

## **Онтологический подход к созданию словаря по основным разделам алгебры**

*Никольская И.Ю., Голод Е.С., Ефременкова В.М.  
(ВИНИТИ РАН)*

### **Введение**

При исследовании явлений различной природы математика служит основой практически всех областей знания. Отражением бурного развития вычислительной математики (computer science), связанного с разработками в области вычислительных систем и информационных технологий, явилось стремительное расширение специальной терминологии и появление новых терминов. Для того, чтобы следить за литературой, а также вследствие развития научных коммуникаций между математиками разных стран, возникает необходимость разработки специальных англоязычных словарей математических терминов по различным разделам математики. Информационные онтологии служат для формального описания и уточнения понятий, определения общей терминологической базы по основным тематическим разделам алгебры.

Банк данных ВИНИТИ РАН дает возможность отбора терминов и фразеологических единиц по фрагменту БД «Математика» в разделах «Линейная алгебра», «Поля и многочлены» и «Гомологическая алгебра», представляющих наибольшие трудности для глубокого понимания текстов публикаций.

### **Краткая история развития РЖ/БД ВИНИТИ по математике и системы классификации по математическим дисциплинам**

Банк данных ВИНИТИ РАН, включающий 27 тематических фрагментов (специализированных БД), формируется на основе Рубрикатора информационных изданий ВИНИТИ. В области математики в банке данных ВИНИТИ

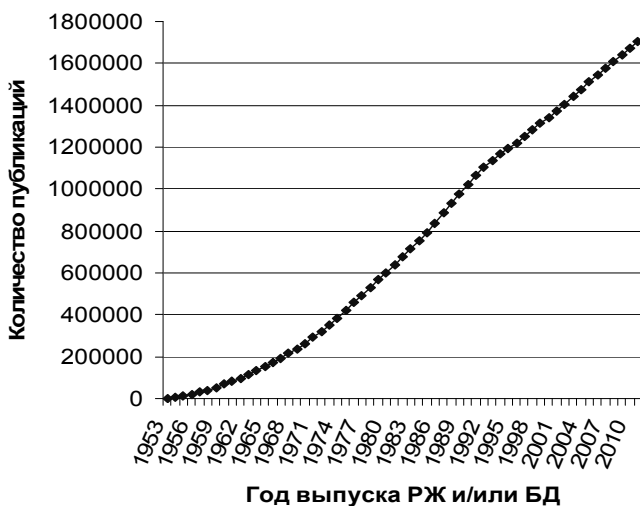


Рис. 1. Рост ретрофонда, отраженного в сводном томе РЖ и/или БД ВИНТИ «Математика» и отдельном выпуске «Вычислительные науки» в период 1953-2012 гг.

представлены 2 раздела: БД «Математика» и БД «Вычислительные науки».

Первый выпуск РЖ «Математика» в ВИНТИ появился в 1953 г. В те годы основными источниками реферативной информации в области математических наук в мире являлись два издания: «Zentralblatt für Mathematik und ihre Grenzgebiete», издаваемый с 1930 г., и «Mathematical Reviews» (MR), издаваемый с 1940 г. Американским математическим обществом (AMS) [с 1969 г. AMS выпускает «Current Mathematical Publications» (CMP), который оповещает о большинстве изданий по математике]. С появлением системы компьютерной верстки TEX — системы электронного представления сложных математических текстов — наряду с печатными реферативными изданиями появились БД: MathSciNet — БД американского математического общества (AMS); Zentralblatt MATH под эгидой Европейского математического общества; а также БД Didactics of Mathematics и Jahrbuch Database. Кроме того, математическая информация содержится в

политематических БД ведущих стран мира: БД INSPEC-C (Великобритания), БД PASCAL (Франция), БД SCI (США), БД SCOPUS и Scirus (Нидерланды).

БД «Математика» предоставляется пользователям ВИНИТИ РАН с 1997 г. [1].

Объем ретрофонда РЖ ВИНИТИ по математике с 1953 г. по 2012 г. насчитывает около 2 млн. публикаций (рис. 1).

Первый Рубрикатор РЖ «Математика» был разработан в 1963 г. Он представлял собой иерархическую классификацию. В 1988 г. введен Рубрикатор информационных изданий ВИНИТИ, разработанный на основе ГРНТИ и совпадающий с ним на первых трех уровнях. Исторически в РЖ ВИНИТИ «Математика» использовался УДК. В настоящее время существует таблица соответствия кодов рубрик Рубрикатора ВИНИТИ и индексов УДК, что позволяет использовать и индексы УДК, и коды рубрик Рубрикатора ВИНИТИ.

На основе УДК и ГРНТИ построены отдельные выпуски сводных томов РЖ ВИНИТИ «Математика» (впоследствии и БД), которые отражают наиболее актуальные проблемы в области фундаментальных математических наук и прикладных областей этой дисциплины.

Распределение суммарных потоков публикаций по математическим дисциплинам определило процесс деления массива сводного тома РЖ «Математика» на отдельные выпуски, в которых отражалась тематика основных разделов в области математики.

Суммарный поток публикаций выпуска БД/РЖ «Общие вопросы математики. Математическая логика. Теория чисел. Алгебра. Топология. Геометрия» (13А/МА01) достаточно стабилен. Анализ терминологии и далее создание предметно-систематических указателей проводился на одном из разделов этого выпуска РЖ/БД «Алгебра».

### ***Система указателей к РЖ ВИНИТИ «Математика»***

Создание и ведение словарей ключевых слов и словосочетаний в определенной предметной области важны для систематизации знания ученого, структурирования и упорядочения документальных массивов, к которым он

обращается в процессе работы. «Словами мы познаем суть вещей» говорил мудрый царь Соломон. Системы ключевых слов быстро отражают те изменения, которые возникают в процессе развития научных направлений. Одним из наиболее эффективных способов упорядочения накопленных терминов является систематическая группировка их по темам и рассматриваемым проблемам. Систематика документов, содержание которых представлено ключевыми словами, распределение их по отраслям знания, проводятся на основе классификации наук. Многозначность представления знания специалистами различных дисциплин отражается в системах классификации, позволяющих посмотреть не только на проблему в целом, но и на все окружение рассматриваемого вопроса.

Для получения перечня упорядоченных по алфавитному, систематическому, нумерационному или какому-либо другому признаку описаний документов, который обеспечивает быстрый поиск сведений в информационном массиве используются **указатели**. По содержанию и назначению различают два больших класса указателей: систематические указатели и предметные.

**Систематический указатель** — это указатель, в котором ключевые слова (дескрипторы) сгруппированы согласно принятой в информационном центре рубрикации. Систематический указатель определяет тематическое содержание словаря определенной области знания, раскрывает и отражает те аспекты рассматриваемой дисциплины, по которым можно с той или иной глубиной детализации проводить поиск. Он дает наглядное представление об общем состоянии терминологии в той или иной области знания, позволяет построить стройную терминологическую модель и учесть, по возможности, все термины и понятия, которые в дальнейшем могут быть использованы при разработке тезауруса. Указатель предназначен для облегчения поиска терминов при составлении поисковых образов документов и запросов путем упорядочения множества ключевых слов по предметному признаку.

**Предметный указатель** — это список ключевых слов, кратко определяющих смысл и тематику определенной области знания.

**Предметизация документов** подразумевает интеллектуальный или автоматический анализ содержательных и формальных признаков документа для его отражения в поисковой системе с помощью языка предметных рубрик.

Для предметных указателей, содержащих сведения о предмете и тематике, всю исследуемую в реферируемых работах тематическую область представляют в виде подобластей, разбиваемых, в свою очередь, на внутренние подобласти и т.д. В этом случае мы имеем иерархическую классификационную систему.

**Предметно-систематический указатель** к информационным массивам — это указатель определенной тематической области, в котором ключевые слова определяют смысл и тематику области знания (или узкотематического ее направления) и при этом сгруппированы в алфавитном порядке по выбранной системе классификации.

Предметно-систематический указатель к РЖ/БД формируется на основе терминов в результате машинной обработки поисковых образов к каждому документу и зависит от массива первоисточников, отражаемого в РЖ/БД. Структура этого типа указателя зависит от отражаемых информационным центром отраслей знания и видов первоисточников.

Предметно-систематический указатель определенной предметной области, в котором устанавливаются соответствия между ключевыми словами, используемыми при координатном индексировании, и разделами рубрик УДК (Рубрикатора БД «Математика»), в которые «попадает» индексируемая работа, оказался наиболее удобным с точки зрения информативности и достоверности. Его содержание определяется, прежде всего, тем потоком литературы, который приходит для дальнейшей обработки (разметки и реферирования) в ВИНТИ. Не менее важна роль специалиста, занимающегося выбором, реферированием и редактированием публикаций для последующего отражения в РЖ/БД.

Таким образом, использована инструментальная реализация и онтологический подход к рассматриваемым разделам алгебры. Иерархически структурированное множество терминов, описывающих предметную область

может быть использовано как исходная структура для базы знаний.

### **Методика формирования предметно-систематических указателей к тематическим разделам РЖ/БД ВИНТИ РАН «Математика»**

В период 2006–2010 гг. на основе массива ключевых слов (КС) разделов БД «Математика» «Линейная алгебра», «Поля и многочлены» и «Гомологическая алгебра» в автоматизированном режиме было получено частотное распределение КС с относящимися к ним кодами УДК и рубриками БД ВИНТИ [2]. Алфавитный список КС, полученный из поисковых образов документов каждого массива насчитывал около 7 тыс. терминов. Анализ этого массива проводился экспертами в несколько этапов. В начале удалялись термины, относящиеся к характеристикам вида документов типа «компьютерный практикум»; общенаучные термины и прочие слова (обзоры, конференции, образование и т.д.); город или страна проведения конференций (Таганрог, Россия и др.); термины, не связанные тематически с рассматриваемым разделом, например, «пакет MATHEMATICA», «физика». Далее КС приводились к нормализованному виду (именительный падеж, единственное число); во множественном числе присутствовали термины, не имеющие единственного числа (числа Фибоначчи, глобальные алгоритмы CG, круги Гершгорина, подобные матрицы и др.).

Кроме того исключалась синонимия, например, из двух синонимов «унитарно инвариантная норма» и «унитарно ковариантная норма» выбирался первый как наиболее точный.

Необходимость периодической корректировки предметно-систематического указателя может быть связана с количественными и качественными изменениями документального потока, поступающего в систему, например, появление нового тематического направления. Процедура ведения указателей предполагает обязательное использование средств автоматизации, позволяющих оперативно производить такие трудоемкие операции, как алфавитная сортировка словаря, частотный анализ лексик.

Фундаментальные понятия в области рассматриваемых разделов вытекают из определения этих разделов.

**Линейная алгебра** — один из разделов высшей математики, являющийся важной в приложениях частью алгебры. Предметом изучения линейной алгебры служат векторы, матрицы, векторные, или линейные пространства, линейные отображения и системы линейных уравнений. Векторные пространства встречаются в математике и её приложениях повсеместно. Этот раздел математики широко используется в абстрактной алгебре и функциональном анализе и находит многочисленные приложения в естественных науках. Построение теории систем линейных уравнений потребовало таких инструментов, как теория матриц и определителей, и естественно привело к появлению теории векторных пространств (<http://ru.wikipedia.org>).

В УДК тематическое направление «линейная алгебра» содержит следующие разделы:

512.64 — **Линейная алгебра** с подрубриками:

512.642 — Векторные пространства. Теория векторных пространств;

512.643 — Матрицы и линейные отображения. Теория матриц;

512.644 — Системы линейных уравнений и неравенств;

512.645 — Линейные отображения;

512.646 — Билинейные, эрмитовы и квадратичные формы.

Тематическое направление «Поля и многочлены» в УДК содержит следующие разделы:

512.62 — **Поля и многочлены** с подрубриками:

512.622 — Многочлены;

512.623 — Общая теория полей;

512.624 — Конечные поля;

512.625 — Локальные поля;

512.626 — Глобальные поля;

512.627 — Теория полей классов, теория Галуа, дзета и L-функции;

512.628 — Поля с дополнительными структурами.

**Гомологическая алгебра** изучает алгебраические объекты, заимствованные из алгебраической топологии, в которой она играет важную роль. Гомологическая алгебра применяется в теории групп, теории алгебр, алгебраической геометрии. Одним из истоков гомологической алгебры явилась теория гомологии топологических пространств, в которой каждому топологическому пространству  $X$  сопоставляется последовательность групп гомологии.

В процессе изучения групп гомологии сложился, прежде всего, язык гомологической алгебры. Появились обозначения отображений с помощью стрелок и коммутативные диаграммы (если в графе отображений два пути, имеющие общие начало и конец, приводят к одному и тому же результату, то такую диаграмму называют коммутативной). Стало обычаем задавать математические объекты одновременно с их отображениями, а наиболее предпочтительными считались соответствия между объектами, сохраняющие отображения, названные *функторами*. В настоящее время язык гомологической алгебры присутствует во многих работах, даже не использующих ее методов.

Тематическое направление «Гомологическая алгебра» представлено в УДК следующими разделами:

512.66 — **Гомологическая алгебра** с подрубками:

512.662 — Цепные комплексы;

512.664 — Производные функторы;

512.666 — Алгебраическая K-теория;

512.667 — Алгебраические аналоги различных конструкций из топологии и алгебраической геометрии.

После экспертного анализа с учетом тематического содержания разделов, фундаментальных понятий в них и терминологии наименований кодов УДК точности описания разделов «Линейная алгебра», «Поля и многочлены» и «Гомологическая алгебра» с помощью КС, кодов классификатора УДК и рубрик разделов 271.17.29, 271.17.27 и 271.17.31 БД «Математика» алфавитный предметно-



систематический указатель каждого из разделов содержал около 2300 слов и словосочетаний.

Мониторинг терминологии разделов был проведен с использованием информационного массива БД ВИНТИ РАН «Математика».

### ***Принципы формирования словарей по отдельным разделам математики***

Отбор для словаря терминов и фразеологических единиц, представляющих наибольшие трудности для глубокого понимания текстов публикаций, сможет помочь специалисту не только правильно понять смысл термина, но и его стилистическую окраску.

На основе созданных двуязычных предметно-систематических указателей к разделам «Линейная алгебра», «Поля и многочлены» и «Гомологическая алгебра» БД ВИНТИ РАН «Математика» появилась возможность составления англо-русских и русско-английских словарей по каждому из разделов математики.

Особенностями таких словарей является расширение терминологической базы рассматриваемых разделов за счет включения новых терминов и фразеологических сочетаний, единообразного их толкования, более широкое представление сложных слов; предметное индексирование терминов с помощью таблиц УДК, используя данные из отражаемых в БД ВИНТИ первоисточников математической литературы.

Для составных терминов принята алфавитно-гнездовая система, по которой термины, состоящие из определяемых слов и определений, следует искать по определяемому словам, например, *элементарная матрица*, *косорефлективная матрица*, *кватернионная матрица* следует искать в гнезде **матрица**.

### ***Заключение***

Информационные онтологии служат для формального описания и уточнения понятий, определения общей терминологической базы в предметных областях. На примере разделов «Линейная алгебра», «Поля и многочлены» и «Гомологическая алгебра» БД ВИНТИ РАН «Математи-

ка» продемонстрирован онтологический подход к созданию словаря терминов и связанных с ними кодов УДК. Банк данных ВИНТИ РАН позволяет проводить отбор терминов и фразеологических единиц по основным разделам алгебры. Данные по накоплению ретрофонда и развитию классификационных систем, созданию двуязычных предметно-систематических указателей к разделам «Линейная алгебра», «Поля и многочлены», «Гомологическая алгебра» и на их основе составление англо-русских и русско-английских словарей по каждому из разделов математики дают возможность проводить сопоставление с мировыми информационными ресурсами.

*Литература:*

1. Никольская И. Ю., Ефременкова В. М. Математика в ВИНТИ РАН: от реферативного журнала к базам данных. Ч.1. История развития Реферативного журнала «Математика» в период 1953-2007 гг. //ИТИ. Сер. 1, 2008. № 5. С. 27-37.
2. Никольская И.Ю., Голод Е.С., Ефременкова В.М., Вахрамеев С.А. Предметно-систематический указатель к разделу «Линейная алгебра» БД/РЖ ВИНТИ РАН «Математика»./ ВИНТИ. М. 2012. 291 с. — Библиогр.: 17 назв.- Рус. — Деп. В ВИНТИ 21.03.12 № 100-В2012.
3. Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтология и тезаурусы: Учебное пособие. — Казань, Москва: Казанский государственный университет, МГУ им. М.В. Ломоносова, 2006.