

УДК 002.2
ББК 73

ФОРМАЛИЗОВАННЫЕ МЕТОДЫ АНАЛИЗА ДОКУМЕНТАЛЬНЫХ ИНФОРМАЦИОННЫХ ПОТОКОВ

© Н.С. Редькина, 2005

*Государственная публичная научно-техническая библиотека
Сибирского отделения Российской академии наук
630200, г. Новосибирск, ул. Восход, 15*

Рассматриваются инструментарий количественного изучения и оценки документальных информационных потоков: специальные методы и подходы, приемы, законы (Распределение Ципфа, Закон рассеяния Бредфорда, уточненная математическая формулировка Закона рассеяния Бредфорда, предложенная Б.К. Бруксом, Закон обратного квадрата А. Лотка, описывающий распределение авторов в зависимости от количества опубликованных ими статей), позволяющие достаточно точно и объективно осуществлять формализованный анализ результатов научной деятельности.

Ключевые слова: документальный информационный поток; методы; библиометрия; информетрия; вебометрия; наукометрия.

Терминологический анализ

Возрастающее значение специальных методов количественного исследования документальных информационных потоков (ДИП) обусловлено их объективной ролью в оценке динамично развивающегося научного потенциала страны, выбора научных приоритетов, исследования закономерностей функционирования научных систем, изучения вклада и продуктивности отдельного ученого/коллектива в развитие науки, прогнозирования дальнейших исследований.

Документальные информационные потоки изучаются такими активно развивающимися направлениями, как библиометрия, информетрия, наукометрия, вебометрия, киберметрия, документометрия, медиаметрия и др., позволяющими анализировать закономерности развития документопотоков. Соответственно перечень методов носит название «библиометрических» /34, 40, 54/, «информетрических» /7, 22, 36/, «наукометрических» /39, 51/, «информационных» /52/. Различие этих методов, по мнению В.И. Горьковой, в «несходстве предметов и сфер применения, объективных количественных закономерностей, а также целей практической деятельности» /36/. Термины являются схожими по значению. В то же время они используются для описания различных подходов и, как отмечают W.W. Hood и C.S. Wilson: «У каждого своя история возникновения и сфера использования» /18. – С. 294/. При этом они не ограничиваются рамками только научной продукции и коммуникаций, а могут применяться в бизнесе, политике и других областях.

Изначально количественные исследования документопотока называли «статистической библиографией». Термин ввел в 1923 г. Е. Хамл, применив его к ранжированию стран по числу журнальных статей в определенных областях. Затем был введен термин «библиометрия», который, по мнению Ю. Гарфилда, впервые встречается в работах П. Отле (1934) и А. Причарда (1969) /15/. Концептуальные положения о библиометрии, ее месте в системе наук и используемых методах остаются, несмотря на многочисленность исследований, весьма актуальными проблемами и сегодня. Зарубежные исследователи чаще всего указывают на работу Ф. Коула и Н. Ильса 1917 г. /8/, в которой был проведен статистический анализ литературы по сравнительной анатомии. О. Воверене, вслед за историком науки С.В. Альтшуллером /31, 34/ отмечает, что несправедливо забыто имя русского ученого П. Вальдена, который еще в 1911 г. впервые применил метод анализа цитирования для изучения вклада ученых отдельных стран в развитие химии /32/ и, таким образом, должен быть признан родоначальником данного метода – наиболее часто используемого в настоящее время при оценке научной результативности. А первые попытки количественного изучения потока отечественной литературы были сделаны А. Шторхом и Ф. Аделунгом /58/. Ими был проанализирован собственный указатель, отражающий литературу 1801–1806 гг. по разным параметрам.

В 1969 г. В.В. Налимов и З.М. Мульченко ввели термин «наукометрия» /51/, в дальнейшем получивший распространение как «*scientometrics*». Он относится к области науковедения, изучающей

закономерности функционирования и развития науки, структуру, динамику научной деятельности, взаимодействие науки с другими сферами материальной и духовной жизни общества, т.е. исследует процессы, явления научной деятельности в более широком контексте.

В 1979 г. в статьях немецких авторов L. Blaskert, S. Siegel и O. Nacke /4, 22/ появился термин «информетрия», который первоначально был определен как комплекс математических методов для исследования объектов информационной науки, описания и анализа их свойств, а также закономерностей в целях оптимизации этих объектов при принятии решений. Десятью годами позже под информетрией стали понимать использование разнообразного математического аппарата для анализа, выявления закономерностей, формулировки законов информационной деятельности и научной информации, а также для принятия решений в информационной практике /24/. Специалист Института научной и технической документации и информации (Гавана, Куба) Мельвин Моралес выделяет следующие характерные аспекты информетрии /50/:

1. Количественный рост литературы.
2. Старение и рассеяние информации.
3. Эффективность информационных продуктов и услуг в сфере производства, науки и техники.
4. Эффективность информационной системы и информационных органов в их совокупности.
5. Роль различных видов документов как средств научной коммуникации.
6. Роль неформальных каналов научной коммуникации.
7. Релевантность и пертинентность информации.
8. Ранжирование периодических и продолжающихся изданий по различным аспектам.
9. Тематическая близость периодических и продолжающихся изданий.
10. Значение практики цитирования между учеными и ее развитие.

11. Внутродисциплинарные и междисциплинарные связи на основе библиографических ссылок.

В основе вновь возникших родственных терминов, таких как «киберметрия», «интернетометрия» и «вебометрия», также находятся библиометрические приемы, только связаны они с исследованием новых форм представления информации. Так, под *киберметрией* понимают анализ потоков киберинформации (всех видов медиаинформации) с использованием наукометрических, библиометрических и информационных подходов /12/. Издается журнал с одноименным названием – «Cybermetrics: International Journal of

Scientometrics, Informetrics and Bibliometrics» (ISSN 1137-5019), отражающий развитие электронных коммуникаций и включающий анализ результатов научной деятельности, проведенных в сети World Wide Web (WWW) /11/.

Термин «вебометрия» впервые был применен в 1996 г. R.H. Abraham /1/ и R.R. Larson /20/ в целях построения когнитивных карт и математических моделей WWW. Первоначально предложенная технология подразумевала «bibliometrics on the

Web», т.е. использование библиометрии при анализе Web. На сегодняшний день существует, как минимум, два его толкования. Первое – вебометрия – это статистика, цитирование веб-сайтов /23/.

Вторая позиция, которую поддерживают Lennart Vjorneborn и Peter Ingwersen /3/, заключается в более углубленной теоретической проработке данной дефиниции. Авторы определяют вебометрию как область, изучающую с помощью библиометрических методов характер и свойства сети Web. Некоторые исследователи предлагают под вебометрией понимать методологию исследования World Wide Web /2/.

Констатируем: изначально библиометрия формировалась со своим инструментарием и подходами, в последующих направлениях применяются чаще всего те же количественные методы, но анализирующие другие формы представления документопотоков.

Сущность формализованных методов анализа ДИП заключается в подсчете, сочетании, интерпретации, сравнении некоторых элементов документопотока, воздействующих друг на друга в процессе развития научного направления.

Методы количественного анализа, базирующиеся на исследовании вторичных источников информации

Основными параметрами количественного анализа вторичного документопотока, в частности метода анализа реферативных журналов, предложенного в 1981 г. А. Причардом и Г. Виттингом /25/, являются: объем документопотока, динамика и тенденции его формирования, закономерности концентрации и рассеяния, изменение во времени, по отраслям науки, научным направлениям, по странам и коллективам, взаимосвязям различных показателей (число публикаций – число авторов; число публикаций – страна – тематическое направление и др.). При этом анализ динамики развития документопотока, рост, старение, ранговое распределение основываются на классических законах и моделях Дж. К. Ципфа (Распределение Ципфа) /29/, С.К. Бредфорда (Закон рассеяния Бредфорда) /5/,

Изначально библиометрия формировалась со своим инструментарием и подходами

Б.К. Брукса (уточненная математическая формулировка Закона рассеяния) /6/, А. Лотки (Закон обратного квадрата, описывающий распределение авторов в зависимости от количества опубликованных ими статей) /21/. Систематизированное обобщение соответствующих законов известно по работам Дж. Солтона, А.И. Яблонского /55, 59/.

Закон обратного квадрата А. Лотки

Важное эмпирическое исследование массива публикаций в целях анализа научной продуктивности, принадлежащее А. Лотке, появилось в 1926 г. /21/. Автор подсчитал число ученых, написавших одну, две и т.д. статьи, на массиве публикаций, отраженных в издании «Chemical Abstracts» за 1907–1916 гг., а также проанализировал справочник открытий по физике за 1600–1900 гг. Ф. Ауэрбаха «Geschichtstafeln der Physik» (1910). В обоих случаях для числа $n(x)$ ученых, написавших x статей, было получено следующее распределение научной продуктивности:

$$n(x) = \frac{A}{x^2}, \quad (1)$$

где A – число ученых, написавших всего одну статью; $x = 1, 2, \dots, x_{\max}$ (максимальная продуктивность ученого).

Данное распределение носит название Закона Лотки или Закона обратных квадратов: число ученых, написавших данное число статей, обратно пропорционально квадрату этого числа статей. Закон многократно проверялся на различных информационных массивах и оказался справедливым для всего диапазона научной продуктивности, кроме самых низких ($n = 1$) и самых больших значений («краевые эффекты»). Таким образом, А. Лотка доказал, что ученые отличаются по числу опубликованных статей, т.е. распределение научной продуктивности носит асимметрический, неравномерный характер. Инвариантность, устойчивость этого закона во множестве научных областей позволяет говорить о нем как об одном из важнейших библиометрических законов.

Закон распределения Дж. Ципфа

Дж. Ципф /29/, собрав огромный статистический материал, открыл закон распределения слов естественного языка. Он установил, что если к какому-либо достаточно большому тексту составить список всех встретившихся в нем слов, затем расположить их в порядке убывания частоты их встречаемости и пронумеровать от 1 до R , то для любого слова произведение его порядкового номера (R) на эту частоту будет величиной постоянной, имеющей примерно одинаковое численное значе-

ние для слов из рассматриваемого списка. Выражается закон Ципфа следующим образом:

$$fr = c,$$

где f – частота встречаемости слова в тексте; r – ранг (порядковый номер) слова в списке; c – эмпирически постоянная величина.

Полученная подчиненность графически выражается гиперболой. Исследовав таким образом самые разнообразные языки, Дж. Ципф для каждого построил указанные зависимости, при этом все кривые имели одинаковую форму – «гиперболической лестницы», т.е. при замене одного текста другим общий характер распределения не менялся. Ранговому гиперболическому закону Ципфа подчиняется распределение не только слов, но и различных явлений социального характера, в том числе распределение ученых (по числу опубликованных ими работ). Ранговый подход позволяет достаточно точно проранжировать распределение ученых по уменьшению продуктивности.

Закон рассеяния С.К. Брэдфорда

В дальнейшем С. Брэдфорд /5/ сформулировал закономерность распределения публикаций по изданиям. Согласно открытому им закону периодические издания можно разделить на группы, равные по общему числу статей в каждой (т.е. на совокупности журналов с одинаковым суммарным числом статей по данной тематике). Первая, наиболее «продуктивная», зона (ядро массива) содержит малое число специальных журналов; вторая, более обширная, содержит статьи в научных журналах по смежным специальностям и, наконец, последняя зона («периферия») включает наиболее обширное число журналов по разным областям, в которых рассеяны немногочисленные статьи по данной тематике.

Если научное направление достаточно развито, то журналы по соответствующей проблеме обычно распадаются на три группы, каждая из которых содержит по одной трети публикаций. В каждой группе должно быть одинаковое количество публикаций, но число периодических изданий в них – различно. С. Брэдфордом была построена кривая рассеяния публикаций, в которой виден начальный участок, т.е. зона «ядерных» периодических изданий, а далее – прямолинейный участок. Выполнимость данного закона считается косвенным признаком «классичности» области исследования и используется для оценки развития и самостоятельности научного направления.

Б. Викери /28/ доработал модель С. Брэдфорда. Он определил, что журналы, выстроенные в порядке уменьшения в них числа статей по конкретной проблеме, можно разбить не на три зоны, а на нуж-

ное количество зон. Закономерность имеет следующее выражение:

$$T_x : T_{2x} : T_{3x} : T_{4x} \dots = 1 : a : a^2 : a^3 \dots,$$

где x – количество статей в каждой группе; a – коэффициент рассеяния, постоянная величина для данного предмета и времени; T_x – количество журналов, содержащих x статей по данному предмету; T_{2x} – количество журналов, содержащих $2x$ статей, т.е. сумма наименований журналов первой и второй групп; T_{3x} – количество журналов, содержащих $3x$ статей; T_{4x} – количество журналов, содержащих $4x$ статей и т.д.

Формула Брукса, предложенная в 1969 г. /6/, также уточняет закон Брэдфорда и состоит из двух уравнений:

$$R_{(n)} = an^b \quad (1 \leq n \leq c) \\ R_{(n)} = k \log_b n/s \quad (c \leq n \leq N),$$

где $R_{(n)}$ – суммарное число публикаций по данной отрасли, содержащееся в n периодических изданиях, проранжированных в порядке уменьшения «продуктивности»; n – ранг (порядковый номер) периодического издания в списке по убывающей продуктивности; a – количество релевантных публикаций в самом продуктивном журнале; c – число периодических изданий в «ядре»; N – общее количество периодических изданий в ранговом списке; b, k, s – эмпирические коэффициенты.

Первое уравнение описывает начальный участок кривой, который имеет экспоненциальный характер, а второе – прямолинейную часть кривой Брэдфорда. В зависимости от соотношения коэффициентов делают заключение о состоянии научного направления. Установлено, что для узкоспециализированных областей науки $s \leq 1$, а величина $c \geq 3$ и связана с s прямой зависимостью /43/.

Б. Брукс объясняет механизм функционирования Закона рассеяния С. Брэдфорда /6/, предлагая следующую гипотезу. Первые статьи по новому предмету исследования публикуют в нескольких наиболее соответствующих тематике периодических изданиях. В результате данные журналы помещают, по мере развития рассматриваемого предмета, все больше подобных статей. В то же время статьи по этому предмету начинают публиковаться и в других периодических изданиях. Если предмет продолжает развиваться, то со временем складывается брэдфордовское «ядро» периодических изданий, в которых помещается наибольшее число публикаций по данной теме.

Истинность вышеперечисленных законов подтверждается огромным количеством последующих исследований. Вместе с тем обобщение практиче-

ского опыта и эксперименты показали, что, например, Закон рассеяния С. Брэдфорда строго выполняется лишь при соблюдении следующих условий /48. – С. 181/:

1. Отрасль, тема или предмет должны быть четко определены.

2. Рассматриваемый список периодических изданий по такой отрасли, теме или предмету, а также учет содержащихся в этих периодических изданиях релевантных публикаций должны быть полными.

3. Время, за которое рассматриваются периодические издания, должно быть четко определено и учтены все помещенные в эти издания релевантные публикации.

Большая заслуга Дж. Ципфа, С. Брэдфорда и А. Лотки состоит в том, что они положили начало строгому исследованию документальных потоков, которые представляют собой совокупности научных документов-публикаций и неопубликованных материалов. Дальнейшие исследования, среди которых видное место занимают работы В.И. Горьковой /36, 37/, показали, что можно определять количественные параметры не только совокупностей научных документов, но и совокупностей элементов признаков научных документов: авторов, терминов, индексов классификационных систем, наименований изданий, т.е. наименований элементов, характеризующих содержание научных документов. Например, можно расположить журналы в порядке убывания числа печатающихся в них авторов, в порядке убывания средней величины публикуемых в них статей или упорядочить совокупность документов по любому ее элементу.

В.И. Горькова /37/ доказала, что одним из критериев, позволяющих оценить развитие отдельных научных дисциплин и их составляющих, может являться средняя скорость роста количества публикаций за интервал времени Δt . Автором предложена следующая формула:

$$V = \frac{\Delta N}{N \Delta t}, \quad (2)$$

где N – число публикаций в базисном году; ΔN – прирост за период Δt ; V – скорость роста числа публикаций, определяемая их приростом (ΔN) во времени (Δt) по отношению к базисному году (N) по данным экспериментальных исследований за ряд лет (более 5).

Данные законы и установленные закономерности являются наиболее распространенными в исследованиях документопотоков, связаны с важными явлениями в научных коммуникациях и имеют большое практическое значение. Значение законов распределения состоит в том, что их основатели

положили начало системному исследованию документопотоков. Можно утверждать, что эти законы являются методологической основой при изучении документопотоков. Их применение в управлении массивами журналов или для оценки научного влияния журналов, авторов, статей, исследований – важное условие информационной поддержки научной деятельности.

Метод моделирования лексического анализа документальных БД

Метод оценки тенденций развития научного направления, основанный на анализе лексики проблемно-ориентированных БД, предложен Е.Ю. Павловска /52/. Автор предлагает выбрать определенное научное направление, для которого в БД можно проследить моменты зарождения научных направлений исследований по какой-либо проблеме, их развитие, распад или трансформацию в новые направления, попытаться выявить закономерности в динамических характеристиках информационного потока в различные периоды его «жизни» и, когда такие закономерности будут найдены, проверить высказанную гипотезу либо на других массивах, либо в другом временном интервале.

Контент-анализ

При исследовании развития научных направлений используется группа методов, основанных на анализе текстов документов, и берущих начало от социологического метода контент-анализа /17, 47, 53/. Различные модификации контент-анализа направлены на получение объективной информации о некоторой совокупности однородных документов (информационного потока) путем фиксации существенных характеристик содержания и их количественного описания. Контент-анализ плодотворен при изучении больших объемов текста, в частности интернет-ресурсов и полнотекстовых изданий и коллекций. Единицей анализа может являться слово или некоторая совокупность слов, выражающая проблему (предмет, отрасль, направление и др.), автор, географическая рубрика и др.

Метод совместной встречаемости ключевых слов

К середине 50-х гг. XX в. исследователи стали обращать внимание не только на простое наличие или отсутствие единиц анализа в тексте, но и на связи между ними, т.е. совместную встречаемость (соосцигенность) слов различных категорий. Для этой цели подсчитывают обычный коэффициент корреляции, который указывает на связи между категориями и знак этой связи. В результате может оказаться, что для некоторых категорий наблюдается тенденция их совместного употребления, а для других – наоборот. На этой основе создан так на-

зываемый метод «co-word analysis», предложенный группой исследователей /10/ в 1984 г. (J.P. Courtial, M. Callon, M. Sigogneau) и применимый к различным информационным ресурсам. Каждое поле в полнотекстовом источнике или БД может быть использовано для анализа, чаще это дескрипторы или ключевые слова. Метод является полезным инструментом для картирования науки, так как способствует выявлению ассоциаций среди дескрипторов или терминов и на их базе построению сетей, отражающих картину эволюции любой дисциплины. Для этой цели создаются указатель включения (inclusion index) и указатель близости (proximity index), используемые для измерения силы связей между отраженными понятиями. С помощью этих указателей понятия объединяются в кластеры и представляются на сетевых картах. Результаты исследований позволяют делать выводы о постоянстве или «смене акцентов» в определенных темах исследовательских областей, проследить «созревание» или утрату значения научных проблем. Данный метод, по мнению ученых /9, 16/, применим ко многим тематическим областям при наличии достаточного «корпуса» текстов и поддерживающего его программного обеспечения.

Метод семантического спектра

Метод семантического спектра позволяет в графической форме исследовать динамику ключевых слов во времени и рассматривать разнообразные аспекты структурных изменений в различных областях /30/. Метод также основан на анализе частот встречаемости ключевых слов. По мере развития научного направления их частота растет, прекращение работ по какому-либо научному направлению приводит к уменьшению частоты встречаемости определенных терминов. Таким образом, в совокупности выделяют три составляющих: низкочастотная – соответствующая развивающемуся научному направлению; высокочастотная – присущая установившейся терминологии, низкочастотная – характерная для научных направлений, работы по которым прекращаются или переходят в стадию производства. Анализ динамики «семантического спектра» позволяет выявлять эти составляющие и, соответственно, делать выводы об интенсивности развития научных направлений или их возникновении; другими словами, динамика «семантического спектра» отражает структурно-частотные изменения в документальных массивах, которые в свою очередь являются индикатором возникновения и развития научных направлений.

Метод логико-смыслового моделирования

Еще один разработанный метод, основанный на анализе текстов документов, – метод логико-смыс-

лового моделирования, предложенный М. Субботиним /56/. Он основан на использовании в качестве исходных элементов любых высказываний, которые могут быть выражены отдельным словом, словосочетанием или целым предложением. Для каждого высказывания выделяются все его непосредственные логические связи с другими высказываниями в данной предметной области, образуя неориентированный граф, вершинами которого являются высказывания, а ребрами – связи между ними, – так называемый ЛС-граф, рассматриваемый как логико-смысловая модель данной области. Метод позволяет следить за развитием предметной области: многосвязные и центральные высказывания в ЛС-графах свидетельствуют о наиболее важных проблемах на данном этапе развития, смещения этих центров на ЛС-графах следующих этапов свидетельствуют об определенных тенденциях развития.

При формировании логико-смысловых графов исследователь должен пользоваться определенными критериями и процедурами, чтобы отличать прямую связь от косвенной. Смежными по смыслу считаются лишь те понятия и утверждения, которые можно объединить при помощи логических связок («есть», «является причиной», «поэтому», «в этих целях» и т.п.). Сегодня логико-смысловые графы приблизились к гипертекстам, их можно читать, вставляя при переходе к смежному узлу соответствующую логическую связку.

Анализ цитирования

Система методов изучения документопотока на основе анализа цитирования, применяемых для получения структурной картины состояния научных исследований и оценки результативности ученых, разрабатывается интенсивно. В данном случае объектом анализа является поток цитируемых и цитирующих публикаций, а предметом – количественные его характеристики. Анализ цитирования чаще всего проводят на базе специальных вторичных источников информации – указателей научных ссылок, например в изданиях Института научной информации США («Science Citation Index» и др.). Ссылки, являясь средством научной коммуникации, позволяют проследить развитие научного направления, дают общие сведения о проблеме, отражают литературу, создающую контекст данной работы /45/. Исследование внутренней структуры областей знания, выявление исторических особенностей и тенденций развития науки и техники, основанных на анализе сетей цитирования публикаций, подробно освещены в ряде работ Ю. Гарфида /13, 14, 35/, В.А. Маркусовой, И.В. Маршковой /41–46/.

Вместе с тем исследователи отмечают ряд значимых недостатков анализа цитирования. Напри-

мер, существует такой феномен, получивший название «Спящая красавица в науке», когда статья после выхода в свет остается незамеченной длительное время, но потом неожиданно привлекает к себе большое внимание /27/. Примером такой работы, сильно опередившей свое время, является исследование Г. Менделя по генетике растений, опубликованное в 1865 г. и остававшееся в тени 34 года. Спорности и «опасности» индекса цитируемости как главного критерия оценки качества научной деятельности посвящена работа О.В. Михайлова /49/. Одним из подходов к более достоверному анализу цитирования является дифференцированный подход. О его необходимости впервые стал говорить А.И. Уемов /57/. В дальнейшем было предложено выделять пять типов цитирования: авторитетное, конструктивное, информирующее, критическое и негативное цитирование. Эта классификация была применена Э. Ваттером для анализа структуры цитирования в области информатики /33/. Еще более детальная классификация цитирований была предложена Е.Д. Гражданниковым и Т.В. Сорокиной в 1976 г. /38/. Сегодня в науковедении выделяют положительное (справочное, обзорное, конструктивное и др.) и отрицательное (критическое, отвергающее, обвиняющее и др.) цитирования, позволяющие более объективно проанализировать цитируемый документопоток.

На основе принципов цитирования предложены дополнительные методы. Наиболее употребляемые среди них следующие: метод ко-цитирования и кластерный анализ, метод библиографического сочетания.

Метод ко-цитирования и кластерный анализ

Используя метод совместного цитирования двух публикаций третьей, так называемого **ко-цитирования** (от англ. «co-citation»), определяют тенденции развития коммуникационных процессов в науке. Данный метод, разработанный одновременно Г. Смоллом и И. Маршковой /26, 44/, позволяет сгруппировать статьи в зависимости от принадлежности к той или иной тематике; вносить коррективы в полученную структуру в процессе выявления новых библиографических ссылок, отражающих изменения во времени исследуемого направления, и тем самым может служить инструментом оперативной коррекции научных проектов.

В основе метода ко-цитирования лежит принцип выделения взаимосвязи между двумя публикациями на основе цитирования их одними и теми же документами. Родство публикаций или авторов публикаций определяется числом работ, цитирующих одновременно обе статьи, т.е. являются невидимые связи, которые при наглядном представлении образуют смысловые сгустки (кластеры).

Начиная с 1981 г. метод ко-цитирования используется в Институте научной информации (США) при построении кластеров публикаций, отражающих активные исследовательские фронты в различных областях знаний. **Кластерный анализ** является действенным инструментом картирования различных свойств, выявления активных исследовательских фронтов или динамики исторического развития определенных научных направлений или дисциплин. Как отмечает С.А. Рожков, «кластер – это своего рода «образ» исследовательской области, созданный всем стоящим за ним мировым научным сообществом ученых, работающих в данном направлении» /54. – С. 29/.

Метод библиографического сочетания

Выявлять связи в документопотоке предлагается также методом библиографического сочетания («bibliographic coupling»), который был предложен М. Кесслером в 1963 г. /19/. В основе этого метода лежит принцип выделения взаимосвязи между двумя публикациями на том основании, что цитируется один и тот же документ, причем интенсивность их взаимосвязи определяется числом библиографических ссылок, общих для обеих публикаций. В соответствии с этим методом сила связи двух публикаций определяется числом приведенных в этих публикациях общих (одинаковых) ссылок, и на этой основе устанавливается связь между публикациями (на основании общих в них ссылок). Еще один вид взаимосвязи между двумя публикациями посредством ссылок – анализ предметного направления, который проводится путем изучения количества ссылок, содержащихся в первой публикации, на работы, содержащиеся во второй. По методу Кесслера две публикации прочно связаны и эта связь не меняется при появлении новых публикаций, т.е. не зависит от изменений, происходящих в науке. Исходя из этого, такую связь называют ретроспективной, в отличие от другого метода – ко-цитирования, где эта связь является проспективной и позволяет исследовать совместно цитируемые публикации в новых работах (рис. 1, 2).

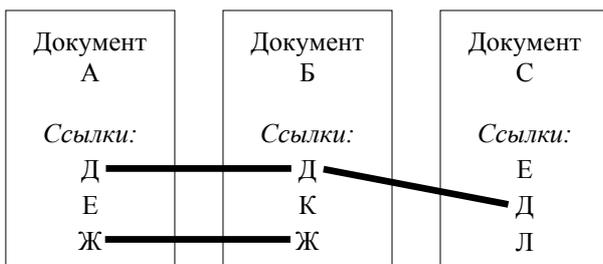


Рис. 1. Библиографическое сочетание документов



Рис. 2. Ко-цитирование документов

Заключение

Все перечисленные методы анализа ДИП имеют свои особенности, достоинства и недостатки. Каждый метод имеет более или менее определенную область применения, в границах которой он эффективен. Методы анализа количественных характеристик первичных документов можно назвать перспективными, учитывая увеличение числа полнотекстовых информационных ресурсов. Методы анализа вторичных источников информации, основанные на классических библиометрических законах, анализе видовой структуры и количественных характеристик документопотоков (мобильность, стабильность, скачкообразность), позволяют оперативно следить и выявлять развивающиеся и затухающие направления. Вопрос о целесообразности применения того или иного метода решается в зависимости от цели исследования, изучаемой проблемы, степени ее разработанности в литературе, от тех информационных ресурсов, которыми располагает исследовательский коллектив.

Список литературы

1. *Abraham, R.H.* Webometry: measuring the complexity of the World Wide Web (1997) [Electronic resource] / R.H. Abraham. – Режим доступа: <http://www.ralph-abraham.org/articles/MS%2385.Web1/>. – Title from the screen.
2. *Almind, T.C.* Informetric analyses on the World Wide Web: A methodological approach to «webometrics» / T.C. Almind, P. Ingwersen // *J. of Doc.* – 1997. – N 9. – P. 22–39.
3. *Bjorneborn, L.* Perspectives of webometrics / L. Bjorneborn, P. Ingwersen // *Scientometrics.* – 2001. – Vol. 50, N 1. – P. 65–82.
4. *Blackert, L.* Ist in ger wissenschaftlichen information Platz für die Informetrie? / L. Blackert, S. Siegel // *Wiss. Zeitschrift Tech. Hochschule.* – 1979. – Bd 25, Hf 6. – S. 187–199.
5. *Bradford, S.C.* Documentation / S.C. Bradford. – L.: Crosly Lockwood, 1948. – 156 p.
6. *Brookes, B.C.* Bradford's law and the bibliography of science / B.C. Brookes // *Nature.* – 1969. – Vol. 224, N 5223. – P. 953–956.
7. *Brookes, B.C.* Towards informetrics: Haitun, Laplace, Zipf, Bradford and the Alvey programme / B.C. Brookes // *J. Doc.* – 1984. – Vol. 40, N 2. – P. 120–143.
8. *Cole, F.J.* The history of comparative anatomy. Part 1. A statistical analysis of the literature / F.J. Cole,

- N.B. Eales // *Science Progress*. – 1917. – Vol. 11, N 10. – P. 1229–1234.
9. Coulter, N. Software engineering as seen through its research literature: A study in co-word analysis / N. Coulter, I. Monarch, S. Konda // *J. Amer. Soc. Inf. Sci.* – 1998. – Vol. 49, N 13. – P. 1206–1223.
 10. Courtial, J.P. Is indexing trustworthy – classification of articles through co-word analysis / M. Callon, M. Sigogneau // *J. Inf. Sci.* – 1984. – Vol. 9, N 2. – P. 47–56.
 11. Cybermetrics: *International Journal of Scientometrics, Informetrics and Bibliometrics* [Electronic resource]. – Режим доступа: <http://www.cindoc.csic.es/cybermetrics/news.html>. – Title from the screen.
 12. Dahal, T.M. Cybermetrics: The use and implications for Scientometrics and Bibliometrics; A study for Developing Science & Technology Information System in Nepal [Electronic resource] / T.M. Dahal // III National Conference on Science & Technology (March 8–11, 1999) / Royal Nepal Academy of Science and Technology. – Режим доступа: <http://www.panasia.org.sg/nepalnet/ronast/cyber.html>. – Title from the screen.
 13. Garfield, E. Citation index of science / E. Garfield // *Amer. Doc.* – 1955. – Vol. 6, N 2. – P. 109–117.
 14. Garfield, E. *Essays of an information scientists*. Vol. 3 (1977–1978) / E. Garfield. – Philadelphia: ISI Press, 1980. – 892 p.
 15. Garfield, E. The intended consequences of Robert K. Merton / E. Garfield // *Scientometrics*. – 2004. – Vol. 60, N 1. – P. 51–61.
 16. He Qin Knowledge discovery through co-word analysis / Qin He // *Libr. Trends*. – 1999. – Vol. 48, N 1. – P. 133–159.
 17. Holsti, O. *Content analysis for the social sciences and humanities* / O. Holsti. – Addison: Wesley, 1969. – 216 p.
 18. Hood, W.W. The literature of bibliometrics, scientometrics, and informetrics / W.W. Hood, C.S. Wilson // *Scientometrics*. – 2001. – Vol. 52, N 2. – P. 291–314.
 19. Kessler, M.M. Bibliographical coupling between scientific papers / M.M. Kessler // *Amer. Doc.* – 1963. – Vol. 14, N 1. – P. 10–25.
 20. Larson, R.R. Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace [Electronic resource] / R.R. Larson. – Режим доступа: <http://sherlock.berkeley.edu/asis96/asis96.html>. – Title from the screen.
 21. Lotka, A.J. The frequency distribution of scientific productivity / A.J. Lotka // *J. Washington Acad. Sci.* – 1926. – Vol. 16, N 12. – P. 317–323.
 22. Nacke, O. Informetria: Name für eine neue Disziplin / O. Nacke // *Nach. Dok.* – 1979. – Bd 30, Hf 6. – S. 219–226.
 23. Owen, Th. Webometric analysis of department of librarianship and information science / Th. Owen, P. Willett // *J. Inf. Sci.* – 2000. – Vol. 26, N 6. – P. 421–428.
 24. Pindlowa, W. Wokol informetrii, bibliometrii i naukoimetrii / W. Pindlowa // *Aktual. Probl. Inf. i dok.* – 1989. – Vol. 34, N 1–2. – С. 3–7.
 25. Prichard, A. *Bibliometrics: A – Bibliography and Index. 1: 1874–1959* / A. Prichard, G. Witting. – Watford: Allm Books, 1981. – 160 p.
 26. Small, H. Co-citation in the scientific literature: A new measure of the relationship between two documents / H. Small // *J. Amer. Soc. Inf. Sci.* – 1973. – Vol. 24, N 4. – P. 265–269.
 27. Van Raan Anthony, F.J. *Sleeping Beauties in Science* / F.J. Van Raan Anthony // *Scientometrics*. – 2004. – Vol. 59, N 3. – P. 467–472.
 28. Vickery, B.C. *Information system* / B.C. Vickery. – L.: Butterworths, 1973. – 350 p.
 29. Zipf, G.K. *Human behavior and the principle of least effort: an introduction to human ecology (1902–1950)* / G.K. Zipf. – Cambridge: Addison-Wesley Press, 1949. – 573 p.
 30. Агеев, Б.А. Анализ развития научных направлений методом «семантического спектра» / Б.А. Агеев, С.Б. Черноног, С.Б. Розуменко // *НТИ. Сер. 1. Орг. и методика информ. работы*. – 1990. – № 5. – С. 16–18.
 31. Альтшуллер, С.В. Вальден П.И. и язык научных ссылок / С.В. Альтшуллер // *Природа*. – 1969. – № 8. – С. 119.
 32. Вальден, П.И. О развитии химии в России // *Дневники 2 Менделеевского съезда (21–28 дек. 1911 г.)* / П.И. Вальден. – СПб.: Б.г. – № 4–8. – С. 124–141.
 33. Ваттер, Э.И.Ф. О структуре научных ссылок / Э.И.Ф. Ваттер // *НТИ. Сер. 2. Информ. процессы и системы*. – 1975. – № 5. – С. 17–20.
 34. Воверене, О. Библиометрия – структурная часть методологии информатики / О. Воверене // *НТИ. Сер. 1. Орг. и методика информ. работы*. – 1985. – № 7. – С. 1–5.
 35. Гарфилд, Ю. Индексы научного цитирования, их анализ и использование: Реф. / Ю. Гарфилд; Ред. Т.В. Горбунова. – М., 1988. – 22 с. – (Специализир. информ. / АН СССР, ИНИОН; 103/1056).
 36. Горькова, В.И. Информетрия: (Количественные методы в НТИ) / В.И. Горькова. – М.: ВИНТИ, 1988. – С. 3–326. – (Итоги науки и техники. Сер. Информатика; Т. 10).
 37. Горькова, В.И. Тенденции развития электроэнергетики по данным анализа документальных информационных потоков / В.И. Горькова. – М.: Информ-электро, 1973. – 81 с.
 38. Гражданников, Е.Д. Наукометрические методы библиографического поиска / Е.Д. Гражданников, Т.В. Сорокина. – Новосибирск, 1976. – 19 с. – (Препр. / Ин-т теплофизики Сиб. отд-ния АН СССР; 06 – 76).
 39. Грановский, Ю.В. Наукометрический анализ информационных потоков в химии / Ю.В. Грановский. – М.: Наука, 1980. – 141 с.
 40. Лазарев, В.С. Библиометрия / В.С. Лазарев // *Вопр. библиографоведения и библиотековедения*. – Минск, 1991. – Вып. 12. – С. 3–18.
 41. Маркусова, В.А. Еще раз об оценках в науке с помощью статистических данных / В.А. Маркусова // *НТИ. Сер. 1. Орг. и методика информ. работы*. – 2000. – № 8. – С. 17–24.
 42. Маршакова, И.В. Анализ и оценка химических журналов по базам данных Института научной информации, США / И.В. Маршакова // *НТИ. Сер. 1. Орг. и методика информ. работы*. – 1993. – № 10. – С. 19–30.
 43. Маршакова, И.В. Выявление тенденций развития науки и техники путем анализа документальных массивов / И.В. Маршакова // *НТИ. Сер. 2. Информ. процессы и системы*. – 1982. – № 2. – С. 1–5.
 44. Маршакова И.В. Система связей между документами, построенная на основе ссылок: (По данным SCI) / И.В. Маршакова // *НТИ. Сер. 2. Информ. процессы и системы*. – 1973. – № 6. – С. 3–8.

45. *Маршакова, И.В.* Система цитирования научной литературы как средство слежения за развитием науки / И.В. Маршакова. – М.: Наука, 1988. – 288 с.
46. *Маршакова-Шайкевич И.В.* Библиометрия как исследовательская техника эпистемологии и философии науки / И.В. Маршакова-Шайкевич // Междунар. форум по информ. и документации. – 1993. – Т. 18, № 3–4. – С. 3–8.
47. Методы сбора данных // Социология. Основы общей теории / Под ред. Г.В. Осипова, Л.Н. Москвичева. – М., 1996. – С. 45–56.
48. *Михайлов, А.И.* Научные коммуникации и информатика / А.И. Михайлов, А.И. Черный, Р.С. Гиляревский. – М.: Наука, 1976. – 435 с.
49. *Михайлов, О.В.* Индекс цитирования ученого: важнейший ли это критерий качества его научной деятельности? / О.В. Михайлов // Наукосведение. – 2001. – № 1. – С. 201–207.
50. *Моралес, М.* Информетрия и ее значение / М. Моралес // Междунар. форум по информ. и документации. – 1985. – Т. 10, № 2. – С. 16–21.
51. *Налимов, В.В.* Наукометрия: Изучение развития науки как информационного процесса / В.В. Налимов, З.М. Мульченко. – М.: Наука, 1969. – 192 с.
52. *Павловска, Е.Ю.* Информационные методы оценки тенденций развития научных направлений / Е.Ю. Павловска. – М.: ВИНТИ, 1990. – С. 3–120. – (Итоги науки и техники. Сер. Информатика; Т. 17).
53. *Пэнто, Р.* Методы социальных наук / Р. Пэнто, М. Гравитц. – М.: Прогресс, 1972. – 607 с.
54. *Рожков, С.А.* Библиометрические методы выявления и анализа научных направлений / С.А. Рожков. – М.: ВИНТИ, 1991. – С. 3–137. – (Итоги науки и техники. Сер. Информатика; Т. 16).
55. *Солтон, Дж.* Динамические библиотечно-информационные системы / Дж. Солтон, Пер. с англ. В.Р. Хисамутдинова. – М.: Мир, 1979. – 557 с.
56. *Субботин, М.М.* О логико-смысловом моделировании содержания управленческих решений / М.М. Субботин // Науч. упр. о-вом. – 1980. – Вып. 13. – С. 203–224.
57. *Уемов, А.И.* Системный подход и общая теория систем / А.И. Уемов. – М.: Мысль, 1978. – 272 с.
58. *Шторх, А.* Систематическое обозрение литературы в России (1801–1806 гг.) / А. Шторх, Ф. Аделунг. – СПб., 1810. – Ч.1. – 352 с.
59. *Яблонский, А.И.* Модели и методы исследования науки / А.И. Яблонский. – М.: Эдиториал УРСС, 2001. – 400 с.

Материал поступил в редакцию 29.04.2005 г.

Сведения об авторе: *Редькина Наталья Степановна* – кандидат педагогических наук, заведующая технологическим отделом, тел. (383) 266-15-36, e-mail: to@spsl.nsc.su

**ГПНТБ СО РАН
предлагает вниманию библиотечарей новое издание**

**Редькина Н.С., Калюжная Т.А.
«ПУТЬ ЭЛЕКТРОННЫХ РЕСУРСОВ В БИБЛИОТЕКЕ»**

В издании отражена технология работы с электронными ресурсами в ГПНТБ СО РАН: порядок комплектования, регистрация, учет, исключение из фондов, каталогизация, организация справочно-поискового аппарата, хранение, сохранность, предоставление пользователям электронных ресурсов и администрирование их пути.

К изданию прилагается перечень и образцы учетно-регистрационной, учетно-статистической и сопроводительной документации, схемы, показывающие в графическом виде последовательность технологических процессов и операций по пути электронных ресурсов, а также методические и технологические инструкции, памятки, конкретизирующие и детализирующие технологию работы с электронными ресурсами на определенном участке.

Издание предназначено для библиотечных работников и является практическим руководством по организации работы с электронными ресурсами в библиотеке.

Издание предназначено для библиотечных работников.

Заявки принимаются по адресу: 630200, г. Новосибирск, ул. Восход, 15, ГПНТБ СО РАН, РИО.

E-mail: rio@spsl.nsc.ru. **Тел.** (383) 266-21-33.

Факс (383) 266-25-85, 266-33-65 (с пометкой «для РИО»)