

НАУКОМЕТРИЯ: МЕТОДЫ АНАЛИЗА ИНФОРМАЦИОННЫХ ПОТОКОВ

In given article modern methods of the analysis of information streams are considered, their characteristic is given, merits and demerits of these methods are revealed

Основной задачей науковедения является создание научной базы для реорганизации науки в направлении ее интенсификации. Решить эту задачу необходимо в ближайшие десятилетия, так как именно в этот временной интервал наука перейдет от стадии преимущественно экстенсивного роста к стадии преимущественно интенсивного роста. Решение этой задачи требует, чтобы науковедение опиралось на количественные методы.

Можно выделить три основные науковедческие количественные дисциплины. *Наукометрия* поставляет эмпирические количественные закономерности и результаты измерения различных параметров объектов науковедческого анализа. *Математическое моделирование* науки, используя выявленные наукометрией закономерности, описывает результаты измерения математическими зависимостями (т.е. моделями). *Теория принятия решений*, используя эти результаты измерения и математические зависимости, ищет значения параметров объектов исследования, оптимизирующие эти объекты в направлении, задаваемом целями данного конкретного исследования. Объект исследования науковедения — научная деятельность — имеет особенности, отличающие его от объектов исследования других наук.

1. Научную деятельность характеризует большая, чем в естественных науках, многофакторность явлений.

2. Большая, чем в естественных науках, субъективность результатов анализа обязывает при исследовании науки прилагать большие усилия к развитию объективных методов анализа, дающих воспроизводимые результаты.

3. Количественный анализ должен базироваться на открытых шкалах и на негауссовской математической статистике.

Наукометрия — это науковедческая дисциплина, осуществляющая воспроизводимое измерение научной деятельности и выявляющая ее объективные количественные закономерности.

В наукометрии, как и в естественных науках, используется в основном количественное измерение. В теории измерения выделяются переменные двух видов — индикаторы и латентные переменные. В качестве латентных переменных используют продуктивность ученого, его научный вклад и т.д. Перечислим наукометрические индикаторы в порядке возрастания: число научных направлений, журналов, организаций, уче-

ных, публикаций, ссылок, сленговых слов. Теперь перейдем непосредственно к методам анализа информационных потоков.

Статистический метод определим как метод, использующий в качестве наукометрических индикаторов все измерители, кроме числа публикаций, ссылок и отдельных слов. Название статистического метода условно, так как все методы имеют статистическую природу. В число измерителей науки статистического метода попадают такие измерители, как число ученых, журналов, заказов на годовые комплекты журналов в библиотеках и информационных центрах, открытий, соавторство, возрастная структура кадров и другие. Например, при исследовании зависимости числа открытий от возраста ученого, в котором они делаются, академик Вальден установил, что большинство крупнейших открытий в физике и химии конца 18-начала 20 в. было сделано учеными в возрасте 25–30 лет [1].

Второй метод — это метод подсчета числа публикаций. Индикатор, определяющий этот наукометрический метод, — это число научных продуктов, т.е. книги, статьи, отчеты и т.п.

Число публикаций может служить, как показывают результаты исследований, индикатором признания ученого, его известности, вклада в науку, продуктивности, престижа научного учреждения. Число публикаций может служить индикатором элитности ученого, его активности, интенсивности. Большое число монографий в данном научном направлении связывают с его насыщенностью, а число журнальных статей — с его конфликтностью. Ценность журнала определяют по частотам отражения их публикаций во вторичных изданиях. Скорость роста числа публикаций связывают с актуальностью данного научного направления или его перспективностью.

По скорости роста числа публикаций данного научного коллектива в данный момент предполагается определять фазы развития данного коллектива. Стационарные распределения числа публикаций также могут быть использованы для выявления проблемных научных направлений. В проблемном направлении публикации рассеяны, например, в меньшем числе журналов, чем в неproblemном.

Метод “цитат-индексов”. В основе этого метода лежит наукометрический индикатор “число цитат и ссылок”. Ссылки, как и публикации, могут подвергаться взвешиванию, в соответствии

с определяемыми средствами социологического и цитатного анализа, рангом публикаций, издания или автора, делающего эту ссылку. При цитировании может учитываться по-разному и соавторство [2,3].

Цитатный анализ возник в таких науках, как теория информационно-поисковых систем, библиометрия и др., направленных на удовлетворение информационных потребностей ученых. Практическая база для использования этого метода появилась с созданием указателей научных ссылок. Современные указатели ссылок восходят к «Празжской библиографии», изданной в 1771–1772 гг. В этой библиографии рецензий материал базируется не по рецензиям, а по рецензируемым книгам.

Другим предшественником указателей научных ссылок является изданный в 1873 г. «Указатель ссылок Федеративного законодательства» Шепарда (США). Он включает в себя списки публикаций, содержащих ссылки на данные судебные решения как претенденты.

В 1949 г. вышел указатель к американскому журналу «Анналы математической статистики», одна из частей которого содержит список статей из других журналов, на которые не менее двух раз ссылались в публикации данного журнала.

В 1963 г. в экспериментальном порядке, а с 1964 г. — регулярно стал выходить «Индекс научных ссылок», издаваемый Институтом научной информации в Филадельфии, возглавляемым Ю. Гарфилдом. Индекс Гарфилда и лежит в основе современных применений метод «цитат-индексов». Он состоит из нескольких частей. Указатель ссылок, представляющий собой основную часть индекса Гарфилда, позволяет установить, кто цитирует фиксированную работу данного автора. Фиксируется лишь фамилия первого автора работы, название работы не приводится, название журнала дается в сокращенном виде. Название работы, фамилии и адреса всех авторов можно найти в указателе источников, который представляет вторую часть индекса Гарфилда. Третья часть — это пермутационный предметный указатель. Он позволяет найти авторов, в заголовках работ которых встречается данное слово. Последняя часть индекса Гарфилда содержит результаты библиометрического анализа охватываемых цитируемых и цитируемых журналов. Каждый журнал характеризуется рядом показателей, среди которых — число ссылок в нем и на него в отдельных журналах и в сумме, так называемый импакт-фактор. Приводятся ранжировки журналов по этим показателям. Собственно индекс научных ссылок охватывает естественные науки. Институтом научной информации Гарфилда с 1970 г. выпускается и индекс ссылок по социальным наукам. При составлении индекса Гарфилда просматриваются наиболее известные

научные журналы и другие периодические и продолжающиеся издания, а также книги и патенты. Данные вводятся в ЭВМ, которая и дает в переработанном виде всю информацию. К сожалению, составители индекса Гарфилда руководствуются, прежде всего, национальными интересами. Это приводит к тому, что различные страны представлены в нем отнюдь не пропорционально их вкладам в мировую науку [4].

Следует отметить субъективность данного метода. Связи научных работ искажаются вследствие либо личностных отношений, либо пристрастия, либо языковых барьеров и т.д. Субъективность метода ссылок в определенной степени гасится на больших массивах публикаций в результате усреднения экспертиз разных авторов. Однако увеличение массива публикаций не снимает некоторых систематических ошибок. Прежде всего это преувеличение роли больших ученых и недооценка малых. При равном количестве статей работы ученых высокого ранга, как установлено, цитируются чаще, чем низкого.

Среди недостатков метода ссылок, кроме отмеченных выше, можно назвать трудность учета отрицательного цитирования, запаздывание с цитированием новых идей, детерминирование числа ссылок числом исследователей в данной области.

Преимуществом является то, что с его помощью можно проводить исследования не только на тех массивах, на которых используются менее дробные измерители, но и на существенно меньших, на которых эти менее дробные измерители дают статистически недостоверные результаты.

Темпы роста числа ссылок могут служить индикатором актуальности и перспективности направления и предметной области. Как показали исследования, метод «цитат-индексов» дает индикаторы качества научной работы, ее ценности, значимости, полезности, важности.

Метод «контент-анализ» происходит от социологического контент-анализа средств массовой коммуникации, развитого в основном американскими учеными. Процедура контент-анализа заключается в сведении рассматриваемого текста к ограниченному набору определенных элементов, которые затем подвергаются счету и анализу. Цель этого метода заключается в том, чтобы зафиксировать в материалах средств массовой информации частоту появления различных символов, определить соотношение во времени и пространстве, сравнивая различные источники информации. При помощи этого статистического метода можно проследить за историей развития специфических языков в отдельных областях знаний и оценить их роль в развитии системы научной информации. Такой анализ позволяет оценить не только широту, но и глубину проникновения новых идей и методов.

Этим методом удастся следить и за развитием отдельных идей или научных школ в пределах одной области знаний.

В основе тезаурусного наукометрического метода лежит тезаурусный метод, широко применяемый в теории и практике информационного поиска. Упор в тезаурусном методе делается на содержательный анализ.

Недостатками последних двух методов являются возможная субъективность решений при отборе слов и их группировке и большой объем трудозатрат.

И в "контент-анализе" и в тезаурусном методе процедуры измерения не обеспечивают воспроизводимости результатов счета, достаточной для наукометрических исследований, направленных на выявление количественных объективных закономерностей науки [5].

Сленговый метод является результатом развития методов "контент-анализа" и тезаурусного. В качестве индикатора берется не число символов, как это делается в контент-анализе, и не число терминов, как это делается в тезаурусном методе, а просто число слов, или так называемых словарных слов, которые следует отличать от словоформ или словоупотреблений. Этот метод опирается на традиции статистической лингвистики, которая изучает частотные распределения слов и создает частотные словари. Энтропия сленговых слов может служить индикатором проблемности данного научного направления. Коэффициенты корреляции сленговых распределений могут служить мерой индивидуального вклада ученого в соавторскую статью. Если не ограничиваться текстами двух ученых, а принимать во внимание тексты научного коллектива, то коэффициент корреляции между сленговыми распределениями текстов данного ученого, с одной стороны, и текстами данного коллектива — с другой, может служить мерой степени связанности тематики исследований и научных интересов ученого и научного коллектива.

Наукометрические исследования документальных информационных потоков были начаты в 60-х гг. в нашей стране в области химии. Однако эти исследования не получили практи-

ческого распространения в других областях в силу многих причин. Прежде всего из-за трудности получения достоверных потоков с ретроспективой не менее 12 лет и полнотой представления не менее 90%. При этом источником получения информации служили, как правило, реферативные журналы, а в последнее время — их машиночитаемые аналоги — интерактивные библиографические базы данных. А также снижение интереса научной общественности к применению наукометрических методов связано с тем, что большинство информетрических методов было направлено на подтверждение данных, хорошо известных экспертам в данных областях.

В настоящее время, когда через телекоммуникационные центры стали доступны более 3000 баз данных, работающих в режиме онлайн, включая полнотекстовые базы с фактографическими данными, высказано мнение о "достоверности использования содержащейся в них информации для научных прогнозов". В последнее время за рубежом резко возрос интерес к прагматическим аспектам применения результатов наукометрических исследований в целях выработки научно обоснованной информационной политики, что особенно важно в условиях сокращения ассигнований на научные исследования.

Литература

1. Налимов В. В., Мульченко В. М. Наукометрия. — М.: Наука, 1969.
2. Добров Г. М. Наука о науке. — Киев: Наукова думка, 1970.
3. Ахманова О. С., Мельчук И. А., Падучева Е. Ф., Фрумкина Р. М. О точных методах исследования языка. — М.: МГУ, 1961.
4. Хайтун С. Д. Наукометрия. Состояние и перспективы. — М.: Наука, 1983.
5. Коренной А. А. Информетрическое моделирование и прогнозирование развития науки // Вопросы информационной теории и практики. — М.: ВИНТИ, 1986. — № 56. — С. 19–33.
6. Горькова В. И. Системно-структурные исследования документальных информационных потоков: Автореферат дисс. на соиск. уч. степ. доктора техн. наук. — М.: ВИНТИ, 1970.