

# ОБРАЗОВАТЕЛЬНЫЕ ЭЛЕКТРОННЫЕ ИЗДАНИЯ И РЕСУРСЫ

## СОЗДАНИЕ ПРОТОТИПА ЭЛЕКТРОННОГО ОБРАЗОВАТЕЛЬНОГО РЕСУРСА ПО ИНФОРМАТИКЕ НА ОСНОВЕ ЛИНГВИСТИЧЕСКОЙ ОНТОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ

**А.В. Абрамов**

Кафедра программного обеспечения  
и администрирования информационных систем  
Курский государственный университет  
*ул. Радищева, 33, Курск, Россия, 305000*

В статье рассматривается механизм создания прототипа электронного образовательного ресурса по информатике на основе лингвистической онтологии предметной области.

**Ключевые слова:** онтология, информатика, термин, информация.

Ранее нами был рассмотрен механизм накопления терминологической базы образовательной предметной области для цели создания лингвистической онтологии по информатике[1]. Напомним некоторые основные моменты, отраженные в предыдущей публикации. Под лингвистической онтологией понимается иерархическая сеть терминов. Каждое понятие связывается отношениями с другими понятиями лингвистической онтологии. Процесс создания лингвистической онтологии состоит из следующих этапов: формирование терминологической базы некоторой предметной области по массиву текстовой информации; анализ полученной информации человеком (экспертом) с целью «фильтрации» терминов и указания определения данных терминов; установление человеком (экспертом) отношений между набором терминов предметной области.

В результате выполнения данных этапов на выходе системы получаем множество  $X'$  слов-«кандидатов» вида:  $X' = \{x, V\}$ , где  $x$  — слово-«кандидат»;  $V$  — вес «кандидата». Далее в модуле редактирования терминов (рис. 1) человеком (экспертом) осуществляется «фильтрация» терминов, указание соответствующих им определений, установление отношений между набором терминов предметной области. На завершающем этапе формируется выходной файл лингвистической онтологии, представляющей собой совокупность терминов, рассматриваемых в данном курсе информатики, их определений и словоформ, а также частоты их встречаемости в обрабатываемом тексте.

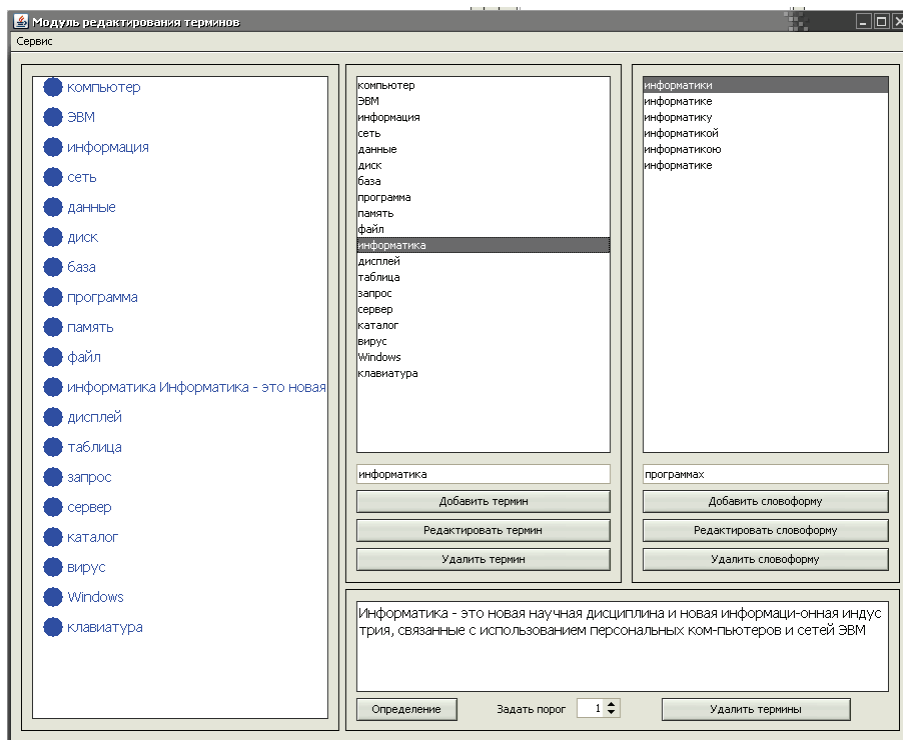


Рис. 1. Модуль редактирования онтологии

Структура данного файла представлена в виде схемы на рис. 2.

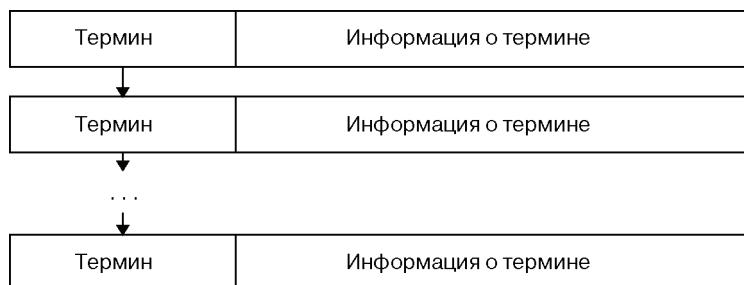


Рис. 2. Структура файла лингвистической онтологии по информатике

После этого файл, содержащий лингвистическую онтологию по информатике, передается в модуль создания HTML-страницы (рис. 3).

При запуске данного модуля файл онтологии подгружается в оперативную память в специально организованную динамическую структуру, оптимизированную для процесса преобразования текста из предметной области в html-страницу.

В данной структуре предусмотрены поля для хранения всех данных, находящихся в файле онтологии, а также поле, хранящее ссылку на список индексов термина. Под словом «индекс» в данном контексте понимается упорядоченная пара целых чисел  $(m, n)$ , где  $m$  — номер предложения, содержащего термин, а  $n$  — номер слова в данном предложении. Данный список динамически заполняется для каждого обрабатываемого текста.

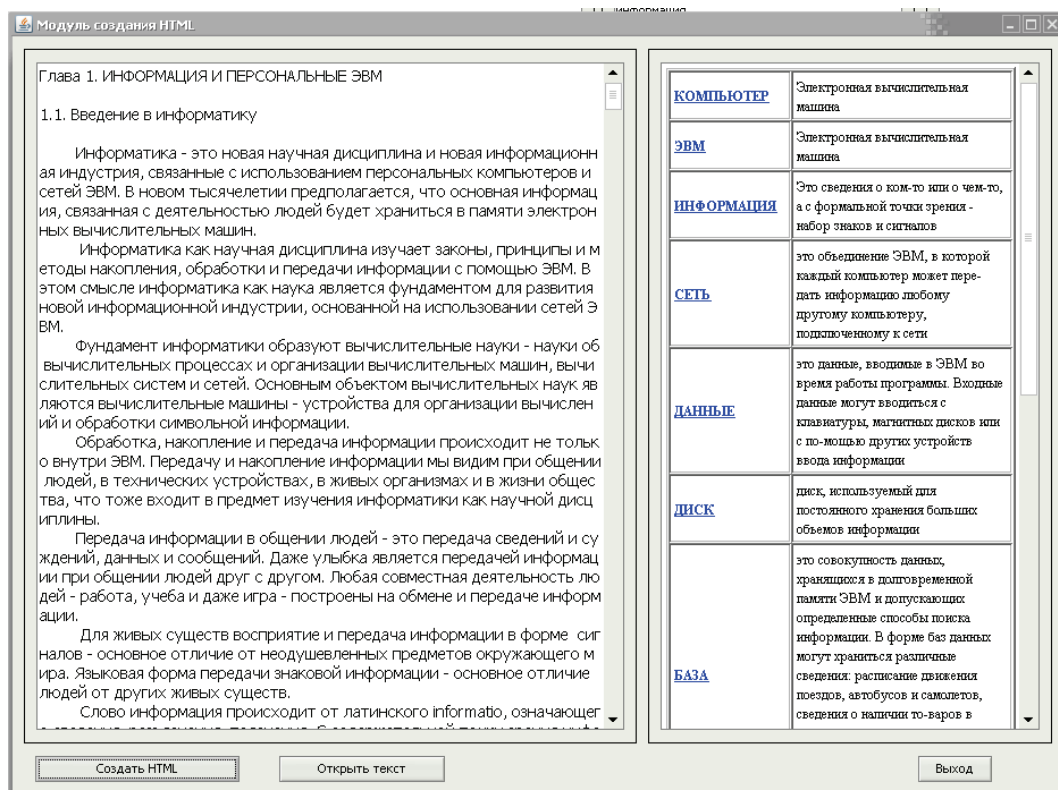


Рис. 3. Модуль редактирования лингвистической онтологии

Таким образом, лингвистическая онтология по информатике с точки зрения организации динамической памяти представляет собой трехмерную совокупность данных: список терминов, список словоформ терминов, список индексов для всех словоформ.

Процесс нахождения всех возможных вхождений терминов онтологии (а также их словоформ) в тексте основывается на его графематическом анализе. Текстовая информация обрабатывается посимвольно, и при каждом событии выделения очередной лексемы проверяется ее эквивалентность (с учетом словоформ) терминам из онтологии. При успешной проверке текущие значения номера предложения и номера слова в предложении заносятся в список индексов для данного термина.

Для хранения номеров используется два временных целочисленных счетчика. Первый счетчик хранит номер предложения. Он изменяется на основе анализа встречаемости точек в тексте. Второй счетчик — номер слова в предложении — инкрементируется при каждом событии выделения лексемы и сбрасывается при увеличении номера предложения. На рисунке 4 изображена обобщенная схема алгоритма индексирования терминов в тексте.

По завершении процесса обработки динамическая структура терминов онтологии содержит всю необходимую информацию для генерации html-страницы, которая будет содержать термины предметной области с их определением, с возможностью перехода по ссылкам ниже на уровень, содержащий подробную информацию о данном термине (рис. 5).

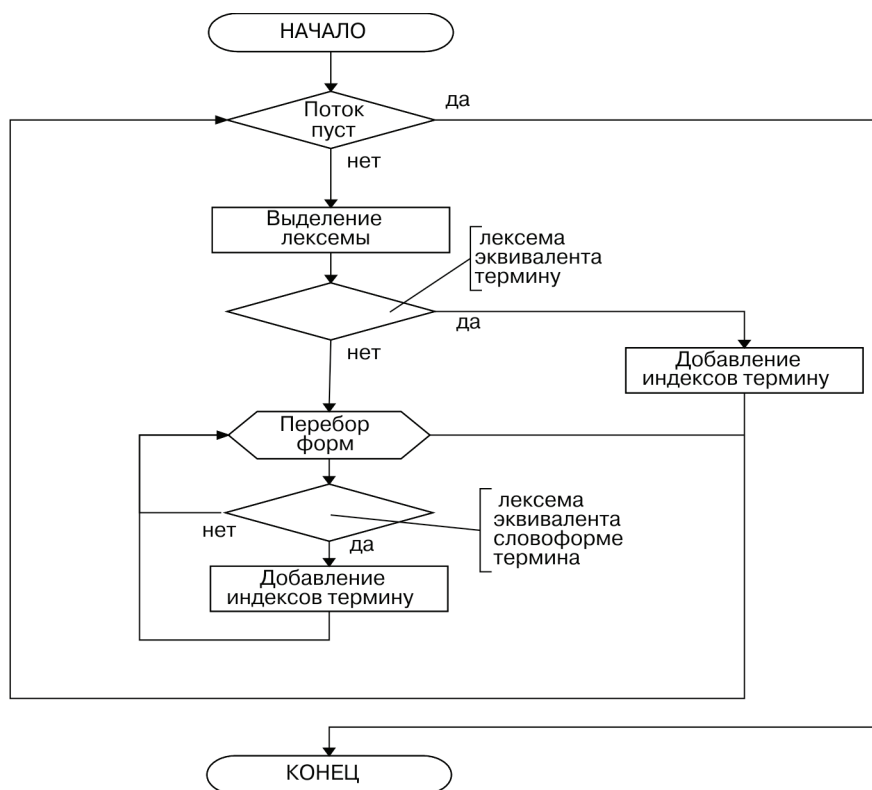


Рис. 4. Схема алгоритма индексирования

<b>КОМПЬЮТЕР</b>	это-электронная вычислительная машина
<b>ЭВМ</b>	Электронная вычислительная машина
<b>ИНФОРМАЦИЯ</b>	Это сведения о ком-то или о чем-то, а с формальной точки зрения - набор знаков и сигналов
<b>СЕТЬ</b>	это объединение ЭВМ, в которой каждый компьютер может пере-дать информацию любому другому компьютеру, подключенному к сети
<b>ДААННЫЕ</b>	это данные, вводимые в ЭВМ во время работы программы. Входные данные могут вводиться с клавиатуры, магнитных дисков или с по-мощью других устройств ввода информации
<b>ДИСК</b>	диск, используемый для постоянного хранения больших объемов информации
<b>БАЗА</b>	это совокупность данных, хранящихся в долговременной памяти ЭВМ и допускающих определенные способы поиска информации. В форме баз данных могут храниться различные сведения: расписание движения поездов, автобусов и самолетов, сведения о наличии то-варов в магазине или на складе, сведения о студентах, преподавате-лях и сотрудниках, информация о книгах и многое, многое другое
<b>ПРОГРАММА</b>	форма представления данных и команд, предназначенных для полу-чения определенных результатов или способа функционирования ЭВМ
<b>ПАМЯТЬ</b>	
<b>ФАЙЛ</b>	это последовательность записей на машинных носителях - магнит-ных или оптических дисках, магнитных или перфокартах и т. п.
<b>ИНФОРМАТИКА</b>	Информатика - это новая научная дисциплина и новая информаци-онная индустрия, связанные с использованием персональных ком-пьютеров и сетей ЭВМ
<b>ДИСПЛЕЙ</b>	устройство отображения информации на электронном экране
<b>ТАБЛИЦА</b>	Структура для выполнения и хранения различных расчетов и каль-куляций на компьютерах
<b>ЗАПРОС</b>	Основой для поиска информации в базах данных служат запросы. Совокупность запросов, по которым можно получить информацию, считается главной характеристикой баз данных.

Рис. 5. HTML-страница с терминами

В данной статье была рассмотрена лингвистическая онтология предметной области, используемая в качестве средства создания электронного образовательного ресурса по информатике на основе гипертекстовой модели, позволяющая получать при анализе только релевантные данные и «не замечать» информацию, не представляющую интереса. Также на примере разработанного программного комплекса DisOnto охарактеризованы этапы создания электронного образовательного ресурса по информатике использующего гипертекстовую модель.

#### **ЛИТЕРАТУРА**

- [1] *Абрамов А.В. Создание лингвистической онтологии образовательной предметной области. 2010. URL: <http://scientific-notes.ru/pdf/014-5.pdf>*

## **CREATION OF THE PROTOTYPE OF THE ELECTRONIC EDUCATIONAL RESOURCE ON COMPUTER SCIENCE ON THE BASIS OF THE LINGUISTIC SUBJECT DOMAIN ONTOLOGIES**

**A.V. Abramov**

Software and administration chair information systems

Kursk state university

*Radishcheva str., 33, Kursk, Russia, 305000*

This article is devoted to consideration of mechanism of creating prototype of electronic educational resource in computer science on the basis of linguistic ontology of subject domain

**Key words:** ontology, computer science, term, information.