

НЕТРАДИЦИОННЫЙ АЛГОРИТМ ИЗУЧЕНИЯ СОСТОЯНИЯ ВОПРОСА В ПИЩЕВОЙ ОТРАСЛИ

Предложен нетрадиционный для технических наук методический подход к изучению научно-технической литературы с использованием контент-анализа и библиометрии. Сформулирован парадокс ключевых слов. Даны примеры пользования авторским алгоритмом предварительного библиографического поиска.

Анализ, литература, обзор, контент-анализ, библиометрия.

Большинство специалистов отрасли сталкиваются в массовом порядке с однотипными задачами, возникающими на начальном этапе анализа так называемого «состояния вопроса в изучаемой области», в то время как известно, что качественно выполненный поиск информации всегда сопряжен с огромными затратами времени, интеллектуальных и вычислительных ресурсов.

Вместе с тем даже предварительный частотный анализ ключевых слов в изучаемой области весьма информативен. Однако при использовании мощных библиотечных серверов для поиска информации мы столкнулись с **парадоксом ключевых слов**: чтобы грамотно сформулировать запрос, т.е. не получать ни невозможное для конструктивного анализа число ссылок, ни отсутствие ссылок по предлагаемым запросам, необходимо в деталях представлять себе структуру проблемной области. Но ведь именно детальное представление и является целью поиска.

Выход из создавшегося положения, общего для всех исследователей, особенно работающих в новейших областях знаний, мы нашли, разработав **алгоритм предварительного библиографического поиска** (рис. 1), состоящий из трех крупных этапов: 1) контент-анализ; 2) библиометрический анализ; 3) содержательный анализ. Первые два этапа являются нетрадиционными для анализа литературы в пищевой отрасли, по крайней мере автор за свою многолетнюю практику ни разу не сталкивался с подобным подходом к изучению состояния вопроса. Поэтому дадим ниже некоторые пояснения.

Контент-анализ – количественный анализ текстов с целью последующей содержательной интерпретации выявленных числовых закономерностей. Основная идея контент-анализа проста и интуитивно понятна. При восприятии текста и особенно больших текстовых потоков мы достаточно хорошо ощущаем, что разные содержательные компоненты представлены в них в разной степени, причем эта степень поддается измерению: ее мерой служит то место, которое они занимают в общем объеме, и/или частота их встречаемости. Замысел контент-анализа заключается в том, чтобы систематизировать эти интуитивные ощущения, сделать их наглядными и проверяемыми и разработать методику целенаправленного сбора тех текстовых свидетельств, на которых эти ощущения основываются. При этом предполагается, что

вооруженный такой методикой исследователь сможет не просто упорядочить свои ощущения и сделать свои выводы более обоснованными, но даже узнать из текста больше, чем хотел сказать его автор, ибо, скажем, настойчивое повторение в тексте каких-то тем может не осознаваться автором, но обнаруживается и определенным образом интерпретируется исследователем. Главной отличительной чертой контент-анализа является его квантитативный характер. Основа контент-анализа – это подсчет встречаемости некоторых компонентов в анализируемом информационном массиве. Отметим, что контент-анализ не отменяет необходимости обычного (т.е. содержательного) анализа документов.



Рис. 1. Алгоритм предварительного библиографического поиска

Уже разработано многочисленное программное обеспечение для контент-анализа текстов, этот вопрос не входит в круг интересов автора, поэтому на нем останавливаться не будем. Констатируем лишь, что методике контент-анализа несколько

десятков лет и к настоящему моменту известно большое количество таких программ, как коммерческих, так и демо-версий. Вернемся к алгоритму поиска. Обработав с помощью любой из доступных программ контент-анализа массив текста по теме, мы получаем ранжированный список ключевых слов, иными словами, проведен частотный анализ текста. Чем больший массив будет обработан, тем достовернее результаты контент-анализа.

В нашем случае через подобную программу был пропущен текст монографии О.Н. Мусиной, посвященной поликомпонентным продуктам [1]. Предпосылкой послужили три причины: 1) соответствие теме изучаемого вопроса; 2) существенный объем текста – несколько сотен страниц; 3) наличие электронной версии документа. В общем случае это могут быть любые научные публикации по изучаемой теме, доступные в электронном виде.

Для абсолютно всех слов монографии, включая союзы, глаголы и т.д., программой была определена частота встречаемости в тексте. Допустим, нас интересует вопрос: каким именно поликомпонентным молочным продуктам чаще всего посвящены научные публикации? Из полученного частотного ряда отбрасываем служебные слова, оставляя термины, относящиеся к интересующему вопросу. Далее объединяем лексико-семантические гнезда, т.е. один термин, находящийся в разных словоформах, отбрасываем слова, встречающиеся менее трех раз, в итоге получаем проранжированную в виде частотного ряда таблицу ключевых слов анализируемого текста (табл. 1).

Таблица 1

Контент-анализ по ключевым словам

Термин	Частота	Термин	Частота
сыр	448	сливки	25
молочное	409	десерт	16
молоко	229	казеин	15
творог	201	коттедж	9
кисломолочные	129	молочнокислые	9
плавленный	116	молочно-растительный	8
напиток	100	сырники	8
масло	86	казеинат	5
сыворотка	63	кефир	5
мороженое	46	крем	4
молочно-белковый	38	курт	4
пастообразный	34	пахта	4
йогурт	27	пудинг	4
сквашенное	27		

Теперь можно приступать ко второму этапу поиска – библиографическому анализу.

Данный подход позволяет не только получить представление об изучаемом вопросе, но и количественно оценить степень изученности различных аспектов темы, определить перечень периодических и отчасти непериодических изданий,

в которых обсуждаются эти аспекты, виднейших ученых отрасли и ведущие учреждения, работающие над изучаемой темой. Главнейшим итогом является научно обоснованное выявление оптимальных с точки зрения библиографии направлений собственных исследований и прикладных разработок.

Для облегчения поиска необходимой библиографической информации по «пищевой» тематике рекомендуем использовать мощные библиотечные серверы Российской государственной библиотеки (<http://www.rsl.ru/>) и Центральной научной сельскохозяйственной библиотеки Россельхозакадемии (<http://www.cnshb.ru/>) либо другие федеральные и локальные библиографические электронные ресурсы на усмотрение исследователя. Автор лично наиболее часто при проведении поисков научно-технической информации обращается именно к этим двум вышеупомянутым библиотекам.

Рекомендация обусловлена следующими причинами. Согласно Федеральному закону Российской Федерации «Об обязательном экземпляре документов» от 23 ноября 1994 года с изменениями и дополнениями от 23 января 2002 года Российская книжная палата направляет в Российскую государственную библиотеку все виды отечественных печатных изданий, также туда поступают все авторефераты и диссертации. Это крупнейшая федеральная библиотека России.

Центральная научная сельскохозяйственная библиотека – одна из крупнейших сельскохозяйственных библиотек мира, выполняющая функции отраслевой национальной библиотеки России по сельскому хозяйству и продовольствию, библиотека-депозитарий ФАО (Международная организация по сельскому хозяйству и продовольствию) ООН – Food and Agriculture Organization of the United Nations (FAO), центр AGRIS/CARIS России. CARIS (the Current Agricultural Research Information System) – информационная система по текущим научным исследованиям в области сельского хозяйства, созданная ФАО для обмена информацией о текущих научных проектах в этой области. Система охватывает научные исследования по всем аспектам сельского хозяйства, в том числе пищевую промышленность, питание человека. AGRIS (Agricultural Research Information System) – Международная информационная система по сельскому хозяйству и смежным с ним отраслям – создана в 1974 г. ФАО с целью координации усилий по сбору, обработке и распространению информации по сельскому хозяйству и продовольственной проблеме в странах мира. Система является как библиографической службой, так и международной сетью, объединяющей 240 национальных центров и международных центров обработки и ввода документов в базу данных AGRIS. Национальные центры отбирают и обрабатывают опубликованные в их странах документы, что обеспечивает максимально полное представление последних мировому сообществу

через AGRIS. Реферативная база данных содержит информацию по технологии производства продуктов питания, питанию человека и т.д., включает информацию о книгах, монографиях, статьях из периодических и продолжающихся изданий и сборников. Особое место занимают научно-технические проекты, отчеты, диссертации, материалы конференций, не публикуемые в широкой печати. Такие материалы составляют около 20 % всей информации в базе данных.

На данном этапе применяют наукометрический/библиометрический подход [2, 3], зародившийся как инструмент анализа науки в начале 60-х годов XX века. Библиометрия представляет собой количественные исследования, направленные на выявление долгосрочных тенденций, связанных со стратегическим слежением (мониторингом) развития науки. Достоинством является то, что при библиометрическом подходе могут быть использованы легкодоступные огромные массивы вторичной информации, представленные в различных базах данных и прежде всего в базах данных библиотек.

В нашем частном случае этот подход состоит в определении количества научных публикаций по теме «Поликомпонентные молочные продукты» посредством поиска по ключевым словам. Проводили поиск по ключевым словам, полученным как выходной параметр Этапа I. Использованы ограничения выборок по времени (15-летний срок), но без ограничений типа публикаций и языка оригинала. В полном объеме библиографический анализ состояния вопроса может служить самостоятельным полноценным исследованием (известны диссертации с подобными темами), имеет множество нюансов и требует специфических навыков, поэтому полученные нами по упрощенному варианту методики результаты являются оценочными и представляют интерес в сравнительном аспекте.

Таблица 2

Распределение публикаций по ключевым словам

Термин	Частота	Термин	Частота
молочное	67 026	молочно-растительный	250
молоко	33 833	молочно-белковый	212
сыр	10 346	кефир	212
сыворожка	9605	пахта	164
мороженое	4387	казеинат	156
кисломолочные	3190	крем	103
казеин	1314	десерт	81
масло	1060	пастообразный	40
молочнокислые	1043	сквашенное	32
творог	745	пудинг	22
плавленный	649	сырники	7
йогурт	642	коттедж	6
напиток	500	курт	6
сливки	490		

Итак, в табл. 2 показано, как за период 1995–2010 гг. распределены публикации по ключевым словам. В структуре библиографии прежде всего вызывает интерес то, что количество публикаций по сыру в целом сопоставимо с публикациями по сыворотке, а отношение частоты публикаций по напиткам (500) лишь в два раза выше частоты публикаций по кефиру (212), в то же время ключевое слово «кисломолочный» в публикациях встречается более трех тысяч раз, также достаточно часто встречается термин «молочно-растительный», даже чаще, чем «молочно-белковый», количество публикаций по мороженому практически в четыре раза больше, чем по маслу, а термин «йогурт» фигурирует чаще, чем «напиток» и «кефир», вместе взятые.

В рамках библиографического анализа нам показалось интересным и конструктивным (с точки зрения дальнейших перспектив исследования) сопоставить количество отдельных видов молочных продуктов в общей их структуре. При анализе таблиц мы ввели коэффициент K – отношение числа публикаций, посвященных ключевому слову, относящемуся к отдельному виду молочных продуктов, к общему числу публикаций, умноженное на 100 %. В порядке убывания коэффициент K составляет: сыр – 0,31, сыворотка – 0,28, мороженое – 0,12. Для казеина, масла, творога, йогурта, напитков, сливок, кефира K колеблется от 0,01 до 0,03, для пахты, казеина, молочных кремов и десертов $K = 0,002...0,005$. Мала относительная доля публикаций, упоминающих пудинги, сырники, сыр коттедж, курт, – для них K составляет тысячные доли единицы.

Безусловно, данный подход к оценке актуальности изучаемых проблем зависит от временной динамики прироста публикаций, выступающей как аналог производной в дифференциальных уравнениях. Особенно важен прирост публикаций в последние годы, отражающий концентрацию интереса исследователей на каком-либо аспекте темы. Библиотечные серверы при поиске по ключевым словам выдают не просто цифру количества публикаций, но и распределение их по авторам, учреждениям, виду публикации, а также другие сведения библиографического характера и иногда рефераты публикаций. Указанная информация позволяет выявить для исследуемой темы крупнейших ученых, ведущие учреждения и важнейшие периодические издания. Эти задачи входят в комплекс библиографических методов анализа литературы, их, безусловно, необходимо решать, что отражено в авторском Алгоритме (см. рис. 1), но ввиду ограниченности объема настоящей работы и принципиальной простоты выполнения пояснять конкретными примерами здесь не будем.

Зная динамику и степень изученности различных аспектов темы, имена авторитетных ученых, учреждения, работающие по актуальной тематике, а также ключевые периодические издания, в которых публикуются результаты таких исследований, можно приступить к Этапу III Алгоритма – содержательному анализу выборки публикаций,

авторы которых, учреждения и наименования изданий уже известны. Таким образом, пара нетрадиционных для пищевой отрасли предварительных этапов при сборе информации о состоянии изучаемого вопроса существенно облегчает и рационализирует выполнение аналитического обзора и формулировку направлений собственных исследований.

Список литературы

1. Мусина, О.Н. Состояние и тенденции развития биотехнологии комбинированных молочных продуктов / О.Н. Мусина, М.П. Щетинин, М.Н. Сахрынин. – Барнаул: Изд-во АлтГТУ, 2006. – 335 с.
2. Редькина, Н.С. Формализованные методы анализа документальных информационных потоков / Н.С. Редькина // Библиосфера. – 2005. – № 2. – С. 51–59.
3. Хайтун, С.Д. Проблемы количественного анализа науки / С.Д. Хайтун. – М.: Наука, 1989. – 280 с.

ГНУ Сибирский НИИ сыроделия
Сибирского отделения Россельхозакадемии,
656016, Россия, г. Барнаул, ул. Советской
армии, 66.
Тел./факс: (3852) 37-56-83

SUMMARY

O.N. Musina

Nonconventional algorithm of studying the condition of a problem in food branch

The nonconventional for technical sciences methodical approach to studying the scientific and technical literature by using the content-analysis and bibliometry has been offered. The paradox of keywords has been formulated. The examples of using the author's algorithm of preliminary bibliographic search are given.

Analysis, literature, review, content analysis, bibliometry.

