

Составляющие семантической технологии проектирования баз знаний

Ивашенко В.П.

Кафедра интеллектуальных информационных технологий, факультет информационных технологий и управления
Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

e-mail: ivashenko@bsuir.by

Аннотация — Рассматривается семантическая технология проектирования баз знаний и лежащие в её основе компоненты, модели и алгоритмы.

Ключевые слова: база знаний, семантические сети, интеллектуальная система

I. ВВЕДЕНИЕ

Для широкого применения интеллектуальных систем, способных повысить качество решения прикладных задач, необходимо большое количество баз знаний. Быстрой разработке достаточного количества баз знаний могло бы способствовать наличие средств разработки интеллектуальных систем, обеспечивающих разработку и проектирование различных компонентов интеллектуальной системы, включая базу знаний. Среди средств, которые могут рассматриваться в качестве основы для разработки баз знаний, можно выделить: оболочки экспертных систем (CLIPS (FuzzyCLIPS, DYNA CLIPS, WxCLIPS), SOAR, OPS83, RT-EXPERT, MIKE, BABYLON, WindExS, ES; ACQUARE, Easy Reasoner, ECLIPSE, EXSYS Professional, SIMER+MIR, AT ТЕХНОЛОГИЯ, CAKE v2.0) 0; инструментальные пакеты для разработки экспертных систем (G2, ART, KEE, Knowledge KRAFT); системы, ориентированные на обработку онтологий 0 – Protégé, WebOnto, OntoEdit, WebODE, OilEd, OntoLingua.

Достоинствами приведенных средств являются: поддержка представления знаний различного вида различными моделями представления знаний в рамках одной системы; наличие средств визуального проектирования баз знаний; наличие средств верификации базы знаний, включая проверку на непротиворечивость; возможность монотонного расширения базы знаний, наличие средств интеграции баз знаний; наличие средств поддержки обмена данными с внешней средой, включая средства обмена данными в реальном времени. Однако для всех указанных средств характерны следующие недостатки: в силу различных ограничений велики сроки разработки баз знаний (отсутствие развитых технологий разработки); узок круг инженеров баз знаний (из-за высоких стартовых требований к разработчику) – от разработчика требуется владение специальными знаниями по моделям и языкам представления знаний; не полностью решен вопрос интеграции баз знаний.

В основе предлагаемой технологии проектирования баз знаний лежат следующие положения: модульное проектирование баз знаний, интеллектуализация средств поддержки проектирования баз знаний и семантическое представление знаний. Эта технология включает: унифицированную модель баз знаний и модель ip-компонентов (повторно используемых фрагментов) баз знаний; среду проектирования баз знаний; методику проектирования и интеграции баз знаний.

II. УНИФИЦИРОВАННАЯ МОДЕЛЬ БАЗ ЗНАНИЙ И МОДЕЛЬ IP-КОМПОНЕНТОВ

Обеспечение унифицированного представления знаний, которые бы отвечали некоторым общим критериям качества в рамках унифицированной модели баз знаний, требует разработки модели представления знаний и выявления в её рамках единой системы мер и критериев оценки качества представленных знаний, которые бы работали для любых видов знаний и не зависели бы от особенностей программной реализации.

Унифицированная модель баз знаний основывается на унифицированной модели представления знаний, которая является частным случаем такой модели представления знаний, как семантические сети и задаётся семейством совместимых sc-языков, использующих унифицированный способ семантического кодирования Semantic Computer code (SC-код) и поддерживающих представление знаний различного вида 0, 0. Особенности SC-кода являются: простой алфавит, содержащий узлы и дуги, простой синтаксис, базовая теоретико-множественная интерпретация. На множестве sc-языков определены отношения sc-подъязыка и трансляции (кодировки). Подъязык, являющийся пересечением этих языков, рассматривается как интегрированный sc-язык представления знаний. Основным принципом построения sc-языков является представление понятий, соответствующих основным классам объектов, описываемых sc-языком, и отношений между этими объектами ключевыми узлами такого sc-языка: каждому sc-языку однозначно сопоставляется конечное множество ключевых узлов (элементов) этого языка. Каждый ключевой узел задаёт ограничения на собственную семантическую окрестность в информационных конструкциях (текстах) этого языка. Характеристиками sc-языка являются: мощность множества ключевых узлов sc-языка; семейство множеств собственных семантических окрестностей ключевых элементов языка; наличие функциональных зависимостей между собственными окрестностями множеств ключевых элементов sc-языка; соотношение алгоритмических сложностей поиска или вычисления элементов собственных семантических окрестностей на основании существующих зависимостей. Построена семантическая онтология существующих языков и моделей представления знаний, в которой указано место унифицированной модели представления знаний и её соотношение с существующими моделями представления знаний.

Унифицированная модель баз знаний описывает базы знаний специального вида, в которых знания представлены с использованием унифицированной модели представления знаний [7]. В унифицированной модели баз знаний база знаний – связанная структурированная информационная конструкция, структура которой состоит не менее чем из одного раздела, каждому из которых принадлежит свой, описываемый в этом разделе (ключевой) элемент этой конструкции, причем в описании хотя бы одного

ключевого элемента присутствует его внешнее обозначение [7]. Между базами знаний определены отношения содержательного и структурного включения одной базы знаний в другую. На множестве баз знаний заданы операции содержательной проекции и темпорализации [7]. Соответствие интеграции каждой паре баз знаний сопоставляет однозначное отображение множества всех знаков, формирующих эти базы знаний, на множество всех знаков результирующей базы знаний. Путём интеграции осуществляется переход от некоторой исходной базы знаний к требуемой оптимизированной базе знаний, которая имеет более предпочтительную величину значения функции критерия качества [7].

В зависимости от соотношения значений собственных характеристик [7] и других признаков база знаний может быть классифицирована как терминологический словарь, таксономия, модель, тезаурус (справочник), теория или прикладная база знаний. Эта классификация, как и классификация фрагментов баз знаний, является основой модели ip-компонентов баз знаний.

Модель ip-компонентов баз знаний описывает в соответствии с классификацией классы специфицируемых баз знаний и их фрагментов и отношения между ними и их частями. Спецификация базы знаний включает указание класса (типа) базы знаний и описание её количественных и качественных характеристик: состав, сертификат, задачно-ориентированный сборник тестовых вопросов, требования полноты и безошибочности базы знаний, информацию о разработчиках, условиях распространения и информацию для сопровождения.

В рамках перечисленных моделей строятся семантические модели и онтологии, как прикладного плана, так и ориентированные на применение в рамках рассматриваемой технологии: семантические модели **верификации** и **интеграции** баз знаний, онтологии языков представления знаний и языков представления исходных текстов баз знаний (SCn) и другие.

III. СРЕДА ПРОЕКТИРОВАНИЯ БАЗ ЗНАНИЙ

Среда проектирования баз знаний состоит из двух составляющих: инструментальных средств проектирования баз знаний и библиотеки ip-компонентов баз знаний. Инструментальные средства основываются на унифицированной модели баз знаний и состоят из базы знаний, машины обработки знаний и пользовательского интерфейса. Библиотека ip-компонентов баз знаний основывается на модели ip-компонентов баз знаний и также состоит из базы знаний, машины обработки знаний и пользовательского интерфейса. Принципы и порядок использования среды проектирования баз знаний регламентируются методикой проектирования и интеграции баз знаний, входящей в состав технологии.

База знаний библиотеки ip-компонентов баз знаний содержит каталогизированный набор ip-компонентов, операциями машины обработки знаний библиотеки ip-компонентов являются операции поиска и редактирования ip-компонентов в этом каталоге, пользовательский интерфейс поддерживает соответствующие команды запуска этих операций. База знаний средств проектирования баз знаний содержит определяемую унифицированной моделью баз знаний метаинформацию о проектируемой базе знаний. Операции машины обработки средств проектирования баз знаний реализуются на основе алгоритмов поддержки решения задач редактирования, интеграции, верификации, тестирования и оценки качества баз знаний, способных осуществлять поиск и выявлять избыточность в базах знаний. Пользовательский интерфейс средств проектирования баз знаний поддерживает необходимый набор соответствующих классов пользовательских команд и использует для общения с пользователем специальные языки (SCg, SCn), обеспечивающие наглядность, совместимость с современными web-технологиями и упрощающие наполнение, анализ и редактирование базы знаний.

IV. ЗАКЛЮЧЕНИЕ

Особенностями и достоинствами предложенной технологии проектирования баз знаний являются: унифицированное представление баз знаний и их фрагментов; перенос акцента от традиционной инженерии баз знаний к интеграции баз знаний 0, т.е. к сборке баз знаний из крупных модулей, основывающейся на унифицированной модели баз знаний; обеспечение совместимости разрабатываемых баз знаний; использование визуальных методов проектирования.

- [1] Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский – СПб: Питер, 2000.
- [2] Sowa J.F. Conceptual Graphs / Sowa J.F., F. van Harmelen, V. Lifschitz, B. Porter // *Handbook of Knowledge Representation*. – 2008. – P. 213-237.
- [3] Представление и обработка знаний в графодинамических ассоциативных машинах /В.В. Голенков [и др.]; под ред. В.В. Голенкова. – Минск : БГУИР, 2001.
- [4] Представление нейронных сетей и систем продукции в однородных семантических сетях. / В.П. Ивашенко // *Известия Белорусской инженерной академии*. – 2003. – №1(15)/1. – С.184-188.
- [5] Применение однородных семантических сетей для представления знаний о нестационарных предметных областях. / В.П. Ивашенко // *Известия Белорусской инженерной академии*. – 2004. – №1(17)/3. – С.77-80.
- [6] Интеграция знаний в информационных системах. / Н.А. Гулякина, В.П. Ивашенко // *Доклады БГУИР*. – 2004. – №6. – С. 113-119.
- [7] Семантические модели баз знаний / В.П. Ивашенко *Информационные системы и технологии (IST'2009): материалы V Междунар. конф.-форума в 2-х ч. Ч.2* – Минск: А.Н.Вараксин, 2009. - с.125-128.