

2. Найханова Л.В., Хаптахаева Р.Б., Янсанова Е.Н. Создание декларативного метода извлечения знаний из терминологических словарей // Информационные технологии. – 2008. – №12. – С. 2–8.
3. Wille R. Concept lattices and conceptual knowledge systems. // Computers and Mathematics with Applications. – №23, 1992.

*И.В. АНТОНОВ*

## **МОДЕЛЬ ОНТОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ СИСТЕМ СЕМАНТИЧЕСКИ-ОРИЕНТИРОВАННОГО ДОСТУПА**

В статье рассматривается модель онтологии предметной области и структура базы данных для хранения онтологии, поддерживающие возможности итерационного автоматизированного формирования структуры онтологии.

Онтологии предметной области в настоящее время находят основное применение в области построения поисковых систем, систем представления знаний, инженерии знаний и при решении задач семантической интеграции информационных ресурсов. Под онтологией понимается «формальная спецификация концептуализации, которая имеет место в некотором контексте предметной области» (Т. Грубер, 1993 [1]). В свою очередь, концептуализация - представление предметной области через описание множества понятий (концептов) предметной области и связей (отношений) между ними. Основным отношением, учитываемым при построении онтологии, является родовидовое отношение между понятиями (отношение гипоним-гипероним), на основе которого формируется таксономия понятий. Представление совокупности понятий предметной области и их отношений в основном реализуется в современных онтологических системах на основе модели семантической сети фреймов. Узлы сети представляют отдельные понятия предметной области, дуги – отношения между понятиями. Отдельное понятие в этой модели представляется фреймом, слоты которого содержат атрибуты понятия. Производные (дочерние) понятия наследуют атрибуты базовых (родительских) понятий. На этапе определения понятий онтологии для их атрибутов обычно задается имя и тип атрибута. Конкретные значения эти атрибуты получают при создании на основе понятий онтологии экземпляров (объектов). Операции по созданию экземпляров понятий поддерживает большинство онтологических систем. При этом экземпляры чаще соответствуют понятиям нижних уровней онтологической иерархии. Таким образом, онтология представляет собой иерархию понятий, характеризующих предметный мир, объекты которого соответствуют преимущественно понятиям нижних уровней онтологии, а промежуточный и верхний ее уровни представляют, как правило, абстракции различной степени обобщения.

Существующие системы, построенные на основе онтологий, рассчитаны, как правило, на работу с онтологией программных агентов, обрабатывающих те или иные информационные запросы. Одним из перспективных направлений развития онтологических систем является построение систем, использующих онтологическую систематизацию как инструмент классификации объектов предметной области, с которыми работают пользователи, и как средство для организации семантически-ориентированного доступа пользователей к этим объектам. К числу потенциальных областей продуктивного применения указанного подхода относится работа пользователей персональных компьютеров с файлами и документами. Традиционные средства доступа к файлам основаны на выборе пользователем папок и файлов в иерархической структуре файловой системы. Инструментом доступа в таком случае является программа, реализующая функции файлового менеджера. С ростом числа файлов и усложнением структуры файловой системы поиск нужного документа и файла становится все более затруднительным для пользователей. Решением проблемы может быть организация доступа посредством семантически-ориентированных интерфейсов,

реализуемых на основе онтологий предметной области. В этом случае документ может быть найден и выбран на основе своих семантических признаков вне зависимости от его физического расположения на каком-либо диске и в какой-либо папке. При реализации указанных систем требуется, прежде всего, решить две задачи – построить онтологию предметной области и снабдить объекты предметной области (файлы и документы в рассматриваемом случае) семантическими аннотациями, на основе которых будет осуществляться доступ к объектам. Аннотированные ресурсы включаются в онтологическую систему в качестве объектов (экземпляров) онтологии. На основе такой системы навигация по коллекции доступных пользователям ресурсов может осуществляться посредством перемещения по уровням иерархического меню, пункты которого соответствуют понятиям предметной области разного уровня обобщения. В качестве частичной аналогии может рассматриваться приложение «Проводник», входящее в состав операционной системы Windows. Однако «Проводник» работает с файловыми путями, а рассматриваемая система - с семантическими категориями, реализуя таким образом «Семантический проводник» по информационным ресурсам. Отбор доступных пользователю объектов также может происходить на основе обработки запросов, задающих шаблоны и ограничения для атрибутов интересующих пользователя ресурсов. Таким образом, онтологическая информационная система может совмещать в себе функции системы навигации и поисковой системы.

Рассмотрим модель онтологии предметной области, на основе которой могут быть построены описанные выше системы семантически-ориентированного доступа.

Модель онтологии может быть представлена кортежем

$$Q = \langle C, M, R \rangle,$$

где  $C = \{c_i\}$  - множество понятий (концептов), образующих онтологию  $Q$ ,  $i = \overline{1, I}$ , т.е.

$$|C| = I;$$

$M_i = \{m_{1_i}, \dots, m_{d_i}\}$  - множество атрибутов понятия  $c_i$  ( $d_i$  - количество атрибутов, описывающих данное понятие);

$R \subseteq C \times C$  - отношение непосредственного наследования.

Отношение  $R$  удобно задавать матрицей размером  $I \times I$ : если понятие  $c_k$  в данной онтологии непосредственно наследует понятие  $c_i$ , т.е.  $(c_i, c_k) \in R$ , то элемент  $r_{ik} = 1$ , в противном случае, если  $(c_i, c_k) \notin R$ , то  $r_{ik} = 0$ .

Дочерние понятия наследуют атрибуты, входящие в состав родительского понятия, расширяя его состав собственными атрибутами:

$$\exists (c_i, c_k) \in R \rightarrow M_i \subset M_k$$

Онтология в общем случае поддерживает множественное наследование понятий, то есть:

$$\exists (c_j, c_k, c_l) : r_{il} = 1 \wedge r_{kl} = 1$$

Корректно построенная онтология не должна содержать ложных отношений наследования, связывающих понятие с предком его же родительского понятия:

$$\bar{\exists} (c_j, c_k, c_l) : r_{il} = 1 \wedge r_{kl} = 1 \wedge M_i = M_k \cap M_l$$

Машинная модель понятия онтологии, используемая для построения прикладных систем, должна включать в себя поля, содержащие имя понятия, состав атрибутов понятия и родовидовые связи понятия.

$$C = \langle N, A, S, D \rangle,$$

где  $N$  – имя понятия,  $A$  – множество атрибутов понятия,  $S$  – множество родительских понятий (суперклассов),  $D$  – множество дочерних понятий (подклассов). При этом  $A \subseteq M$ .

Атрибут понятия, в свою очередь, характеризуется именем, типом и значением:

$$A = \langle N_A, T, V \rangle,$$

где  $N_A$  – имя атрибута,  $T$  – тип атрибута,  $V$  – значение атрибута.

Атрибуты  $c_i$  понятий онтологии имеют определенный тип данных  $T$ . В качестве основных типов данных атрибутов онтологии можно выделить числовой, текстовый, логический и тип ссылки на объект:

$$T \in \{Int, Num, Txt, Bool, ObjR\}$$

На рисунке 1 изображена иерархия онтологической системы, включающая в себя понятия онтологии и экземпляры объектов. На каждом уровне представленной иерархии явно определяются только те атрибуты понятий, которые дополняют набор атрибутов, унаследованный от родительских понятий. На вершине иерархии находится корневое понятие, включающее в себя универсальные атрибуты, присущие любому понятию онтологии.

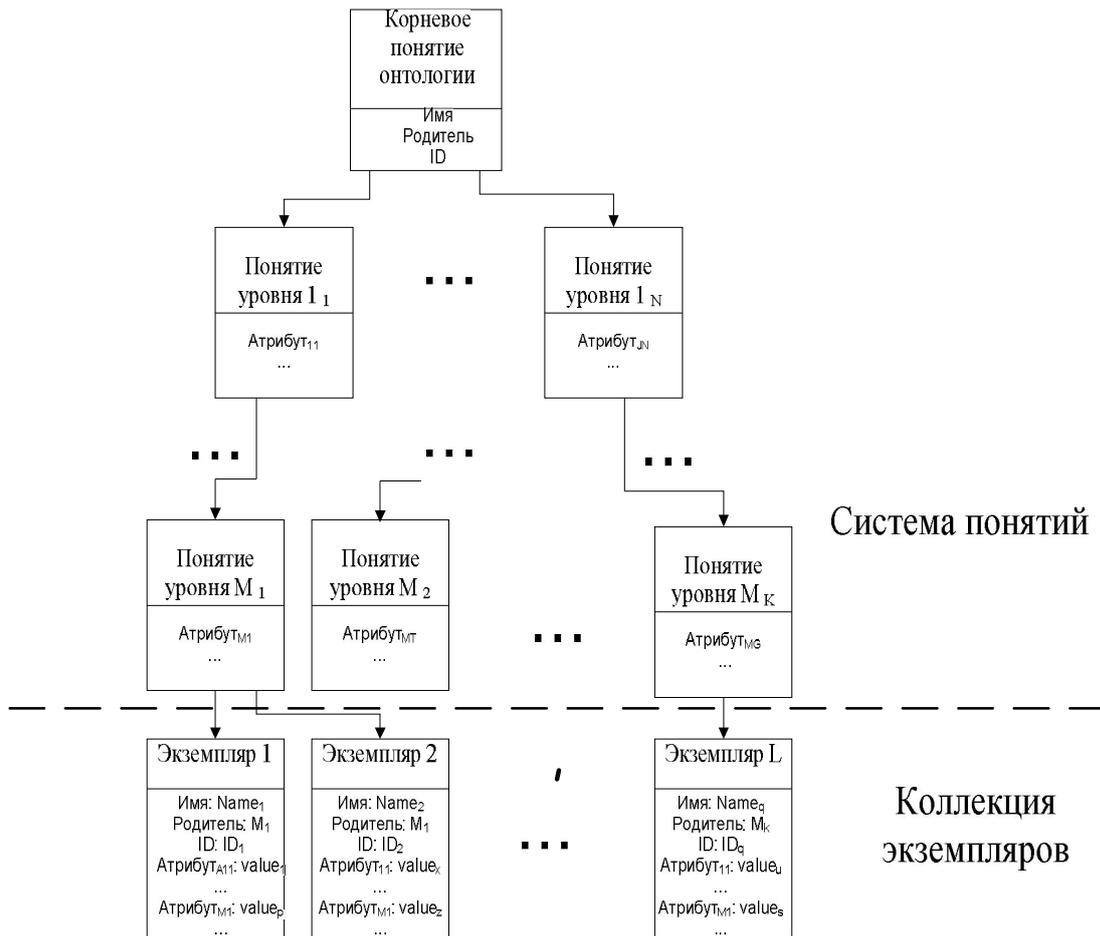


Рис. 1. Иерархия онтологической системы

Для машинного представления данных онтологии может быть использована реляционная база данных, включающая в себя ряд связанных между собой таблиц. Структура базы данных представлена на рисунке 2. Таблица *Classes* содержит идентификаторы, имена и описания понятий, включенных в онтологию. Записи таблицы *Parents* задают отношения наследования между понятиями онтологии. Таблица *Fields* содержит состав атрибутов каждого из понятий. В этой таблице для каждого из понятий указываются лишь атрибуты, определенные на уровне данного понятия. Наследуемые от родительских понятий атрибуты описываются в составе родительских понятий. Таблица *Attributes* содержит данные, определяющие полное множество атрибутов зарегистрированных в системе понятий. Таблица *AttrTypes* содержит определение набора типов атрибутов, используемых в системе.

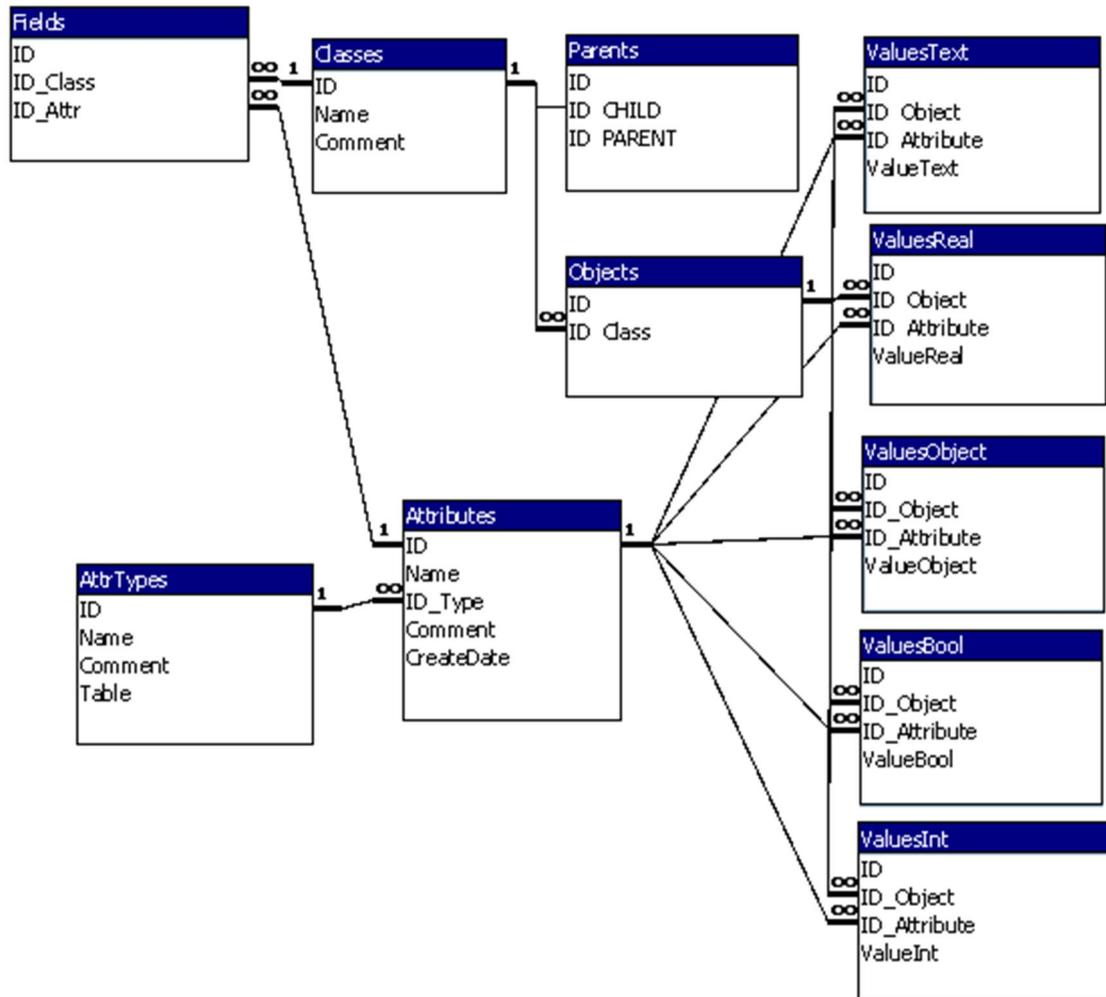


Рис. 2. Структура базы данных онтологии предметной области

Следующие шесть таблиц базы данных предназначены для хранения коллекции объектов онтологии. Таблица *Objects* содержит информацию о наборе зарегистрированных в системе объектов и их соответствии определенным понятиям. Таблицы *ValuesInt*, *ValuesReal*, *ValuesBool*, *ValuesText*, *ValuesObject* содержат значения атрибутов объектов, относящиеся к различным типам данных – целому, вещественному, логическому, текстовому и типу объектной ссылки.

На основе перечисленного выше набора таблиц может быть сформировано машинное представление онтологии предметной области, соответствующее рассмотренной математической модели онтологии.

Традиционным инструментом формирования онтологий являются редакторы онтологий, самым известным представителем которых является редактор Protégé [2]. В широко используемых сейчас редакторах не реализованы средства автоматизированного построения онтологий. В работе [3] рассматриваются методы автоматизированной поддержки формирования онтологий и возможность построения информационных систем на основе этих методов. Предлагаемая структура базы данных обеспечивает возможность использования рассмотренных в [3] методов автоматизированного формирования онтологий.

Предложенная машинная модель онтологии поддерживает динамическое формирование структуры входящих в состав онтологии предметной области моделей понятий и объектов, позволяя тем самым эффективно реализовать операции поэтапного формирования онтологий и редактирования ее структуры. На основе рассмотренной модели могут создаваться редакторы онтологий нового поколения, обеспечивающие автоматизированную поддержку операций онтологического инжиниринга.

**ЛИТЕРАТУРА**

1. Gruber T. R., A Translation Approach to Portable Ontology Specifications. // Knowledge Acquisition. – 1993, 5(2). – pp. 199–220.
2. Муромцев Д.И. Онтологический инжиниринг знаний в системе Protégé // СПб.: СПб ГУ ИТМО, 2007.
3. Антонов И.В., Воронов М.В. Формирование онтологических моделей предметной области для электронных обучающих систем // Информационные технологии в обеспечении нового качества высшего образования. Сборник научных статей. – Кн. 2. – М.: Исследовательский центр проблем качества подготовки специалистов, 2010. – С. 48–55.

*Ю.В. БРУТТАН*

### **РОЛЬ ФОРМАЛЬНЫХ СИСТЕМ СЕМАНТИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ТЕКСТОВЫХ ОПИСАНИЙ ПРИ ПРОЕКТИРОВАНИИ ЛИНГВИСТИЧЕСКИХ ПРОЦЕССОРОВ**

В статье рассмотрены вопросы понимания компьютером текстовых описаний на естественном языке. Определена роль формальных систем семантического представления текстовых описаний при проектировании лингвистических процессоров.

Пытаясь формализовать понимание компьютером даже довольно простых текстовых описаний, исследователи пришли к выводу о том, что для решения их частных задач необходимо предварительно иметь теоретические методы, относящиеся к произвольным текстам группы естественных языков (ЕЯ), например, русского или английского. Следовательно, требуется разработать такие формальные языки для представления знания о предметных областях и построения семантического представления (СП) принадлежащих им текстов на ЕЯ, чтобы можно было конструировать СП в виде структур, отражающих смысловое содержание этих текстов. Другими словами, нужны формальные языки (или формальные системы, поскольку множество их правильно построенных выражений образует язык) для описания смыслов ЕЯ-текстов, обладающие выразительными возможностями, близкими к возможностям естественного языка.

Формальные модели языка рассматриваются как компоненты различных прикладных систем. Компонента системы, реализующая формальную лингвистическую модель и осуществляющая смысловую обработку текстов на естественном языке, называется лингвистическим процессором (ЛП). Со стороны своего внутреннего устройства лингвистический процессор представляет собой многоуровневый преобразователь, состоящий из трёх уровней пофазного представления текста – морфологического, синтаксического, семантического [1]. Каждый из уровней обслуживается соответствующим компонентом модели – массивом правил и словарями. На каждом из уровней предложение имеет формальный образ, именуемой в дальнейшем структурой – морфологической, синтаксической и семантической структурами.

Под морфологической структурой понимается последовательность входящих в анализируемое предложение слов с указанием части речи и морфологических характеристик (падежа, числа, рода, одушевлённости и т.п.).

Под синтаксической структурой понимается дерево зависимостей, в узлах которого стоят слова данного ЕЯ с указанием части речи и грамматических характеристик, а дуги соответствуют специфичным для данного языка отношениям синтаксического подчинения (например, у синтаксического анализатора Link Parserfor Russian: Хр - связь между началом и концом предложения, Sp - связь между существительным и глаголом и т.д.)

Автором статьи предложена семантическая структура, рассмотренная в [2]. Эту структуру можно назвать «семантическим образом» текстового описания, т.к. она