

## Методы библиометрического анализа научных публикаций

### Bibliometric approaches to scientific publication analysis

*Елена Павловска*

*Университет библиотековедения и информационных технологий,  
София, Болгария*

*Elena Pavlovska*

*State University of Library Studies and Information Technologies,  
Sofia, Bulgaria*

Рассмотрены основные методы библиометрического анализа, которые сгруппированы следующим образом: методы количественной оценки элементов документального информационного потока (ДИП); методы цитатного анализа; методы лексического анализа. Дана характеристика каждой группы. Отмечено, что на современном этапе развития информационных технологий используется все многообразие современных методов библиометрического анализа для структурирования науки или отдельных отраслей знания, слежения за их развитием, принятия решений о финансировании отдельных научных коллективов и проектов, при формировании научной политики.

Main approaches to bibliometric analysis are regrouped as follows: quantitative evaluation of document information flow; citation analysis methods; lexical analysis methods. Each group is characterized. The author emphasizes that at the current stage of information technology development, the whole variety of bibliometric methods are used to structure sciences and individual disciplines, tracking their advancements, making decisions on financing research teams and projects, formulating science policy.

В настоящее время в исследованиях, проводимых в информатике, науковедении и социологии, большое внимание уделяется вопросам прогнозирования и оценки развития научных дисциплин, выявления новых областей науки, выделения приоритетных научных направлений. Повышенный интерес к этим проблемам вызван необходимостью в современном, рациональном планировании и организации научных исследований, такой организации науки, при которой органы управления и планирования периодически получают информацию о структуре области знания или науки в целом, о состоянии исследований в конкретных областях науки в своей стране и мире.

Обычно подобную информацию органы управления наукой получают от экспертов – ведущих специалистов в конкретных научных направлениях. Но метод экспертизы, как известно, субъективен. Эксперты, являясь представителями «зрелых», сформировавшихся направлений, довольно часто не замечают нового направления в исследованиях или даже отвергают его. Это связано с тем, что «новое» опровергает систему представлений, сложившуюся за многие годы работы у ученого-эксперта. Для восприятия радикально новой идеи необходимо изменение научного мышления, а это достаточно трудно, поскольку ему свойственна консервативность. Именно поэтому объясним интерес исследователей к использованию специальных библиометрических методов, обеспечивающих большую достоверность и объективность результатов.

Известно, что развитие науки проявляется в росте ее количественных информационных параметров, поэтому количественные характеристики информационных потоков используются как косвенные данные о тенденциях и темпах развития конкретных отраслей наук, при формировании новых научных направлений, при выявлении наиболее перспективных из них. При этом исходят из того, что новые проблемные направления, формирующиеся внутри сложившейся отрасли науки, отличаются от нее скоростью роста числа публикаций, степенью концентрации новой научной информации, частотой появления новых научных терминов, фамилий авторов и т. д. [1].

К настоящему моменту в науковедении уже накоплен значительный арсенал **теоретических и экспериментальных методов библиометрического анализа**, используемых для выявления актуальности, перспективности различных научных направлений. Эти методы могут быть сгруппированы следующим образом:

– Методы количественной оценки элементов документального информационного потока (ДИП);

– Методы цитатного анализа;

– Методы лексического анализа.

Группировка эта условна, так как и цитаты, и термины также являются элементами документального потока, если рассматривать его на другом уровне подробности. Тем не менее, рассмотрим каждую из групп методов.

### **1. Методы количественной оценки элементов ДИП**

При анализе состояния и тенденций развития научных направлений наиболее часто используются методы количественной оценки элементов ДИП, которые представляют собой «продукты» научной деятельности ученых (книги, статьи, диссертации, обзоры, научные журналы, патенты и т.п.). Значимость этих видов документов различна, поэтому некоторые ученые предлагают при их подсчете учитывать «вес» каждого из них. Так, иногда, одну книгу приравнивают к четырем, шести или даже восемнадцати статьям.

Методы этой группы часто используют для определения структуры потока документов в различных научных дисциплинах [2] или конкретных проблемах [3, 4]. При этом часто используются реферативные издания, в которых подсчитывается количество публикаций из источников данного направления, отраженных в соответствующем разделе реферативного журнала [5]. Оценивая научное направление, многие авторы предлагают использовать известный закон Брэдфорда – Викери, выполнимость которого считается косвенным признаком «классичности» области исследования и используется для оценки развития и самостоятельности научного направления. О степени «классичности» научного направления можно судить и по возможности применения для описания этого направления других ранговых распределений, в частности известного закона Лотки. Этот закон описывает распределение авторов в зависимости от количества опубликованных ими статей.

Для ДИП характерна закономерность роста числа научных публикаций, журналов, научных открытий, числа ученых, отражающая состояние той или иной научной области и позволяющая определить границы отдельных этапов развития науки, прогнозировать тенденции ее дальнейшего развития, выявлять как рост науки в целом, так и отдельных ее направлений.

Одна из наиболее распространенных и общепризнанных моделей роста числа научных публикаций – модель экспоненциального роста, впервые предложенная Д. Прайсом. Эта модель согласуется с фактическими данными о росте числа публикаций за относительно небольшой промежуток времени (10–15 лет). Причем, чем шире рассматриваемая научная область, тем длительнее период, в течение которого выполняется экспоненциальный закон [6]. Критерием, позволяющим оценить развитие отдельных научных дисциплин, может служить, средняя скорость роста количества публикаций за определенный интервал времени (показатель, предложенный В. Горьковой [7]).

Одной из характеристик актуальности направления принято считать время удвоения продуктов научной деятельности. П. Оже, Д. Прайс, М. Конрад в своих работах определяли период удвоения числа научных открытий, числа журналов, числа ученых для конкретных областей науки.

Актуальность отдельных научных направлений можно оценивать также по:

– распределению ученых по научным отраслям и по странам [8, 9];

– распределениям публикаций по научным направлениям [7, 10];

– по странам и языкам, на которых они опубликованы [11, 4];

– распределениям спроса на литературу по научным направлениям с целью оценки их сравнительной актуальности [12].

Следует отметить, что оценивая состояние какого-либо научного направления, недостаточно выявлять только статистические закономерности, присущие ДИП. Необходимо учитывать временную эволюцию распределения публикаций и других продуктов научной деятельности. Такой пространственно-временной анализ распределения публикаций действительно представляет интерес, так как динамика информационных потоков отражает закономерности развития науки.

## 2. Методы цитатного анализа

Цитатный анализ или анализ библиографических ссылок в научных публикациях, позволяет определять связи между публикациями, выявлять структуру областей знания и даже прогнозировать их развитие. Новые научные направления возникают, как правило, внутри уже существующих, поэтому их формирование отражается в библиографическом аппарате публикаций. Анализ пристатейных библиографий дает возможность выявить уже сформировавшееся, но формально еще не выделившееся направление, а может быть и новые проблемы, часть из которых затем перерастет в самостоятельные разделы науки.

Практическая база для использования этого метода появилась с созданием в 1964 г. Ю. Гарфилдом указателей научных ссылок: Science Citation Index (SCI) – Индекс цитирования в области естественных и точных наук, Social Sciences Citation Index (SSCI) – Индекс цитирования по социальным наукам, Arts & Humanities Citation Index (A&HCI) – Индекс цитирования по искусству и гуманитарным наукам и справочника Journal Citation Reports (JCR) – указателя цитируемости журналов, а также междисциплинарного указателя научных обзоров Index to Scientific Reviews (ISR).

С начала 80-х годов XX века цитатный анализ начинают активно использовать для выявления тенденций в развитии науки. В этом направлении развивается деятельность Института научной информации (ИНИ) США по созданию карт и атласов науки. С 1990 г. ИНИ является собственностью корпорации Thomson Reuters, которая предлагает указатели цитируемости SCI, SSCI, A&HCI и JCR в виде баз данных.

Сравнительный анализ темпов роста числа библиографических ссылок в разных научных направлениях помогает выбрать наиболее перспективные из них [2].

Возрастное распределение ссылок (старение публикаций) представляет большой интерес в связи с выявлением «затухающих» научных направлений. Известно, что хотя вероятность цитирования данной статьи растет с ее возрастом, новые статьи цитируются значительно чаще. Это не значит, что старые работы не представляют ценности для современных исследований. В любой науке есть передовая литература и «архивная», содержащая проверенные временем знания. Соотношение «архивной» и новой литературы в ссылках различно в разных дисциплинах [1]. Д. Прайс предложил характеризовать дисциплины «индексом Прайса» – долей ссылок на работы последних пяти лет, т.е. отношением количества «оперативных» ссылок к количеству «архивных». Старение публикаций описывают также «периодом полужизни», т. е. числом лет, за которое количество ссылок на публикации падает вдвое. Период полужизни зависит от предметной области. Например, установлено, что этот показатель имеет наименьшее значение в биомедицине (три года), а наибольшее – в географии (16 лет). Это значит, что значительная часть публикаций по биомедицинским наукам стареет быстрее, чем публикации в других областях.

Исследование цитируемости ученых [6, 13, 14] проводится не только для определения их научного вклада, но и с целью определения состояния научной дисциплины, в которой они работают, и составления карт определенной научной области на данный момент. Для этой цели наиболее часто используют такие методы цитатного анализа, как метод библиографического группирования (М.Кесслер) и метод социитирования (И. Маршакова, Г. Смолл).

В последнее время в связи с внедрением рейтинговых систем оценки деятельности отдельных ученых, научных коллективов и университетов методы цитатного анализа все шире применяются как один из инструментов анализа и управления исследованиями [15]. С их помощью, стало возможно создавать карты науки, выявлять наиболее быстро развивающиеся области науки и новые направления исследований в мире, оценивать темпы развития различных направлений. Но при этом необходимо учитывать основные принципиальные **ограничения** методов цитатного анализа, которые заключаются в следующем:

– цитатный анализ в определенной степени субъективен, поскольку данный метод исследует ссылки в каждой статье, указывающие место данной работы во всем массиве публикаций, оценивая тем самым и саму работу автора. С этой точки зрения цитатный анализ является неявным методом экспертной оценки, в котором роль эксперта выполняет сам автор работы;

– большая часть научных документов читается, но не цитируется: из 100 произвольно выбранных материалов – 40 вообще не цитируются, 50 цитируются раз в год и только 10 цитируются чаще одного раза;

– в среднем публикации начинают цитироваться через 1–1,5 года после их выхода в свет, а «пик» цитирования нередко наступает еще позже [15]. Поэтому для оперативных оценок цитатный анализ не подходит;

– большой «вес» журнальных статей по сравнению с книгами;

– различия в написании иностранных фамилий;

– существуют определенные затруднения при применении цитатных методов: необходимо иметь Указатель библиографических ссылок или доступ к базам данных, предоставляющим наукометрические данные: Web of Science или SCOPUS;

– неполный охват мирового потока научных публикаций в базах данных научного цитирования, причем отдельные страны представлены в них непропорционально их вкладом в мировую науку. В первую очередь это относится к неанглоязычным странам. Таким образом, подавляющая часть публикаций ученых этих стран остается «невидимой» и недоступной для мирового сообщества ученых. Решить эту проблему можно только создавая национальные индексы цитирования. Европейская Комиссия уже приняла решение о создании альтернативной базы данных по европейским научным публикациям, где были бы полнее представлены издания на немецком, французском, итальянском, испанском и других европейских языках.

### 3. Методы лексического анализа научного текста

Эти методы берут начало от социологического метода **контент-анализа**, развитого Г. Лассуэлом и Ч. Осгудом [16, 17]. Процедура контент-анализа заключается в сведении рассматриваемого текста к ограниченному набору определенных элементов (символов), которые потом подсчитываются и анализируются. Особое внимание уделяется подсчету частоты употребления символов.

В последнее время этот метод используют не только в целях изучения самих текстов, а и для изучения создающих эти тексты систем. В применении к анализу науки, например, это означает исследование научной деятельности по научным текстам.

По мере своего обособления и углубления каждая область знаний вырабатывает свой особый язык, делающий ее малодоступной для специалистов в других областях. Поэтому статистический анализ лексики публикаций позволяет проследить глубину проникновения новых идей и методов, которая может быть задана частотой появления слов, кодирующих целые концепции.

Развитием методов лексического анализа, при которых в качестве индикатора используется «символ» или «термин» (дескриптор) является так называемый **«сленговый метод»**, предложенный С. Хайтуном [1]. Здесь в качестве измерителя автор предлагает использовать словарное слово, объединяющее соответствующую группу словоформ. Метод опирается на статистическую лингвистику, изучающую частотные распределения слов и создавшую частотные словари.

Необходимо отметить, что в последние годы ученые, решая самые разнообразные задачи в области информатики и науковедения, все чаще прибегают к разработке и использованию **формализованных методов**, обеспечивающих воспроизводимость результатов, точность и достаточную объективность. Этому способствует создание и бурное развитие в последние десятилетия большого числа электронных ресурсов в различных областях знания и внедрение новых информационных технологий. Электронные ресурсы, в т.ч. документальные базы данных, являясь реальным отображением процессов, происходящих в науке, содержат элементы, анализ которых позволяет оценивать и прогнозировать ее развитие. В качестве примера можно привести исследования, связанные с частотными анализами документальных баз данных: по автоматическому индексированию документов по текстам их рефератов и автоматическому составлению тезаурусов, по выявлению основных тенденций в изучении конкретных тематических проблем.

Следует отметить метод логико-смыслового моделирования (ЛСМ), позволяющий выделять наиболее значимые проблемы для развития данной предметной области, предложенный М. Субботиным [18]. Метод ЛСМ основан на использовании в качестве исходных элементов любых высказываний, которые могут быть выражены отдельным словом, словосочетанием или целым предложением.

Интересная методика прогнозно-аналитического исследования библиографических информационных ресурсов, предложена Е. Стась в [19]. В качестве входных параметров автор рассматривает интенсивность роста числа элементов в системе (изданий, документов, терминов), в качестве выходных параметров – интенсивность использования элементов системы, в качестве важнейшей характеристики системы потока документов – ее структуру. Закономерности развития науки проявляются в изменении структуры потока и его входных и выходных параметров.

Формализованный метод анализа документальных баз данных с целью выявления перспективных научных направлений, основанный на изменениях *динамических* характеристик лексики, используемой для создания научных публикаций, предложен Е. Павловской [20, 21]. Метод анализа лексического состава базируется на основных положениях теории неравновесной термодинамики открытых систем, разработанной И. Пригожиным. Метод апробирован на потоке документов в новом развивающемся научном направлении.

Информационные ресурсы, содержащие электронные научные публикации в последние годы стали основой для проведения анализа и обобщения содержащихся в них сведений, для нахождения соответствующих закономерностей и выработки рекомендаций к действиям для различных категорий потребителей. Органы управления наукой большинства европейских стран и США активно используют все многообразие современных методов библиометрического анализа научных документов для структурирования науки или отдельных отраслей знания, в слежении за их развитием, при принятии решений о финансировании отдельных научных коллективов и проектов, при формировании научной политики.

#### Список литературы:

1. Хайтун, С. Д. Наукометрия: Состояние и перспективы. – М. : Наука, 1983. – 344 с.
2. Грановский, Ю. В. Наукометрический анализ информационных потоков в химии. – М. : Наука, 1980. – 141 с.
3. Арский, Ю. М., Маркусова, В. А., Чумакова, Н. Ф. Науки о Земле: объемы и рейтинги отечественных публикаций. // Вестник РАН, 2007, т. 77, № 10, с. 887–895.
4. Высочанская, О. А. Анализ раздела «Информационный поиск» Реферативного журнала «Информатика» за 15 лет. // Науч. и техн. информация, Сер. 2, 1985, № 1, с. 22–29.
5. Зайченко, Н. Я. Выявление приоритетных направлений развития науки путем анализа научных публикаций, прореферированных в УРЖ «Джерело» на протяжении 1999–2002 гг. // Библиотеки национальных академий наук : Проблемы функционирования, тенденции развития. – Вып. 2. – Киев, 2003, с. 49–59.
6. Налимов, В. В., Мульченко, З. М. Наукометрия. – М. : Наука, 1969. – 422 с.
7. Горькова, В. И. Информетрия. // Итоги науки и техники. Сер. Информатика / ВИНТИ, 1988, т. 10. – 327 с.
8. Vlachy, J. Planning in Physics. // Czech. J. Phys., vol. B 18, 1968, N 11, p. 1479–1501.
9. Слащева, Н. А. Публикации ученых как объект наукометрических исследований. // Роль книгоиздания в развитии международных научных и культурных контактов : Материалы междунар. науч. конф., Москва, 21–23 сентября 2005 г. – М. : Наука, 2005, с. 132–135.
10. Dimov, A., Ilieva, S., Pavlova, I., Eskenazi, A. Software and Services – Research and Education in Bulgaria. // Proc. of National Conference «Education in Information Society», Plovdiv, 12–13 May, 2009, p. 9–16.
11. Trimble, R. F. The Journals of Inorganic Chemistry. // J. Chem. Doc., 1963, vol. 3, p. 79–81.
12. Бойченко, Е.Н. Периодические издания по техническим наукам в Интернете: доступность электронных ресурсов. // Науч. и техн. библиотеки, 2009, № 5, с. 31–45.
13. Garfield, E. Citation Analysis as a Tool in Journal Evaluating. // Science, vol. 178, 1972, N 4060, p. 471–479.
14. Маркусова, В. А. Информационные ресурсы для мониторинга российской науки // Вестник Российской Академии наук, 2005, т. 75, № 7, с. 607–612.
15. Маркусова, В. А. Новые наукометрические данные по России и другим странам. [Электронный ресурс] <http://www.sibai.ru/index.php?Itemid=853&id=543&option>
16. Lasswell, H. D. The Future of Political Science. – London : Tovistock Publ., 1964. – 119 p.
17. Osgood, Ch. E. The Representation Model and Relevant Research Methods. // Trends in Content-Analysis. – Urbana, 1959, p. 33–88.
18. Субботин, М. М. О логико-смысловом моделировании содержания управленческих решений. // Научное управление обществом, 1980, вып. 13, с. 203–224.

19. Стась, Е. В. Прогнозно-аналитические исследования документального информационного потока в автоматизированном режиме. // Науч. и техн. информация, Сер. 2, 1986, № 11, с. 25–31.
20. Pavlovska, E. Early Identification of Development Trends in Science. // Intern. Forum on Information and Documentation, vol.16, 1991, N 1, p. 28–33.
21. Павловска, Е. Приоритетни научни направления. Прогнозиране, откриване и оценка. – София : Унив. изд. Св. Климент Охридски, 2009. – 150 с.