | При анализе используются международные системы цитирования, не используется РИНЦ. Отсутствует возможность автоматического построения результирующего списка публикаций и цитирований                                            |

Из проведённого анализа можно сделать следующие выводы:

- отсутствуют программные средства, способные агрегировать данные о публикационной активности одновременно из различных систем цитирования с учётом данных РИНЦ и из источников данных, подготовленных пользователем;
- отсутствуют программные средства, обладающие возможностью автоматического построения результирующего списка публикаций, цитирований и расчёта наукометрических показателей в условиях неполной определённости.

При анализе программных средств с различным характером проводимого анализа было выявлено, что ни одно программное средство не позволяет проводить анализ накопленных данных о публикационной активности в условиях неполной определённости.

Для устранения недостатков, выявленных при рассмотрении программных средств, предоставляющих дополнительные возможности по работе с системами цитирования, предлагается разработать автоматизированную программную систему сбора библиометрических и наукометрических показателей учёных из различных систем цитирования и баз данных, с целью получения достоверной, непротиворечивой и полной информации для последующего анализа.

Практической базой при построении программной системы служат метаданные статей, размещенных в Российском индексе научного цитирования и в международных базах данных Web of Science, Scopus, Agris, GeoRef,

Chemical Abstracts Service и др., авторами которых являются научные сотрудники учреждений высшего образования.

Проект будет основан на использовании библиометрических и наукометрических показателей, которые позволяют проводить объективную и независимую оценку результативности интеллектуальной деятельности ученых, а так же выявление скрытых взаимосвязей между результатами интеллектуальной деятельности ученых различных междисциплинарных профилей. Программная система будет основана на базе существующих критериев оценки уровня результативности научной деятельности ученых по ряду наукометрических показателей, таких как: индекс цитируемости, индекс Хирша, импакт-фактор в Российском индексе научного цитирования (РИНЦ) и в международных базах данных (Web of Science, Scopus, Agris, GeoRef, Chemical Abstracts Service и т.д.), а так же на разработке интеллектуальных методов обработки полученных данных по привлечению образовательных учреждений к участию в программах инновационного развития промышленных компаний.

С помощью методов интеллектуального анализа планируется провести изучение проблемы определения перспективности научных исследований, используя результаты интеллектуальной деятельности образовательных учреждений.

#### *Список литературы*

- 1. Ковалевский, В. П. Формирование информационно-образовательной среды вуза: опыт Оренбургского государственного университета / В.П. Ковалевский, В.В. Быковский Е.В. Дырдина, Т.В. Волкова // Информатизация образования и науки, 2015. - N 2 (26). - С. 15-23.*
- 2. Брайнин Н.А., Оценка эффективности исследовательской деятельности экономического университета на основе анализа публикационной активности: принципы, методы, инструменты / Н.А. Брайнин // Известия Российского экономического университета им. Г.В. Плеханова. – 2012. – № 5 (10). – С. 096-107.*
- 3. Овчинников М.Н., Об оценивании деятельности университетов и показателях эффективности программ развития / М.Н. Овчинников // Университетское управление: практика и анализ. – 2012. – № 1. – С. 25-30.*
- 4. Полозов А.А., Рейтинг вуза: эволюция и проблемы / А.А. Полозов // Университетское управление: практика и анализ. – 2011. – № 2. – С. 85-89.*
- 5. Спиридонова А.А., Мониторинг процессов в системе менеджмента качества вуза / А.А. Спиридонова, Е.Г. Хомутова // Университетское управление: практика и анализ. – 2011. – № 6. – С. 37-43.*
- 6. Болдырев, П.А. Технология управления электронными ресурсами библиотеки в информационно-образовательной среде вуза / П.А. Болдырев // Библиотековедение. – 2014. – № 4. – С. 36-40.*

7. Научная электронная библиотека eLIBRARY.RU [Электронный ресурс] : научная электронная библиотека. - Электрон. дан. - Москва : 2000. - Режим доступа: <http://elibrary.ru>. - Загл. с экрана. - 15.12.2015.
8. Scopus [Электронный ресурс] : библиографическая и реферативная база данных. - Электрон. дан. - Режим доступа: <http://www.scopus.com/>. - Загл. с экрана. - 15.12.2015.
9. Web of Science [Электронный ресурс] : поисковая платформа. - Электрон. дан. - Режим доступа: <http://isiknowledge.com>. - Загл. с экрана. - 15.12.2015.
10. Jacsó, P. Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the Publish or Perish software. SAVVY SEARCHING: Online information review, 2009, v. 33, n. 6, p. 1189-1200.
11. 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 № 6, P. 363-371.
12. HistCite [Электронный ресурс]. – Режим доступа: <http://www.histcite.com/>. - Загл. с экрана. - 15.12.2015.
13. CiteSpace [Электронный ресурс]. – Режим доступа: <http://cluster.cis.drexel.edu/~cchen/citespace/>. - Загл. с экрана. - 15.12.2015.
14. Yang, Kiduk. Citation Analysis: A Comparison of Google Scholar, Scopus, and Web of Science / Kiduk Yang, Lokman I. Meho // *Proceedings of the American Society for Information Science and Technology*. – 2006, V. 43, I. 1. – P. 1-15.
15. BibExcel [Электронный ресурс]. – Режим доступа: <http://homepage.univie.ac.at/juan.gorraiz/bibexcel/index.html> - Загл. с экрана. - 15.12.2015.