

УДК 025.347

ББК 78.527в648

DOI 10.25281/2411-2305-2024-6-45-54

С.С. Захарова

Новые подходы к аппарату научного библиографического указателя (содержание журнала «Почвоведение» за 1999–2023 гг.)

Реферат. Библиографический справочник «Материалы, опубликованные в журнале «Почвоведение» за 100 лет» (1999) представляет собой уникальный информационный ресурс. В настоящее время дирекцией Института физико-химических и биологических проблем почвоведения РАН принято решение о подготовке продолжения (за период с 1999 по 2023 г.). Одной из основных проблем является разработка предметных рубрик для наиболее оптимального отражения материалов журнала в этом справочнике, поскольку актуальные вопросы развития как самого почвоведения, так и смежных отраслей знания очень сильно изменились за прошедшие десятилетия (особенно в ряде комплексных проблем). Автор подчеркивает также необходимость кардинально переработать основные рубрики предметно-тематического указателя. В статье рассматриваются возможности применения для этой цели формализованных методов оценки документных потоков.

Ключевые слова: журнал «Почвоведение», библиографический справочник, справочный аппарат, контент-анализ, тематический анализ, интеллектуальный анализ текста, предметно-тематический указатель.

Для цитирования: Захарова С.С. Новые подходы к аппарату научного библиографического указателя (содержание журнала «Почвоведение» за 1999–2023 гг.) // Библиография и книговедение. 2024. № 6. С. 45–54. DOI: 10.25281/2411-2305-2024-6-45-54.

Интеллектуальный анализ данных становится важным направлением информационно-библиографического обслуживания в современных библиотеках научно-исследовательских институтов и университетов. При проведении экспертизы публикаций и деятельности конкретного ученого необходим тща-



**Светлана Сергеевна
Захарова**

Библиотека по естественным
наукам (БЕН) РАН,
старший научный сотрудник
ул. Знаменка, д. 11/11,
Москва, 119991, Россия
ORCID 0000-0002-6351-6405;
SPIN 8914-6161
bibl_impb@rambler.ru



Рис. 1. Обложка библиографического справочника «Материалы, опубликованные в журнале «Почвоведение» за 100 лет» (1999) [4]

тельный анализ закономерностей развития выбранного им научного направления, при этом возрастает роль методов количественного изучения документальных информационных потоков.

Сотрудники библиотек широко и успешно применяют библиографический метод, позволяющий не только составить описание документов и представить их содержание в аннотации, но и сгруппировать по различным признакам для характеристики потока или массива документов (а также оформить в виде библиографического ресурса) [1–3]. В частности, спра-

вочник «Материалы, опубликованные в журнале “Почвоведение” за 100 лет» [4] представляет собой своего рода путеводитель по истории науки почвоведения (не только в России). С его помощью удалось ввести в научный оборот огромный и ценный массив информации (во многом ставшей сейчас малодоступной). Этот справочник — отличный пример библиографического указателя.

Издание (рис. 1) содержит перечень всех публикаций (13 664) из 834 номеров журнала «Почвоведение» (за периоды, когда он выходил: 1899–1916 и 1924–1998 гг.). Даны библиографические описания оригинальных научных статей, рецензий на книги, хроникальных заметок о научной жизни. Все записи расположены по годам и выпускам журнала, приводятся последовательно, в порядке единой сквозной нумерации.

Аппарат издания объединяет несколько вспомогательных указателей: авторский — список всех авторов публикаций в алфавитном порядке, включая соавторов (всего 7120 фамилий), указатель персоналий — 450 фамилий ученых с указанием публикаций о них (биографии, юбилеи, анализ творчества, некрологи), а также предметно-тематический, исчерпывающе раскрывающий содержание журнала.

Колоссальная и кропотливая работа, выполненная составителями этого справочника (д-ром геогр. наук И.В. Ивановым и канд. биол. наук Т.С. Луковской), нашла наиболее яркое отражение именно в предметно-тематическом указателе, включающем около 1000 рубрик. Распределение выполнено на основании названий публикаций; внутри крупных рубрик вы-

деляются подрубрики с соблюдением логической последовательности от общего к частному. Родственные или близкие по содержанию понятия связываются перекрестными ссылками. Этот указатель напоминает классический предметный каталог (так тщательно составители раскрывают суть статей). При тематическом поиске легко убедиться, насколько грамотно работают ссылки и отсылки, подсказывая связи и разграничения между предметными рубриками. Скорее всего, в правильном сочетании этих моментов и заключается длительная востребованность «Материалов, опубликованных в журнале “Почвоведение” за 100 лет» [5].

В 2023 г. дирекцией Института физико-химических и биологических проблем почвоведения (ИФХиБПП) РАН было принято решение о подготовке продолжения этого справочника (за период с 1999 по 2023 г.).

В резолюции VIII съезда Общества почвоведов им. В.В. Докучаева (проходил 10–14 августа 2022 г. на базе Института биологии Коми научного центра Уральского отделения РАН и Сыктывкарского государственного университета им. Питирима Сорокина) предлагается способствовать организации работ по подготовке и изданию толковых словарей и энциклопедии по почвоведению, создав рабочую группу для этих целей. Выполняя рекомендации, директор ИФХиБПП РАН член-корреспондент РАН А.О. Алексеев издал приказ от 16.02.2023 о составлении второго тома справочника «Материалы, опубликованные в журнале “Почвоведение” за 25 лет (1999–2023 гг.)».

Один из старейших академических журналов России (рис. 2)

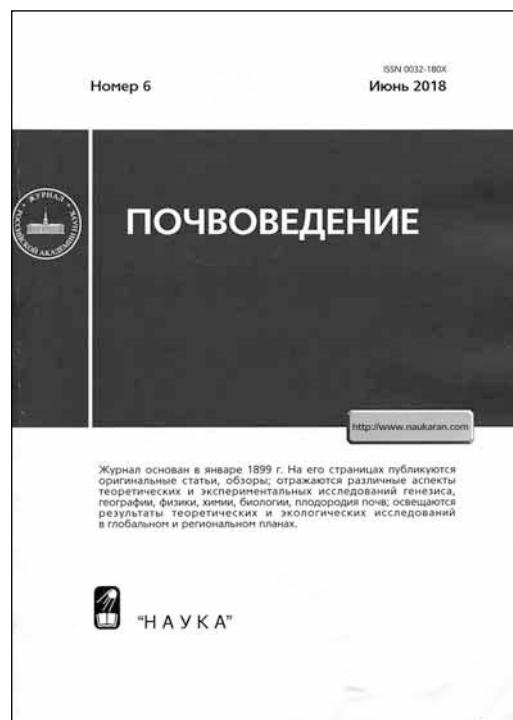


Рис. 2. Журнал «Почвоведение»

не только продолжает публикацию оригинальных статей и обзоров, в которых отражаются различные аспекты теоретических и экспериментальных исследований генетики, географии, физики, химии, биологии, плодородия почв; освещаются результаты теоретических и экологических исследований в глобальном и региональном планах.

Журнал основан в январе 1899 г. На его страницах публикуются оригинальные статьи, рецензии, обзоры, сообщения о конференциях, методические материалы, теоретические и экспериментальные исследования генетики, географии, физики, химии, биологии, плодородия почв; освещаются результаты теоретических и экологических исследований в глобальном и региональном планах.

“НАУКА”



Рис. 3. Страница сайта журнала «Почвоведение»

Citation Index, WOS и Scopus (переводные версии). Главный редактор член-корреспондент РАН П.В. Красильников и редколлегия очень серьезно относятся к подготовке материалов.

В настоящее время разработка предметных рубрик для многоаспектного отражения материалов журнала «Почвоведение» (1999–2023) является важной проблемой, так как существенно изменились магистральные направления развития данной науки и смежных отраслей знания, появились новые разделы почвоведения и активно формируются комплексные проблемы исследований (в том чис-

ле в Российской Федерации). Все это вынуждает кардинально перерабатывать рубрики предметно-тематического указателя для составления второго тома справочника.

В данной статье рассматриваются возможности формализованных методов оценки публикационных потоков для разработки предметных рубрик, оптимально раскрывающих содержание журнала «Почвоведение» и соответствующих современным правилам информационно-библиотечного сопровождения научных исследований.

ГОСТ 7.78–99. «Издания. Вспомогательные указатели» вспомога-

тельный указатель определяет следующим образом: «Упорядоченный (по алфавиту или какому-либо другому признаку) перечень личных имен, предметов, географических названий, событий, публикаций, цитат, аббревиатур, символов, формул и других информационных объектов с указанием их местонахождения на страницах издания». Основными видами таких указателей являются предметный (его разновидности — тематический, именной, географический) и хронологический [6, с. 1–2]. Все они выполняют конкретные функции, но предметно-тематический указатель наиболее удачно воспроизводит структуру истории и современного состояния отрасли знания (в данном случае почвоведения). Перед составителями ставится задача создать путеводитель по огромному материалу, хотя и ограниченному хронологическими рамками. Методика составления хорошо известна [7; 8], но специфика развития библиотечных и информационных технологий, естественно, требует определенных корректировок.

Обновленная схема предметно-тематического указателя нового справочника за 1999–2023 гг. должна показать изменения в структуре почвоведения. Помогают отразить специфику темы соблюдение логической последовательности разделов и подразделов, четкое их разграничение и выработка системы ссылок, соединяющих смежные рубрики. Все это необходимо тщательно проверить и осмыслить уже на новом уровне.

Поэтому сформирована комиссия ведущих ученых в области почвоведения, однако пока проблема находится в стадии разработки. В этой связи по-

пытаемся разобраться, есть ли возможность составить предметно-тематический указатель с применением методов анализа текста. В первую очередь представляют интерес методы формализованного анализа документов и использование их для получения информации из имеющихся данных [9; 10].

Контент-анализ, пожалуй, самый очевидный вариант. Один из его основоположников (Бернард Берельсон) определяет метод как «объективное, систематическое и количественное описание очевидного содержания коммуникаций» [11, с. 18]. Публикаций по применению контент-анализа в разных областях знаний достаточно, он успешно используется на протяжении продолжительного времени [12; 13]. Какие возможности этого метода применимы в сфере библиографических ресурсов?

Контент-анализ исследует массивы данных, переводя текстовую информацию в показатели с их последующим подсчетом (различают его качественный и количественный варианты). Количественный метод определяет частоту появления в тексте ряда параметров содержания, в то время как качественный позволяет делать обобщения на основе единичного присутствия или отсутствия в тексте заявленных элементов содержания. Иными словами, применение контент-анализа сводится к переводу текста в определенные символы, затем их сводят воедино и производят подсчеты, обращая внимание на частоту употребления [14–16]. Методы лексического анализа научного текста являются вариантом контент-анализа и широко используются для изучения содержания публикаций, в качестве основного элемента

выбирая словарное слово («сленговый метод») или словосочетание, предложение («логико-смыслоное моделирование»). Таким образом, контент-анализ трансформируется в зависимости от решаемых задач.

Для разработки предметно-тематического указателя использование специальной схемы кодирования и категоризации, позволяющей выявить ключевые понятия и проследить взаимосвязи внутри текстов определенной тематики, является вполне перспективным направлением.

Тематический анализ (категоризация и кодирование данных для выявления повторяющихся тем, концепций) показывает основные закономерности и темы текстового документа для углубленного изучения данных. Универсальность этого метода может быть полезна и для вспомогательных указателей, так как позволяет разрабатывать коды и темы на основе данных в процессе логического вывода на основе перехода от частного к общему. Переход от кодов к темам не является простым механическим действием. В процессе анализа данных можно выявить новые подразделы, поэтому необходимо указывать, как именно заявленные коды были классифицированы.

Тематический анализ и контент-анализ имеют разные цели, но они оба направлены на извлечение информации из документального массива данных. Цель контент-анализа — анализ распределения и частоты конкретных элементов или кодов в данных, которые классифицируются и подвергаются систематической количественной оценке. Цель тематического анализа — выявление и последующий анализ по-

вторяющихся тем и закономерностей в данных. Если в контент-анализе используются заранее определенные категории и схемы для количественной оценки и анализа данных путем систематического кодирования и категоризации, то в тематическом поиске кодирование и анализ проводятся интерактивно, что позволяет изучить точки зрения участников. Анализ содержания и взаимосвязей между переменными — ключевые составляющие контент-анализа, который часто проводится на больших данных, а тематический анализ ограничивается изучением конкретной проблемы. Таким образом, контент-анализ особое внимание уделяет количественным данным, а тематический анализ — субъективности, которая часто встречается в теоретических исследованиях и концепциях и относится к отличной от других интерпретации исследования [17; 18].

Все же и контент-анализ, и тематический анализ не могут решить все задачи, важные для разработки предметно-тематического указателя. Видимо, целесообразно их использовать комплексно, поскольку для анализа данных они требуют систематического и структурированного подхода. В частности, следует создать четкую схему кодирования или категоризации. Также категории анализа должны обеспечивать полноту характеристик основных понятий, исключать возможность повтора одного содержания в различных категориях в том же объеме, учитывать разногласия специалистов по наполнению рубрик.

Возможные ошибки программного обеспечения напрямую связаны с недочетами классификатора, составлен-

ного с нарушением логики или по недостаточно отработанной инструкции (особенно, если инженер-программист сам как следует не апробировал схему). Таким образом, в идеале библиографу рекомендуется совместно с ИТ-специалистом тщательно проанализировать содержание каждой статьи, вырабатывая нужные классификационные стратегии. Нельзя оставить без внимания и перспективы развития новых методов: так, М.Ю. Нещерет отмечает, что «цифровизация библиографической деятельности выражается в появлении новых методов, предполагающих использование инновационных программных инструментов в качестве альтернативы традиционному библиографическому подходу» [19, с. 44].

В этой связи обращает на себя внимание *интеллектуальный анализ текста* (ИАТ, англ. text mining), который подразумевает применение автоматических методов для анализа и структурирования текстовых данных с целью создания полезного знания из ранее неструктурированной информации. В процессе ИАТ используются такие технологии, как извлечение информации, ее обобщение, категоризация, кластеризация и визуализация. Они повторяют методологию формализованных методов оценки текста, но используют более точные системы извлечения информации, учитывая значительный прогресс в этой области за последнее время [20].

Именно этот метод дает возможность разработать схему классификации, а затем отдать ее на редактирование группе авторитетных ученых (для экспертизы предложенного макета). Например, процесс категоризации текста состоит из предварительной об-

работки (preprocessing), индексирования, сокращения или уменьшения размерности и классификации. Целью категоризации является подготовка классификатора на основе известных примеров (неизвестные затем классифицируются автоматически). Метод кластеризации помогает находить группы документов с аналогичным содержанием. Если в пределах одного кластера цель достигнута, а между различными кластерами есть отличия, то качество кластеризации считается наивысшим.

Технология кластеризации, используемая для группировки похожих документов, отличается от категоризации, гарантируя, что нужный результат не будет исключен в ходе поиска (документы могут отражаться в нескольких подразделах). В интеллектуальном анализе текста графическая визуализация улучшает поиск информации, выстраивая текстовые источники в визуальной иерархии, позволяя взаимодействовать с документом путем увеличения и масштабирования [21–23].

В статьях С.С. Хачатуровой представлено инновационное программное обеспечение «Анализатор текста» для интеллектуального анализа текстовых документов, использующее методы частотного анализа для определения степени схожести смысловой близости документов по определенной теме. Выполнение интеллектуального анализа текста расписано по шагам «в процедуре подсчета значимых ключевых, однокоренных и других информативных характеристик анализируемого текста» [24; 25].

В публикации Е.А. Немчиновой и ее соавторов показаны возможно-

сти применения нейронных сетей для классификации. Пример разработки информационной базы, содержащей словарь анализируемой предметной области, был очень полезен для работы по составлению библиографического справочника по почвоведению [26]. В статье Л.В. Зиминой последовательно охарактеризованы два направления использования искусственного интеллекта (ИИ) в книжном деле. Одно из них (генерация текста) перспективно для дальнейшей работы с указателем (если учитывать важнейшее свойство нейронных сетей — самообучение) [27].

Таким образом, комплекс методов количественных оценок и технологии ИИ, целью которых является извлечение наиболее значимой информации из текстовых документов, вполне применим для составления вспомогательных указателей к библиографическим ресурсам.

Подчеркнем, что предметно-тематический указатель обеспечивает четкую навигацию для более полного раскрытия разнообразия разрабатываемых исследователями проблем. На подготовительном этапе составления предметно-тематического указателя (с использованием формализованных методов оценки публикационных потоков) технологии создания библиографических ресурсов и задачи программного обеспечение похожи. Для распределения данных с применением методов тематического анализа и контент-анализа программист должен сначала создать четкую схему кодирования или категоризации, а основная задача библиографов — составление схемы классификации, позволяющей раскрыть тематику, специ-

фику и комплексность проблематики исследований по почвоведению за указанные годы.

Список источников

1. ГОСТ Р 7.0.76—2022. Система стандартов по информации, библиотечному и издательскому делу. Библиографирование. Библиографические ресурсы. Термины и определения. Москва : Рос. ин-т стандартизации, 2022. 24 с.
2. Калинина Г.П. Изменение требований к государственным библиографическим указателям : новый стандарт СИБИД ГОСТ Р 7.0.61—2022 // Библиография и книговедение. 2023. № 1. С. 39–53. DOI: 10.25281/2411-2305-2023-1-39-53.
3. Калинина Г.П. Правила формирования заголовка библиографической записи в новом национальном стандарте // Библиография и книговедение. 2023. № 3. С. 30–61. DOI: 10.25281/2411-2305-2023-3-30-61.
4. Материалы, опубликованные в журнале «Почвоведение» за 100 лет : библиогр. справ. / Рос. акад. наук, Докучаев. о-во почвоведов, Науч. совет по проблемам почвоведения, Ин-т фундам. проблем биологии ; отв. ред. Г.В. Добропольский ; сост.: И.В. Иванов, Т.С. Луковская. Москва : Наука, 1999. 670 с.
5. Иванов И.В., Луковская Т.С. Норметрический анализ публикаций журнала «Почвоведение» за 100 лет (1899–1998 гг.) // Почвоведение. 2003. № 1. С. 113–125.
6. ГОСТ 7.78—99. Система стандартов по информации, библиотечному и издательскому делу. Издания. Вспомогательные указатели. Москва : ИПК «Изд-во стандартов», 2000. 17 с.
7. Призмент Э.Л., Динерштейн Е.А. Вспомогательные указатели к книжным изданиям. 2-е изд., испр. и доп. Москва : Книга, 1988. 207 с. (От рукописи — к книге).

8. Кунце Х. Составление вспомогательных указателей / пер. с нем., предисл. и примеч. Э.Л. Призмента. Москва : Книга, 1977. 62 с.
9. Редькина Н.С. Формализованные методы анализа документальных информационных потоков // Библиосфера. 2005. № 2. С. 51–59.
10. Митина О.В., Евдокименко А.С. Формализованные методы исследования текстов: опыт применения к анализу технической документации // Вестник Томского гос. ун-та. Филология. 2010. № 1. С. 60–69.
11. Berelson B. Content Analysis in Communications Research. New York : The Free Press, 1952. 220 р.
12. Курдюмова И.А. Контент-анализ как метод изучения архивных документов // Донецкие чтения 2018: Образование, наука, инновации, культура и вызовы современности : материалы III Междунар. науч. конф. Донецк : Донецкий нац. ун-т, 2018. Т. 1. С. 332–333.
13. Семёнова А.В. Контент-анализ в социологии: методологические новации // Вестник Вятского гос. гуманит. ун-та. 2011. № 3 (1). С. 11–15.
14. Пашиян И.А. Контент-анализ как метод исследования: достоинства и ограничения // Научная периодика: проблемы и решения. 2012. № 3. С. 13–18.
15. Везетиу Е.В., Вовк Е.В. Контент-анализ как метод качественно-количественного анализа содержания документов // МедиаВектор. 2022. № 3. С. 9–12.
16. Кузьмин А.М., Высоковская Е.А. Контент-анализ — один из инструментов изучения и анализа содержания источников информации // Методы менеджмента качества. 2022. № 12. С. 51.
17. Завери А. Контент-анализ и тематический анализ: Более пристальный взгляд. URL: <https://mindthegraph.com/blog/ru/content-analysis-vs-thematic-analysis/> (дата обращения: 16.11.2024).
18. Как проводить тематический анализ и когда его использовать. URL: <https://hr-portal.ru/story/kak-provodit-tematicheskiy-analiz-i-kogda-ego-ispolzovat> (дата обращения: 16.11.2024).
19. Нещерет М.Ю. Цифровая библиография: библиотеки в поисках инновационных инструментов библиографической деятельности // Научные и технические библиотеки. 2021. № 7. С. 33–50. DOI: 10.33186/1027-3689-2021-7-33-50.
20. Федюшкин Н.А., Федосин С.А. Понятие, проблемы и разновидности интеллектуального анализа текста // Проблемы и достижения в науке и технике : сб. науч. тр. по итогам междунар. науч.-практ. конф. Омск, 2016. 206 с.
21. Харченко Б.В., Нестругина Е.С. Искусственный интеллект как инструмент для работы с текстами // Вестник Донецкого нац. ун-та. Серия Г: Технические науки. 2023. № 4. С. 29–40.
22. Воронович С.И., Каракулько Ю.О. Использование тезауруса AGROVOC для анализа тематической структуры поступлений национальных документов в фонд Белорусской сельскохозяйственной библиотеки // Библиотечно-информационный дискурс. 2023. Т. 3, № 2. С. 40–47. DOI: 10.5281/zenodo.10418844.
23. Воскобойников М.Л. Диспетчер научных рабочих процессов: методы и средства визуализации данных // Современные наукоемкие технологии. 2024. № 1. С. 16–21. DOI: 10.17513/snt.39902.
24. Хачатурова С.С. Анализатор текстовых документов // Вестник педагогических наук. 2023. № 1. С. 110–113.
25. Хачатурова С.С. Анализатор текста для подборки электронных документов // Стратегия России: взгляд в завтрашний день : сб. ст. X Междунар. науч. конф. «Абакинские чтения», 26–27 апр. 2021 г. Москва, 2021. С. 187–192.

26. Немчинова Е.А., Плотникова Н.П., Федосин С.А. Подготовка и обработка нормативно-справочной текстовой информации для классификации с помощью искусственных нейронных сетей // Нелинейный мир. 2019. Т. 17. № 2. С. 27–33. DOI: 10.18127/j20700970-201902-03.
27. Зимина Л.В. Нейросети в книгоиздательской индустрии // Библиография и книговедение. 2023. № 2. С. 48–66. DOI: 10.25281/2411-2305-2023-2-48-66.3

Иллюстрации предоставлены автором статьи.

Статья поступила в редакцию 03.05.2024; одобрена после рецензирования 18.05.2024;
принята к публикации 04.12.2024.

FROM PRACTICAL EXPERIENCE

Original article

New Approaches to the Apparatus of the Scientific Bibliographic Index (Contents of the Journal “Soil Science” for 1999–2023)

Svetlana S. Zakharova

Library for Natural Sciences RAS, Moscow, Russia

ORCID 0000-0002-6351-6405; SPIN 8914-6161; bibl_impb@rambler.ru

Abstract. The bibliographic reference book “Materials Published in the Journal “Soil Science” for 100 Years” (1999) is a unique information resource. Currently, the directorate of the Institute of Physico-Chemical and Biological Problems of Soil Science of the Russian Academy of Sciences has decided to prepare a continuation (for the period from 1999 to 2023). One of the main problems is the development of subject headings for the most optimal reflection of the journal’s materials in this handbook, since topical issues of the development of both soil science itself and related branches of knowledge have changed a lot over the past decades (especially in a number of complex problems). The author also emphasizes the need to radically redesign the main headings of the subject and thematic index. The article discusses the possibilities of using formalized assessment methods for evaluation of document flows.

Keywords: journal “Soil Science”, bibliographic reference, reference apparatus, content analysis, thematic analysis, intellectual text analysis, subject and thematic index.

Citation: Zakharova S.S. New Approaches to the Apparatus of the Scientific Bibliographic Index (Contents of the Journal “Soil Science” for 1999–2023), *Bibliografiya i knigovedenie* [Bibliography and Bibliology], 2024, no. 6, pp. 45–54. DOI: 10.25281/2411-2305-2024-6-45-54.

The article was submitted 03.05.2024; approved after reviewing 18.05.2024;
accepted for publication 04.12.2024.