

**Сравнительный анализ программного обеспечения
для библиометрических исследований и построения карт науки**
Comparing software for bibliometric studies and science mapping

**Порівняльний аналіз програмного забезпечення
для бібліометричних досліджень та побудова карт науки**

Н. А. Мазов

*Институт нефтегазовой геологии и геофизики им. академика А. А. Трофимука СО РАН,
Новосибирск, Россия*

В. Н. Гуреев

*Государственный научный центр вирусологии и биотехнологии «Вектор»,
Кольцово, Новосибирская область, Россия*

Nikolay Mazov

*A. A. Trofimuk Institute of Oil and Gas Geology and Geophysics,
Siberian Division of the Russian Academy of Sciences,
Novosibirsk, Russia*

Vadim Gureev

*VECTOR State Research Center of Virology and Biotechnology,
Koltsovo, Novosibirsk Region, Russia*

М. О. Мазов

*Институт нефтегазовой геологии та геофизики ім. академіка А. О. Трофімука СВ РАН,
Новосибірськ, Росія*

В. М. Гурєєв

*Державний науковий центр вірусології та біотехнології «Вектор»,
Кольцово, Новосибірська область, Росія*

В настоящее время накоплены огромные объемы библиографической информации различного вида, требующей качественно новых форм аналитико-синтетической обработки. В информационной деятельности наблюдается возрастающий интерес, привлекаемый к библиометрическим исследованиям, в частности к картографированию науки или к библиографическому картографированию. Отчасти это связано с более открытым и публичным доступом к наукометрическим базам данных. Несмотря на то, что лидеры производства и исследований наукометрических баз данных – Thompson Reuters (БД Web of Knowledge) и Elsevier (БД Scopus) предоставляют необходимый сервис для анализа публикаций, тем не менее, он остается весьма ограниченным. В настоящей работе представлен сравнительный анализ программного обеспечения, предназначенного для различных аспектов библиометрического анализа и библиометрического картографирования.

Ключевые слова: Библиометрический анализ, Картографирование науки, Web of Science, Scopus, Программное обеспечение, сравнительный анализ.

Today, vast arrays of bibliographic information is accumulated which demands whole new forms of analyses and synthesis. Interest towards bibliometric studies, in particular towards science mapping or bibliographic mapping, is increasing. Partially it is due to the more open and public access to scientometric databases. Thompson Reuters (Web of Knowledge database) and Elsevier (Scopus database) both provide relevant service for publications analysis but it is still limited though. The authors compare software for various aspects of bibliometric analysis and bibliometric mapping.

Keywords: bibliometric analysis, science mapping, Web of Science, Scopus, software, comparative study.

Сьогодні накопичено величезні об'єми різної за видами бібліографічної інформації, яка потребує якісно нових форм аналітико-синтетичної обробки. В інформаційній діяльності спостерігається зростаючий інтерес до бібліометричних досліджень, зокрема до картографування науки або до бібліографічного картографування. Частково це пов'язано з більш відкритим і публічним доступом до наукометричних баз даних. І хоча лідери виробництва та досліджень наукометричних баз даних –

Tomson Reuters (БД Web of Knowledge) та Elsevier (БД Scopus) надають необхідний для аналізу публікацій сервіс, однак, він лишається досить обмеженим. У даній роботі представлено порівняльний аналіз програмного забезпечення, що призначене для різних аспектів бібліометричного аналізу та бібліометричного картографування.

Ключові слова: Бібліометричний аналіз, Картографування науки, Web of Science, Scopus, Програмне забезпечення, порівняльний аналіз.

В настоящее время существует несколько онлайн-библиографических (а также библиометрических) баз данных, содержащих метаописания научных работ, а также их цитирования. Эти источники библиографической информации позволяют находить и извлекать информацию о большинстве научных областей. Безусловно, наиболее важные библиографические базы данных – это Web of Science (WoS) (<http://www.webofknowledge.com>), Scopus (<http://www.scopus.com>), Google Scholar (<http://scholar.google.com>) и NLM's MEDLINE (<http://www.ncbi.nlm.nih.gov/pubmed>).

В настоящее время существует немало программ для решения проблем библиометрического картографирования. Подбор программного обеспечения для успешного решения этих является очень важным вопросом. Ранее авторами был представлен подробный обзор программного обеспечения, предназначенного для библиометрических исследований, часть из которого предназначена и для библиометрического картографирования [1].

Картографирование науки, или библиометрическое картографирование является важной исследовательской проблемой в области библиометрии. Она предпринимает попытки представить интеллектуальные связи в динамически изменяющейся системе научного знания. Другими словами, картографирование науки стремится отразить структурные и динамические аспекты научного исследования [2-4].

Общий процесс картографического анализа науки предполагает различные этапы:

- получение данных,
- предварительную обработку,
- извлечение сети,
- нормирование,
- построение карты,
- анализ и графическое представление.

В конце этого процесса исследователю необходимо интерпретировать результаты и сделать из них некоторые выводы. Ниже подробно рассмотрим каждый из этапов.

Для получения данных могут быть использованы различные библиометрические БД.

Данные, извлеченные из библиографических источников могут содержать ошибки, такие как, например, орфографические ошибки в написании имени автора, заглавия журнала или списке литературы. Иногда в изначальные данные необходимо добавить дополнительную информацию, например, если место работы автора неполное или ошибочное. По этой причине картографический анализ науки не может применяться напрямую к полученным из библиографических источников данным, то есть необходим процесс предварительной обработки полученных данных. В действительности, процесс предварительной обработки, возможно, один из наиболее важных этапов в усовершенствовании качества единиц анализа (главным образом, авторы и слова) и, следовательно, в получении лучших результатов картографического анализа науки. Чтобы провести качественный картографический анализ науки, для подготовки данных можно применять следующие различные процессы предварительной обработки:

- Определение дублирующих записей и записей с орфографическими ошибками.
- Квантование времени для разбиения данных на различные временные периоды, или кванты времени, для анализа эволюции исследуемой научной области.
- Выбор наиболее значимых данных.

Предварительная обработка сетей может использоваться для выбора наиболее значимых вершин сети и отношений между единицами анализа (удаление изолированных вершин, удаления менее значимых связей между вершинами и пр.).

Когда построена сеть взаимоотношений между выбранными единицами анализа, данные вначале преобразуются для выявления подобия данных или, говоря точнее, для нормирования данных. При нормировании текста для каждого термина определяется его вес в соответствии с его значимостью во всем тексте. Могут применяться различные системы измерений при нормировании текста: частота встречаемости термина и пр.

Этап создания карты является самым важным. Сам по себе процесс ответствен за создание карты посредством применения алгоритмов картопостроения ко всей сети, сформированной посредством взаимоотношений между выбранными единицами анализа. Для построения карты могут использоваться различные технические приемы [4]. Информация на выходе или вид построенной карты будут зависеть от применяемого метода [5].

После построения карты для извлечения полезных знаний могут применяться различные виды анализа: Сетевой анализ, временной анализ, геопространственный анализ.

Дальнейшее графическое представление очень важно для ясного понимания и лучшей интерпретации результатов. По завершении картографического анализа пользователю необходимо интерпретировать результаты и карты, используя свой опыт и знания.

Хотя картографический анализ науки можно проводить с использованием общего программного обеспечения, созданного для анализа социальных сетей [4], существуют другие специальные программы, разработанные специально для проведения картографического анализа науки [1].

Мы представим десять типичных программ, специально созданных для анализа научных отраслей средствами построения карт науки. Ниже приводится список этих программ.

- Bibexcel (Persson et al., 2009)
- CiteSpace II (Chen, 2004, 2006)
- CoPalRed (Bailón-Moreno et al., 2005, 2006)
- IN-SPIRE (Wise, 1999)
- Leydesdorff's Software
- Network Workbench Tool (Börner et al., 2010; Herr et al., 2007)
- Sci2 Tool (Sci² Team, 2009)
- VantagePoint (Porter & Cunningham, 2004)
- VOSViewer (van Eck & Waltman, 2010)
- SciMAT (M.J. Cobo et al., 2011)

Для проведения сравнительных исследований проанализировано девять программ, во внимание принимались следующие аспекты:

- а) методы предварительной обработки,
- б) доступные библиометрические сети,
- в) используемые методы нормирования,
- г) виды анализа, прочие второстепенные аспекты.

В табл. 1 показаны сравнительные сводные характеристики программ по построению карт науки. Можно заметить, что SciMAT является одним из наиболее полных инструментариев, если принять в расчет все перечисленные выше характеристики [6].

Таблица 1

Сводные характеристики программных продуктов по построению карт науки

Программный инструмент	Предварительная обработка	Сети	Нормирование	Анализ
BibExcel	Уплотнение данных и сетей	DBCA, ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA и др.	Косинус Солтона, индекс Жаккара или показатели Владуца и Кука	сетевой
CiteSpace	Квантование времени, уплотнение данных и сетей	DBCA, ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA и др.	Косинус Солтона, сила Дайса или Жаккара	определение прорывных областей, геопространственный, сетевой, временной
CoPalRed	Устранение дублетов, квантование времени, уплотнение данных	CWA	индекс эквивалентности	сетевой, временной
IN-SPIRE	Уплотнение данных	CWA	Обусловленная вероятность	определение прорывных областей, сетевой, временной
Loet Leydesdorff's Software		ABCA, JBCA, ACAA, CCAA, ICAA, ACA, CWA	Косинус Солтона	
Network Workbench Tool	Устранение дублетов, квантование времени, уплотнение данных и сетей	DBCA, ACAA, DCA, CWA, DL	Задается пользователем	определение прорывных областей, сетевой, временной
Science of Science Tool	Устранение дублетов, квантование времени, уплотнение данных и сетей	ABCA, DBCA, JBCA, ACAA, ACA, DCA, JCA, CWA DL и др.	Задается пользователем	определение прорывных областей, геопространственный, сетевой, временной
VntagePoint	Устранение дублетов, квантование времени, уплотнение данных	ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA и др.	r Пирсона, косинус Солтона или наибольшая пропорциональность	определение прорывных областей, геопространственный, сетевой, временной
VOSviewer			Сила зависимости	сетевой
SciMAT	Устранение дублетов, квантование времени, уплотнение данных	DBCA, ABCA, JBCA, ACAA, ACA, DCA, JCA, CWA и др.	Сила зависимости, индекс эквивалентности, индекс включения, индекс Жаккара и косинус Солтона	сетевой, временной, эффективности

ABCA = авторская библиографическая связка; DBCA = документная библиографическая связка; JBCA = журнальная библиографическая связка; ACAA = соавторы по автору; CCAA = соавторы по стране; ICAA = соавторы по институту; ACA = социрование по автору; DCA = социрование по документу; JCA = социрование по журналу; CWA = совместно появляющееся слово; DL = прямая связь.

В заключении следует отметить, что представленный список программного обеспечения является исчерпывающим по картографическому анализу науки, которые используются по всему миру. Эти программы имеют различные характеристики; например, некоторые из них направлены лишь на графическое представление, тогда как другие обладают различными модулями предварительной обработки. Не было ни одной программы, которую можно было бы признать лучшей. Следовательно, мы считаем, что исчерпывающий картографический анализ определенной области науки должен проводиться с использованием нескольких из этих программ, что позволит собрать все важные знания с различных углов зрения. Такая кооперация программ дает положительный эффект, позволяющий получать знание, скрытое в данных. Рассмотренный набор программ может быть рекомендован специалистам в областях, связанных с библиометрическими исследованиями для построения карт науки.

Список литературы:

1. Мазов Н.А. Свободно распространяемые программы для наукометрических и библиометрических исследований // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: 19-я междунар. конф. «Крым 2012» (2–10 июня 2012 г., г. Судак): Труды конф. – М.: Изд-во ГПНТБ России, 2012. – С. 1–6.
2. Morris, S., Van Der Veer Martens, B. Mapping research specialties // *Annual Review of Information Science and Technology*, – 2008. – 42(1), P. 213–295.
3. van Eck, N.J., Waltman, L. Bibliometric mapping of the computational intelligence field // *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. – 2007. – 15(5), P. 625–645.
4. Börner, K., Chen, C., Boyack, K. Visualizing knowledge domains // *Annual Review of Information Science and Technology*. – 2003. – 37, P. 179–255.
5. Cobo M.J., López-Herrera A.G., Herrera-Viedma E., Herrera F. Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools // *J. of Amer. Soc. for Inf. Sci. and Technol.* – 2011. – 62(7), P. 1382–1402
6. Cobo M.J., López-Herrera A.G., Herrera-Viedma E., Herrera F. SciMAT: A New Science Mapping Analysis Software Tool // *J. of Amer. Soc. for Inf. Sci. and Technol.* – 2012. – 63(8), P. 1609–1630.