

УДК 004.04

М.В. Новожилова, И.А. Михеев, С.А. Ямницкий

Харьковский национальный университет строительства и архитектуры, Харьков

МОДЕЛИРОВАНИЕ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В ЭЛЕКТРОННЫХ ТЕРМИНОЛОГИЧЕСКИХ СИСТЕМАХ

В данной статье предложена модель графического представления знаний в электронных терминологических системах. Создание модели осуществлялось на примере информационной системы «Лексикон «Бетонведение», разработанной в Харьковском национальном университете строительства и архитектуры. Реализация модели осуществлена как независимое программное обеспечение, которое может быть использовано совместно с электронными терминологическими системами разнообразных предметных областей.

Ключевые слова: терминологические словари, семантическая связь, электронные терминологические системы, программа «Интерпретатор иерархических связей», информационная система «Лексикон «Бетонведение».

Введение

Актуальность исследования. Быстрое развитие науки, как теоретической, так и прикладной, появление новых научных направлений и сфер практической деятельности человека, экспоненциальный рост объема и энтропии информации – эти факторы обуславливают необходимость логического упорядочивания терминологических систем и их реализации конструктивными средствами современных информационных технологий, интеллектуальных систем обработки данных и знаний.

В настоящее время активно ведется обработка существующих терминологических словарей и банков данных с целью создания электронных терминологических систем, важным свойством которых является обеспечение возможности их использования как специалистами в данной предметной области, практической работе, так и в процессе обучения.

Одним из практически реализуемых подходов к решению данной задачи является моделирование представления знаний в электронных терминологических системах путем визуализации одноуровневых семантических связей между терминами.

Целью настоящего исследования является создание конструктивных средств графического представления знаний в электронных терминологических системах (на примере информационной системы «Лексикон «Бетонведение») [1].

Разработка графовой модели направлена на представление семантических связей между элементами информационной терминологической системы.

Важным свойством элемента терминологической системы является то, что каждый термин имеет строгое определение, и для его понимания необходимо знать определения всех терминов, присутствующих в его определении, вплоть до базовых, неопределяемых, понятий.

Связи между элементами системы отражают иерархию понятий в их соотношении друг с другом. Связи между терминами, как и сами термины, определяются отраслевыми специалистами (экспертами) совместно с логиками и лингвистами. Характер установленных связей может быть различным. Автоматизация поиска и формализация связей между терминами возможна при семантическом анализе дефиниций элементов терминологической системы.

Анализ существующих исследований и публикаций. Существующие информационные терминологические системы отличаются по назначению и функциональным возможностям. Модели представления знаний в таких системах, как правило, являются вариациями семантических сетей и онтологий.

Среди наиболее известных электронных терминологических систем выделяют:

1) «WordNet» – семантическая сеть для английского языка, разработанная в Принстонском университете, и выпущенная вместе с сопутствующим программным обеспечением со свободной лицензией [2].

2) «РусГез» – тезаурус русского языка (был создан Центром информационных исследований как инструмент для автоматического индексирования; разрабатывается с 1997 года по настоящее время, 45 тыс. понятий, 107 тыс. слов и выражений, 177 тыс. отношений) [3].

3) «Большая энциклопедия Кирилла и Мефодия (БЭКМ)» – электронная мультимедийная энциклопедия. Содержит сведения по всем основным отраслям знания – из области науки, техники, литературы и искусства, а также наиболее важную историческую, социально-экономическую, географическую информацию по странам мира и регионам России [4].

4) «Викисловарь (Wiktionary)» – тезаурус и многофункциональный многоязычный словарь в котором содержатся грамматические описания, толкования и переводы слов. Кроме того, в статьях мо-

жет отражаться информация об этимологии, фонетических свойствах и семантических связях слов [5].

5) «АвиаОнтология» – информационный ресурс АвиаОнтология является лингвистической онтологией, разработанной для специальной прикладной области. Данная прикладная область включает знания об авиации и специализирована на знаниях об информационных событиях и процессах в авиационной практике. Прежде всего это касается событий и процессов функционирования бортовой аппаратуры, а также экипажа в различных полетных режимах [6].

6) «Gold book» IUPAC – справочник по технической терминологии, применяемой в химии, созданный Международным союзом теоретической и прикладной химии. Проект является терминологической системой, содержащей более 7000 терминов химической предметной области [7].

7) Одной из успешных реализаций информационных терминологических систем является система «Лексикон «Бетонovedение» [1], разработанная специалистами Харьковского национального университета строительства и архитектуры.

Информационная терминологическая система бетонovedения, ориентированная на профессионалов в области бетонovedения, реализована таким образом, что позволяет обеспечить ее применение не только в научной и практической деятельности, но и в учебном процессе. Система представляет собой эргономичную многофункциональную терминологическую систему с возможностью использования как в локальном, так и в удаленном режимах.

Изложение основного материала

В информационной терминологической системе бетонovedения терминологический фундамент на теоретико-множественном уровне представляется как целенаправленная система S вида:

$$S = \langle M, R, P \rangle,$$

где M – конечное множество элементов системы: терминов и их определений, представляющих понятия (знания) предметной области бетонovedения;

R – множество связей между элементами системы – терминами бетонovedения;

P – множество свойств системы, позволяющее достичь заданных целей [1]: – релевантности по отношению к запросу пользователя, актуальности поисковых индексов, надежности, быстродействия обслуживания пользователя.

Рассмотрим множество M . Термин M_i как элемент терминологической системы должен быть:

1. свободен от субъективности жизненного опыта. Недопустимо, чтобы разные исследователи при произнесении одного термина представляли себе разные вещи;

2. однозначен. Недопустимо, чтобы один и тот же термин в рамках одной и той же научной дисциплины описывал в разных случаях разные сущности;

3. должен иметь точно определенную область значений, то есть должно существовать строго определенное множество объектов, описываемых этим термином.

На основе анализа структуры множества R предложена следующая формализация связей между терминами:

$$C_P \rightarrow C_T \rightarrow C_D,$$

где первый список (C_P) содержит термины, которые в своем описании ссылаются на текущий анализируемый термин M_i ; второй список (C_T) содержит термин M_i ; третий список (C_D) содержит термины, на которые встречаются ссылки в описании текущего термина.

Таким образом, при реализации данного упорядочения для каждого термина M_i необходимо создать результирующий файл, содержащий три списка (C_P, C_T, C_D).

Блок-схема алгоритма создания первого (так называемого «родительского») списка приведена на рис. 1.

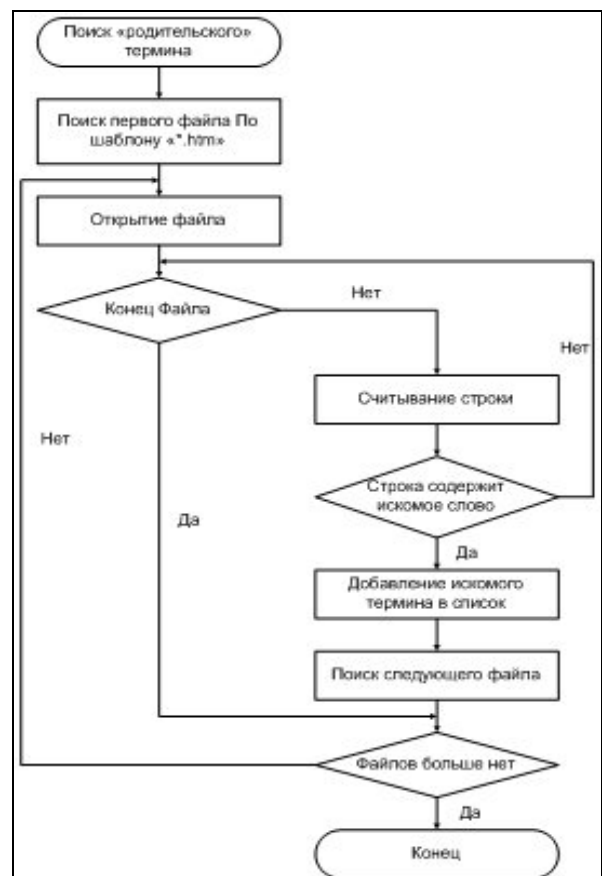


Рис. 1. Блок-схема алгоритма создания «родительского» списка

Блок-схема алгоритма создания третьего («дочернего») списка приведена на рис. 2.

Созданные списки терминов при помощи html-тегов формируют интерактивную таблицу иерархических связей.

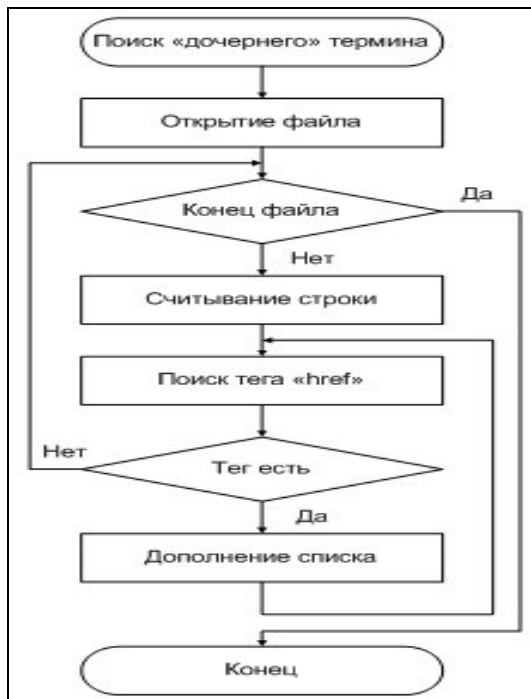


Рис. 2. Блок-схема алгоритма создания «дочернего» списка

Обобщенная блок-схема алгоритма работы программы «Интерпретатор иерархических связей» приведена на рис. 3.



Рис. 3. Обобщенная блок-схема алгоритма работы программы

Программа «Интерпретатор иерархических связей»

В рамках технического задания предложенный алгоритм был реализован как программа «Интерпретатор иерархических связей» на основе использования интегрированной среды разработки приложений Microsoft Visual C++ 6.0.

Использование программного обеспечения предусмотрено в качестве компонента информационной терминологической системы бетоноведения.

Помимо компонента «Интерпретатор», основными компонентами информационной терминологической системы бетоноведения (рис.4) являются база данных и клиентское приложение.

В информационной терминологической системе бетоноведения предусмотрено два режима работы: пользовательский и администраторский.

Программа «Интерпретатор иерархических связей» используется только в режиме администратора, т.к. разработанное программное обеспечение имеет консольное управление.



Рис. 4. Интерфейс информационной системы «Лексикон «Бетоноведение»

В силу того, что «Лексикон «Бетоноведение» является открытой системой и на уровне компонент, и на уровне множества терминов, то применение программы «Интерпретатор иерархических связей» в режиме администратора является необходимым при включении новых терминов в терминологическую базу.

В пользовательском режиме реализован доступ к базе данных, основанный на представлении информации в виде гипертекста и гипермедиа с использованием языка разметки HTML, а также более сложных процедуральных структур, взаимодействующих с пользователем на уровне локальных приложений с достаточно развитой функциональностью.

База данных представляет собой упорядоченный набор данных по терминологии бетона, представленных в виде отдельных файлов и систематизированных каталогов. Такая система организации данных обеспечивает основные требования, которые

выдвигались перед разработкой информационной системы: простота поддержки актуальности информации; возможность быстрой модификации данных; наглядность и понятность организации файлов, каталогов и подкаталогов; отсутствие необходимости использования серверных версий дорогого программного обеспечения для работы информационной системы [8].

Программа «Интерпретатор иерархических связей» запускается из клиентского приложения терминологической системы бетоноведения, при этом исполняемый файл программы находится в одной директории с базой данных терминологической системы бетоноведения.

Пример работы программы «Интерпретатор иерархических связей» для термина «Бетон» приведен на рис. 5.

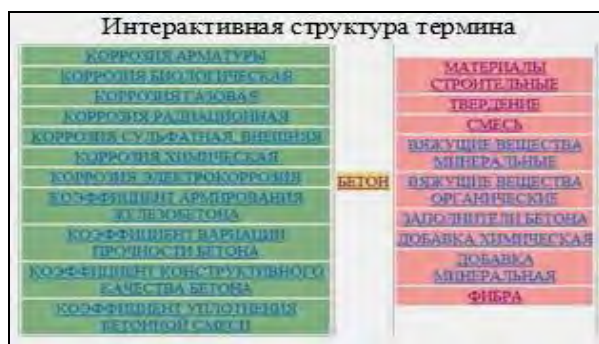


Рис. 5. Пример работы программы «Интерпретатор иерархических связей» для термина «Бетон»

Заклучение

Предложенная в данной работе модель позволяет создать интерактивную структуру семантических связей элементов информационной терминологической системы.

Структура программного продукта является достаточно гибкой и может быть использована в других предметных областях и информационных терминологических системах.

Список литературы

1. Создание информационной системы «Лексикон «Бетоноведение» / А.В. Уиеров-Маршак, М.В. Новожилова, Е.В. Латорец, И.А. Михеев. – К.: УкрНЦ РИТ, 2010. – С. 76-78.
2. WordNet [Электронный ресурс]. Princeton university. – Режим доступа: wordnet.princeton.edu/.
3. Онтологии для автоматической обработки текстов: Описание понятий и лексических значений [Электронный ресурс]. Международная конференция по компьютерной лингвистике «Диалог». – Режим доступа: <http://www.dialog-21.ru/dialog2006/materials/html/Dobrov.htm>.
4. О Проекте [Электронный ресурс]. Мегаэнциклопедия Кирилла И Мефодия. – Режим доступа: <http://www.megabook.ru/About.asp>
5. Викисловарь [Электронный ресурс]. Энциклопедия Википедия. – Режим доступа: <http://ru.wikipedia.org/wiki/Викисловарь>.
6. АвиаОнтология: анализ современного состояния ресурса [Электронный ресурс]. Международная конференция по компьютерной лингвистике «Диалог». – Режим доступа: www.dialog-21.ru/Archive/2004/Lukashevich.htm.
7. Справочник химической терминологии [Электронный ресурс]. Международный союз теоретической и прикладной химии. – Режим доступа: <http://goldbook.iupac.org>.
8. Латорец Е.В. Понятийное единство в бетоноведении – состояние, перспективы / Е.В. Латорец, И.А. Михеев, М.В. Новожилова // Информационная система «Лексикон».

Поступила в редколлегию 5.12.2011

Рецензент: д-р техн. наук, проф. И.В. Гребенник, Харьковский национальный университет радиоэлектроники, Харьков.

МОДЕЛЮВАННЯ ПРЕДСТАВЛЕННЯ ЗНАТЬ В ЕЛЕКТРОННИХ ТЕРМІНОЛОГІЧНИХ СИСТЕМАХ

М.В. Новожилова, І.А. Міхеев, С.О. Ямницький

У даній статті запропоновано модель графічного представлення знань в електронних термінологічних системах. Створення моделі здійснювалося на прикладі інформаційної системи «Лексикон «Бетонознавство», розробленої в Харківському національному університеті будівництва і архітектури. Реалізація моделі запропонована у вигляді незалежного програмного забезпечення, яке може бути використане спільно з електронними термінологічними системами різноманітних предметних областей.

Ключові слова: термінологічні словники, семантичний зв'язок, електронні термінологічні системи, програма «Интерпретатор ієрархічних зв'язків», інформаційна система «Лексикон «Бетонознавство».

DESIGNS OF REPRESENTATION OF KNOWLEDGES IN ELECTRONIC TERMINOLOGY SYSTEMS

M.V. Novozhilova, I.A. Miheev, S.A. Jamnickij

In this article a model of the graphical representation of terminological knowledge in electronic systems is offered. Creating the model was carried out by an information system "Lexicon "Betonovedenie", developed at the Kharkov National University of Engineering and Architecture. Implementation of the model proposed as an independent software that can be used in conjunction with the electronic systems of various terminological domains.

Keywords: terminology dictionaries, semantic connection, electronic terminology systems, the informative system is «Lexicon «Betonovedenie».