

**Свободно распространяемые программы
для наукометрических и библиометрических исследований
Freeware for Scientometric and Bibliometric Studies
Програми для наукометричних і бібліометричних досліджень,
що розповсюджуються вільно**

Н. А. Мазов

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

Nikolay Mazov

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

Н. А. Мазов

*Институт нефтегазовой геології та геофізики
ім. академіка А. А. Трофимука СВ РАН,
Новосибірськ, Росія*

В последние годы в информационной практике наблюдается возрастающий интерес, привлекаемый к наукометрическим и библиометрическим исследованиям. Отчасти это связано с тем, что накоплены колоссальные объемы библиографической информации различного вида, требующей качественно новых форм аналитико-синтетической обработки, а с другой стороны это связано с более открытым и публичным доступом к наукометрическим базам данных. Несмотря на то, что лидеры производства и исследований наукометрических баз данных – Tompson Reuters (БД Web of Knowledge) и Elsevier (БД Scopus) предоставляют необходимый сервис для анализа публикаций, тем не менее, он остается весьма скудным и ограниченным. В настоящей работе представлен обзор свободно распространяемого программного обеспечения, предназначенного для различных аспектов наукометрических и библиометрических исследований.

In the recent years, interest towards scientometric and bibliometric studies has been increasing. Partly this is owing to the fact that vast amounts of various type bibliographic information have been accumulated and, on the other hand, it is owing to the more open and public access to scientometric databases. Despite leading scientometric databases producers and researchers, e. g. Tompson Reuters (ISI Web of Knowledge) and Elsevier (DB Scopus) provide necessary services for the analysis of publications, nevertheless, these services still remain very poor and limited. The author reviews freeware intended for various aspects of scientometric and bibliometric studies.

В останні роки в інформаційній практиці спостерігається зростаючий інтерес до наукометричних і бібліометричних досліджень. Частково це пов'язано з тим, що накопичено колосальні обсяги бібліографічної інформації різного виду, які вимагають якісно нових форм аналітико-синтетичної обробки, а з іншого боку, це пов'язано з більш відкритим і публічним доступом до наукометричних баз даних. Незважаючи на те, що лідери виробництва і досліджень наукометричних баз даних - Tompson Reuters (БД Web of Knowledge) та Elsevier (БД Scopus), надають необхідний сервіс для аналізу публікацій, однак, він залишається дуже мізерним і обмеженим. В даній роботі представлено огляд програмного забезпечення, що вільно поширюється, і призначеного для різних аспектів наукометричних і бібліометричних досліджень.

Подбор программного обеспечения для успешного решения инфометрических исследовательских задач в области наукометрии и библиометрии является очень важным вопросом. В настоящее время существуют мощные коммерческие наукометрические программы с закрытыми исходными кодами, например,

1. IN-SPIRE™ Visual Document Analysis [1], программное обеспечение для исследования и визуализации текстовых данных и аналитический инструментарий для получения различных временных трендов. Используется для исследования научно-технической литературы.

2. VantagePoint [2], мощный инструмент для глубокого анализа текста с целью обнаружения знаний в результатах поиска в библиографических и патентных базах данных, позволяет быстро проанализировать большие объемы информации, отобранные в результате поиска и превратить отобранную информацию в знание.

Однако, существует большое количество свободно распространяемых программ, как с открытыми, так и с закрытыми кодами, речь о которых пойдет в настоящей работе. В статье приводятся небольшие обзоры свободно распространяемых программ, предназначенных для наукометрических и библиометрических исследований:

1. HistCite,
2. Bibexcel,
3. CiteSpace,
4. Sci2 Tool,
5. Loet Leydesdorff,
6. Publish or Perish,
7. VOSviewer,
8. InterDisciplinary Research (IDR),
9. Network Workbench.

Ниже дано краткое описание перечисленных программных продуктов.

1. HistCite [3] – очень гибкий программный продукт, создание и развитие которого инициировал основатель Института научной информации США и изобретатель Science Citation Index® доктор Юджин Гарфилд. Настоящий программный продукт предназначен для анализа и визуализации результатов поиска в БД Web of Science (файл в формате plain text) по различным критериям: по ключевым словам, авторам и цитируемым авторам, журналам, странам и организациям, от которых авторы публикуют свои статьи и др. Программа легка и прозрачна в установке и использовании.

2. BibExcel [4-5] – программа предназначена для анализа библиографических данных, или любых данных текстовой природы, отформатированной соответствующим образом. Идея состоит в том, что необходимо подготовить файлы данных, которые могут быть импортированы в Excel, или любую программу, работающую с табличными данными, для дальнейшей обработки. Эта программа включает множество инструментальных средств, часть которых видима в окне, а часть других скрыта в меню. В качестве исходного файла для библиометрического исследования может быть использован файл результатов поиска в БД Web of Science, сохраненный в формате comma delimited. Программа разработана шведским ученым Persson, O.D.

3. CiteSpace [6] – свободно распространяемое программное обеспечение для визуализации и анализа тенденций и направлений в научной литературе. Приложение разработано в виде инструмента для визуализации прогрессивной области знаний [7]. Приложение направлено на поиск критических точек в развитии области или сферы, особенно интеллектуальных критических точек и важнейших моментов. Детальное предметное исследование можно найти в [8] и в других публикациях. CiteSpace имеет набор функций, облегчающих понимание и способствующих интерпретации сетевой или исторической модели, включая определение быстро растущих предметных областей, обнаружение «горячих точек» цитирования в области публикаций, разделения сети на кластеры, автоматическое маркирование кластеров в терминах, взятых из цитирующих статей, геопространственные модели совместных проектов, уникальные области международных совместных проектов. CiteSpace поддерживает структурный и временной анализ различных сетей, возникающих из научных публикаций, включая сети совместных проектов, сети авторского социотирования и сетей социотирования документов. Первичным источником входных данных для CiteSpace являются результаты поиска в БД Web of Science, сохраненные в файле. CiteSpace предоставляет также несколько простых интерфейсов для получения данных из других систем, например из PubMed. Приложение можно использовать для генерации слоя географической карты, основанного на местоположениях авторов, который в дальнейшем можно посмотреть в Google Earth.

4. Science of Science (Sci2) Tool [9] является блочным набором инструментов, специально разработанным для изучения науки. Он поддерживает временной, геопространственный, тематиче-

ский и сетевой анализ и визуализацию набора данных на различных уровнях (микро - индивидуальном, мезо - локальном и макро - глобальном). Пользователи могут:

- воспользоваться он-лайнowymi наборами данных по науке либо загрузить свои собственные данные;
- проводить различные типы анализа с наиболее эффективными доступными алгоритмами;
- использовать различные воплощения для интерактивного объяснения и понимания специфических наборов данных;
- распределять наборы данных и алгоритмы сквозь научные границы.

Sci2 Tool построен на оболочке Cyberinfrastructure Shell (CIShell) (Cyberinfrastructure for Network Science Center, 2008) – открытом программном обеспечении для простого объединения и использования наборов данных, алгоритмов, инструментов и вычислительных ресурсов.

5. Loet Leydesdorff [10] представляет собой набор ДОО программ открытого доступа для разбора, преобразования и анализа библиометрических данных, полученных из таких источников, как Scopus, Web of Science и Google Scholar. Можно проводить анализ на соавторство, сети совместных проектов между странами, организациями и городами, анализ на одинаковые ключевые слова, цитирование, библиографический анализ и пр. Хотя непосредственно в приложение и не включены инструменты визуализации, оно подготавливает данные для создания реляционной базы данных и визуализации в других программах.

6. Publish or Perish [11] – программное обеспечение для поиска и анализа научных цитирований. Оно использует Google Scholar для получения необработанных цитат, а затем анализирует их и выводит следующие статистические данные:

- общее число статей;
- общее число цитирований;
- среднее число цитирований на статью;
- среднее число цитирований на автора;
- среднее число статей на автора;
- среднее число цитирований в год;
- индекс Хирша с относящимися к нему параметрами;
- индекс Эгга;
- современный h-индекс;
- взвешенный по возрасту показатель цитирования;
- два варианта индивидуальных h-индексов.

Результаты доступны с экрана, их также можно скопировать в буфер обмена Windows (для использования в других приложениях) или сохранить в одном из нескольких выходных форматов (для последующего упоминания и дальнейшего анализа). В Publish or Perish включено детальное руководство пользователя с подсказками и дополнительной информацией о системе показателей цитирования.

7. VOSviewer [12] – программа открытого доступа, которую можно использовать для различных целей. VOSviewer можно использовать для создания карт, основанных на сети данных. Карты создаются с использованием технологии создания карт VOS и технологии выделения кластеров VOS. VOSviewer можно использовать для обозрения и исследования карт. Программа выводит карту различными способами, каждый из которых выделяет ее различные аспекты. Предложены такие функции, как увеличение, прокрутка и поиск, которые облегчают тщательное исследование карты.

Изначально VOSviewer задумывался для анализа библиометрических сетей. Программа, например, может использоваться для создания карт по публикациям, статьям или журналам, основанных на сети цитирования, или создавать карты ключевых слов, основанных на их одновременном появлении в сети.

8. InterDisciplinary Research («IDR») [13] – данный веб-сайт предлагает новый инструмент – слои на карте науки в качестве метода исследования степени междисциплинарности набора публикаций. Технология наложения слоев показывает распространение публикаций на глобальной карте науки, т.е. структуре науки, какой она получается на основе анализа перекрестного цитирования

между дисциплинами. Междисциплинарное исследование часто считается существенным для научного и технологического развития. Однако, междисциплинарность – это неоднозначное и многоаспектное понятие: существует несколько допустимых способов ее выделения, и нет единого мнения по поводу того, какое из определений наиболее приемлемо.

Анализ можно провести на различных единицах объединения: например, для университета или корпорации, для темы исследования или для исследовательской программы или финансирующей организации. Располагая публикации на карте науки, можно понять разнообразие вовлеченных дисциплин. Научное приписывание публикаций к дисциплинам проблематично и противоречиво, карты со слоями надежны только тогда, когда на них много чисел.

Карты позволяют интуитивно понять различные аспекты дисциплинарного разнообразия. Во-первых, количество включенных дисциплин. Во-вторых, соотношение дисциплин, т.е. распределены ли публикации равномерно или с преобладанием какой-либо дисциплины. В-третьих, и это важно, включено когнитивное расстояние между вовлеченными во взаимодействие дисциплинами – покрывает ли рассматриваемое исследование далекие или родственные области науки. Этот аспект несоответствия – ключевое преимущество карт: они проводят границу между междисциплинарностью малого радиуса (например, химия и физика) и большого радиуса (например, общественные науки и биология). Измерения междисциплинарности можно связать с этими картами [14-15].

9. Network Workbench (NWB) [16] – поддерживает сеть научных исследований, преодолевая научные границы. Пользователи NWB имеют он-лайн доступ к большинству сетевых наборов данных или могут загружать собственные сети. Они имеют возможность представить сетевой анализ посредством наиболее эффективных доступных алгоритмов. Кроме того, у них имеется возможность создавать, запускать и проверять достоверность сетевых моделей для улучшения их представления о структуре и динамике специфических сетей. NWB предоставляет продвинутый инструментальный для визуализации, позволяющий интерактивно исследовать и понимать специфические сети, а также их взаимодействие с другими типами сетей.

В заключении следует отметить, что представленный список программного обеспечения не является исчерпывающим. На основе вышеописанных свободно распространяемых программ возможно проведение качественного инфометрического исследования. Прежде всего, это открывает путь для самостоятельной работы специалистов в рамках активно развивающихся наукометрических и библиометрических исследований и позволяет им вносить вклад в дальнейшее развитие программного обеспечения. Если говорить о закрытых и платных программах, то в таких системах пользователь-исследователь весьма ограничен в творчестве рамками соответствующего лицензионного соглашения.

Рассмотренный набор программ может быть рекомендован специалистам в областях, связанных с наукометрическими и библиометрическими исследованиями.

Список литературы:

1. <http://in-spire.pnnl.gov/>
2. <http://www.thevantagepoint.com/>
3. <http://www.histcite.com/>
4. <http://www8.umu.se/inforsk/Bibexcel/>
5. Persson, O.D., et al. 2009. How to use Bibexcel for various types of bibliometric analysis. In *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday*, p 9-24. Leuven, Belgium: International Society for Scientometrics and Informetrics.
6. <http://cluster.cis.drexel.edu/~cchen/citespace/>
7. Chen C. Searching for intellectual turning points: Progressive knowledge domain visualization // *PNAS*. 2004. - vol. 101. - suppl. 1. – P. 5303–5310.
8. Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature // *J. of the Amer. Soc. for Inf. Sci. and Technol.* - 2006. - 57(3). - P. 359-377.
9. <https://sci2.cns.iu.edu/>
10. <http://www.leydesdorff.net/>
11. <http://www.harzing.com/pop.htm>
12. <http://www.vosviewer.com/>
13. <http://idr.gatech.edu/>

14. Porter A.L. et al. Measuring researcher interdisciplinarity // *Scientometrics*. 2007. – Vol. 72, No. 1. – P. 117–147.
15. Rafols I., Meyer M. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience // *Scientometrics*. 2010. – vol. 82. – P. 263–287.
16. <http://nwb.cns.iu.edu/>