

ОБРАЗОВАТЕЛЬНЫЕ ЭЛЕКТРОННЫЕ ИЗДАНИЯ И РЕСУРСЫ

РАЗРАБОТКА И ОПИСАНИЕ ГИПЕРТЕКСТОВОГО ИНФОРМАЦИОННО-ПОИСКОВОГО ТЕЗАУРУСА ПО АЛГЕБРЕ

И.В. Кузнецова

Коряжемский филиал Поморского государственного
университета им. М.В. Ломоносова
пр. Ленина, 9, Архангельская область, Коряжма, Россия, 165653

С.В. Лесников

Сыктывкарский государственный университет
Октябрьский пр., 55, Сыктывкар-1, Россия, 167001

В статье описана разработка гипертекстового информационно-поискового тезауруса по алгебре в процессе конструирования гипертекстового информационно-поискового тезауруса метаязыка науки.

Ключевые слова: алгебра, гипертекст, дефиниция, математика, метаязык, тезаурус.

Особенностью моделируемого информационно-поискового тезауруса (далее ИПТ) является то, что сама его структура дает возможность пользователю осуществлять поиск информации не только по отдельным ключевым словам, но и по смысловым полям, разделам математики и терминологическим парадигмам.

Суть гипертекстового ИПТ по алгебре заключается в расширении возможностей современных поисковых систем для обеспечения запросов пользователей. Используя тезаурус, можно с достаточной степенью эффективности получать результаты поиска, соответствующие информационной потребности пользователя, уместности и состоятельности результатов. Программа выдает пользователю не только информацию о ресурсах, содержащих введенные ключевые слова, но и семантически связанные с ними термины, соответствующие дескрипторы, дефиниции, иллюстрации, интерпретации, леммы, объяснения, определения, понятия, разъяснения, толкования, трактовки, формулировки, экскурсы и эксцерпции. Программно сконструировав гиперссылки, на базе такого тезауруса можно также подсказать пользователю, по каким еще вокабулам целесообразно продолжить поиск (например, антоним, гипероним, гипоним, квазисиноним, лексико-семантический вариант, мероним, холоним).

Современной тезаурус — это особая разновидность словарей общей или специальной лексики, в которых указаны семантические отношения (синонимические, антонимические, паронимические и т.п.) между лексическими единицами. Тезаурусы в электронном формате являются одним из действенных инструментов для описания отдельных предметных областей и создания информационно-поисковых систем. Первым тезаурусом фактически был *Thesaurus of English Words and Phrases*, составленный британским лексикографом Питером Марком Роже около 1805 г. и опубликованным в 1852 г. [5].

Словарь П. Роже — это свод английской лексики, в котором выделены шесть тематических классов, разбитых, в свою очередь, на 24 подкласса, каждый из которых представляет собой объединение более мелких семантических групп (1000 тем). При этом возможны пересечения классов. Темы — типовые семантические контексты для слов; класс семантических контекстов слова — множество тем, которые могут быть ему приписаны.

В гипертекстовом ИПТ по алгебре (в формате компьютерного, гипертекстового тезауруса (ГИЗАУРУСА) — базы данных с перекрестными гипертекстовыми ссылками) дается определение (дефиниция, иллюстрация, интерпретация, объяснение, понятие, пояснение, разъяснение, толкование, трактовка, формулировка, цитата, экскурс и/или эксцерпция) ключевых понятий-терминов, относящихся к сфере алгебры и ряду смежных дисциплин, которые в той или иной мере связаны с ее проблемами. Конструирование тезауруса обусловлено необходимостью единства в терминологии — нередко ученые одно и то же явление называют по-разному, с другой стороны, один и тот же термин бывает полисемичен даже в рамках одной научной области. Кроме этого, при анализе уже опубликованных терминологических словарей определенной науки бросается в глаза лакунарность словников и некая тенденциозность в подаче научных материалов авторами-составителями. Фактически речь идет о составлении метаязыка науки — особого языка, объектом которого является содержание и выражение другого языка, т.е. в нашем случае метаязык — это язык второго порядка как специальная семиологическая система, употребляемая тогда, когда надо говорить о естественном (или искусственном) языке, выступающем в качестве языка-объекта.

В метаязыке можно выделить следующее:

— собственно термины (слова, которые или совсем не употребляются в языке-объекте, или приобретают после заимствования из языка-объекта особое значение);

— сочетания слов, которые характеризуют собственно метаязык определенной науки в его полном виде (например, составные термины);

—определенный социально-предметный аспект (например, когда метаречь отражает специфику эпохи, того или иного научного направления, школы).

Одноязычный ИПТ — словарь лексических единиц дескрипторного языка, основанный на лексике одного естественного языка, отображающий семантические отношения между лексическими единицами (выбранные для использования в ИПТ слова, словосочетания или лексически значимые компоненты сложных слов естественного языка) и предназначенный для организации поиска информации путем индексирования документов и/или запросов. Главной целью создания ИПТ

является повышение показателей качества поиска информации в автоматизированных информационно-поисковых системах. Тезаурусы, как правило, используют в своей работе специалисты, в той или иной степени связанные с подготовкой и обработкой текстовой информации: переводчики, редакторы, научные работники, референты и др. В свою очередь, ИПТ используются при индексировании документов и запросов.

Индексирование можно осуществлять следующими способами:

— бестезаурусно-поисковый образ документа представляется в виде списка ключевых слов, выбранных непосредственно из текста документа (требуется меньше времени и менее трудоемко, но возможна синонимия и нельзя отразить парадигматические отношения);

— с тезаурусным контролем лексики для выражения содержания документа и запроса — используются только дескрипторы, входящие в ИПТ (позволяет исключить синонимию и учесть парадигматические связи, что, естественно, повышает эффективность поиска).

ИПТ подразделяют на два типа:

— выделяющие среди своих лексических единиц дескрипторы (дескриптор — лексическая единица ИПТ, предназначенная для использования в поисковых образах документов и/или запросов) и аскрипторы (недескриптор — лексическая единица ИПТ, которая в поисковых образах документов (запросов) подлежит замене на дескриптор при поиске или обработке информации);

— все лексические единицы, которые являются дескрипторами.

ИПТ может быть применен для перевода текста с естественного языка на дескрипторный информационно-поисковый язык путем замены ключевых слов соответствующими дескрипторами, при этом исключается многозначность; отображения парадигматических связей между дескрипторами.

Правила разработки, структура, состав и форма представления ИПТ определены ГОСТ 7.25-80 (постановление Государственного комитета СССР по стандартам от 31 декабря 1980 г. № 6340; введены в действие с 1 января 1982 г.) и ориентированы на использование лексики русского языка и разрабатываемых в рамках автоматизированных систем научно-технической информации.

Построение ИПТ включает следующие этапы:

- 1) определение тематического охвата ИПТ;
- 2) сбор массива лексических единиц;
- 3) формирование словника ИПТ;
- 4) построение словарных статей и указателей;
- 5) оформление ИПТ;
- 6) экспертиза и регистрация ИПТ.

В базе данных разработанного ИПТ по алгебре выделено несколько разделов: основные алгебраические структуры, кольцо многочленов от одной переменной, матрицы, отношения и отображения, системы линейных уравнений, многочлены над числовыми полями, а также отдельно сформирован электронный словарь-справочник основных понятий, определений и теорем по алгебре.

При разработке данного ИПТ были использованы программные средства Visual Basic и СУБД Microsoft Access. Перед тем как начать работу с ИПТ по алгебре, необходимо установить программу и базу данных на компьютере. Для этого двойным щелчком запускаем мастер установки программы (файл «Установка ИСС»). После соблюдения всех пунктов установки можно приступить к работе с данной программой. Работа с ИПТ по алгебре начинается с главной формы, на которой написано ее название. На приветственной форме находятся две кнопки: «Запустить программу» и «Выход».

После запуска программы нужно заполнить базу данных интересующей нас информацией. Для этого можно или сначала создать новую базу данных или использовать уже имеющуюся. На панели инструментов находится кнопка «Создать базу данных», при нажатии на нее появится окно, в котором необходимо указать путь сохранения новой базы данных и указать ее имя в строке «Имя файла».

После того как создали базу данных, появляется форма, в которой спрашивается, сколько таблиц будет содержать создаваемая база данных. Поскольку будет создаваться только одна таблица, после нажатия на кнопку «Нет» появится еще одна форма, в которой нужно указать имя создаваемой таблицы. Для того чтобы вносить информацию, необходимо создать столбцы и строки, далее необходимо указать поля таблицы, для этого нужно щелкнуть правой кнопкой мыши по заголовку таблицы. В раскрывшемся списке необходимо выбрать пункт «Добавить новое поле», после этого появится столбец, необходимо указать его название и нажать клавишу «Ввод (Enter)» для подтверждения имени поля. Эту же операцию повторять столько раз, сколько необходимо иметь столбцов. В сконструированном ИПТ по алгебре имеем четыре столбца со следующими именами: «Понятия и теоремы», «Формулировка», «Литература», «Некоторые примеры и доказательства теорем». Аналогично тому как создавались столбцы для таблицы, создаются строки и заполняются соответствующей информацией. Количество строк в данной таблице не ограничено.

После заполнения всей таблицы необходимо сохранить базу данных. Для этого на панели инструментов существует кнопка «Сохранить изменения». При нажатии на эту кнопку база данных сохранится. Но если вдруг забыли сохранить все то, что ввели, и решили создать или открыть другую базу данных, то появится предупреждение о том, что несохраненные данные будут утеряны. После сохранения всей внесенной информации можно закрывать программу. Вся информация будет храниться в таблице базы данных, которую создавали в начале работы с программой.

Итак, база данных создана, таблица заполнена соответствующей информацией и сохранена. Теперь можно приступить к самой работе программы с использованием ранее созданной базы данных. После запуска программы появляется приветственная форма, на которой написано ее название. Также на этой форме находится кнопка «Запуск программы», при нажатии на которую происходит запуск самой программы. Но данная форма пустая, для того чтобы ее заполнить, нужно открыть соответствующую базу данных. Сделать это можно с помощью кнопки на панели инструментов «Открыть базу данных». После нажатия на дан-

ную кнопку появится окно «Открыть файл». В открывшемся окне можно указать путь сохранения базы данных.

Существует еще один способ открытия базы данных. Ниже панели инструментов находится строка с надписью «Быстрый доступ к базе данных» и поле со стрелочкой. При нажатии на эту стрелку появляется список. В раскрывшемся списке находятся пути хранения баз данных, которые были созданы с помощью данной программы. В этом списке находим нужную базу данных, наводим на нее курсор мыши и левой кнопкой мыши активизируем ее. При нажатии в первом столбце по любому разделу во втором столбце появятся соответствующие данному разделу понятия и теоремы.

Поиск дефиниций в созданной базе данных может осуществляться двумя способами. Первый способ предусматривает последовательный просмотр вручную всех имеющихся в ИПТ данных. Для этого с правой стороны таблицы есть полоса прокрутки, с помощью которой можно двигаться сверху вниз, и внизу таблицы, чтобы двигаться слева направо. При нажатии в первом столбце по любому разделу во втором столбце появятся соответствующие данному разделу понятия и теоремы.

Второй способ состоит в наборе нужного слова в специальном окне. Для этого на панели инструментов есть поле, в которое необходимо ввести искомое понятие (рисунок). После ввода необходимой для поиска информации надо нажать на кнопку «Начать поиск», которая находится слева от строки ввода.

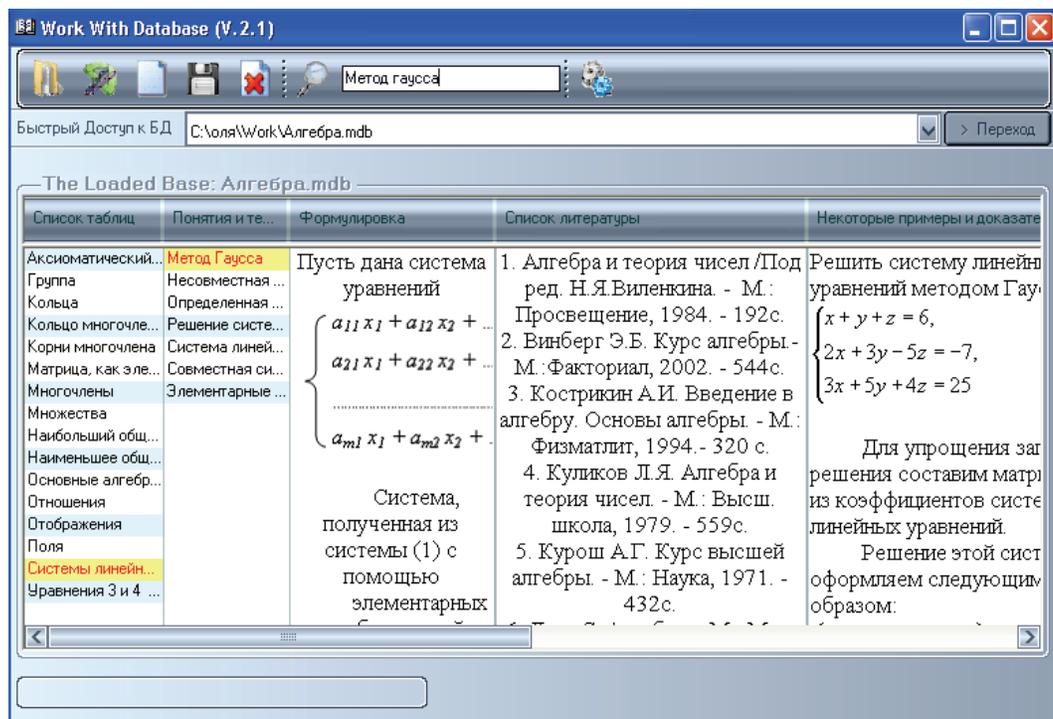


Рис. Результат поиска дефиниций в ИПТ

Сконструированный гипертекстовый ИПТ по алгебре совместим с разными версиями, например Windows XP и Windows 7.

Для редактирования ИПТ необходимо переключиться в режим редактора с помощью кнопки «Редактировать» на панели инструментов. И только после этого можно вносить любые изменения: добавлять, удалять, исправлять названия таблиц, строк и столбцов и т.д.

Допустим, что при создании базы данных была допущена ошибка в названии таблицы. Чтобы устранить эту ошибку, нужно навести курсор мыши на данное название и правой кнопкой мыши нажать по нему. В раскрывшемся списке необходимо выбрать пункт «Переименовать таблицу», после этого можно изменить название таблицы и подтвердить ввод клавишей «Ввод (Enter)». А если нужно удалить одну из строк какого-либо раздела, то сначала нужно выбрать раздел, в котором находится эта строка, навести на нее курсор и правой кнопкой мыши активизировать соответствующий раздел. В раскрывшемся списке выбрать пункт «Удалить строку полностью». Данная строка из этого раздела удалится.

При добавлении строк в базу данных необходимо выбрать в раскрывшемся списке пункт «Добавить новую строку». Новая строка появится в самом конце таблицы. В эту строку соответственно столбцам вносим необходимую информацию. После всех внесенных в базу данных изменений необходимо сохранить базу данных, иначе она останется такой, какой была до изменения.

Таким образом, в режиме «Редактор» можно выполнять много действий над базой данных. Это очень удобно, позволяет держать базу данных в актуальном состоянии, обновлять ее, добавлять недостающую информацию или наоборот удалять из нее, что-то лишнее. На панели инструментов есть еще одна кнопка «Настройки», которая находится после строки ввода для поиска. При нажатии на данную кнопку появляется форма, которая разбита на две части — левую и правую. В левой части формы находится цветовая схема. С помощью данной схемы можно менять цвет оформления самой программы. Первые пять цветов уже заданы в самой программе, а шестой можно задать самим. Чтобы выбрать один из заданных цветов, нужно с левой части схемы навести курсор на кружок и левой кнопкой мыши выбрать тот цвет, который нужен, после выбора цвета нужно нажать внизу формы на кнопку «Применить».

Если нужно, чтобы база данных загружалась автоматически при загрузке программы, то необходимо установить галочку напротив надписи «Загружать БД по умолчанию» и ниже в поле ввести путь хранения той базы данных, которая должна загружаться, нажать кнопку внизу формы «Применить». Это удобно тогда, когда постоянно работаете только с одной базой данных. Если работаем с несколькими базами данных, то лучше всего поставить галочку перед надписью «Вести учет загружаемых файлов в списке быстрого доступа». Тогда все базы данных, которые были обработаны с помощью данной программы, будут отражаться в списке быстрого доступа к базе данных. Это очень удобно для переключения с одной базы данных на другую, не придется затрачивать много времени на поиск необходимой базы данных.

Разработанный гипертекстовый ИПТ по алгебре является своего рода инструментальной программной оболочкой, к дидактическим функциям которой относятся информационная, систематизирующая и самообразовательная составляющие. Использование ИПТ в учебном процессе по алгебре позволит индивидуализировать обучение, сформировать информационную культуру, подготовить обучаемых к жизни в современном информационном обществе.

Процесс использования гипертекстового ИПТ в целом напоминает обычную работу в системе Интернет, но в локальном варианте на персональном компьютере. В перспективе предполагается, что после необходимой апробации в локальном варианте в режиме диалога сконструированный тезаурус будет доступен и в Интернете, например, по адресам <http://lsw.ru>, <http://ЛСВ.РФ>.

ЛИТЕРАТУРА

- [1] *Кондаков Н.И.* Логический словарь-справочник / Отв. ред. Д.П. Горский. — М.: Наука, 1975.
- [2] Новейший словарь иностранных слов и выражений. — М.: Современный литератор, 2005.
- [3] *Смирнов В.* Полный словарь иностранных слов, вошедших в русский язык с общедоступным толкованием их значения и употребления, и со включением отдельных слов и фраз, употребляющихся в устной и письменной речи в их оригинальной иностранной форме. — М.: Издание торгового дома В. Коновалова и К, 1908.
- [4] *Шайкевич А.Я.* Тезаурус // Лингвистический энциклопедический словарь. — М.: Советская энциклопедия, 1990.
- [5] *Roget P.M.* Thesaurus of English Words and Phrases. — New ed. N.Y., 1964. URL: <http://www.gutenberg.org/files/10681/10681-body.txt>

WORKING OUT AND THE DESCRIPTION OF THE HYPERTEXT INFORMATION RETRIEVAL THESAURUS ON ALGEBRA

I.V. Kuznetsova

Korjzhma Branch of the Pomor State University named after M.V. Lomonosov
Lenina str., 9, Arkhangel Region, Koryazma, Russia, 165653

S.V. Lesnikov

Chair of mathematical modeling and cybernetics
Syktyvkar state university
55 October str., Syktyvkar-1, Russia, 167001

In article working out of the hypertext information retrieval thesaurus on algebra in the course of designing of the hypertext information retrieval thesaurus of a meta language of a science is described.

Key words: algebra, the hypertext, a definition, mathematics, a meta language, the thesaurus.