

ТЕХНОЛОГИЯ ПРОЕКТИРОВАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННО-ПОИСКОВЫХ МАШИН

В.А. Житко, К.А. Уваров

Белорусский государственный университет информатики и радиоэлектроники, Минск, Республика Беларусь, zhitko.vladimir@gmail.com.

Abstract. This paper explains some questions about semantic technology designing search engine for an intelligent system, based on OSTIS technology. Geometry help system is showed as an example of intelligent help systems with such search engine based on OSTIS technology. Semantic question language, ip-components shared library are also described.

Основной задачей интеллектуальных справочных систем (ИСС) является предоставление информации пользователю по его запросу. Основными средствами, при этом, являются операции навигации и поиска по семантическим сетям. Главной особенностью таких средств является то, что они могут использовать различные подходы для поиска нужной информации (подход с поиском по шаблону или интеллектуальным поиском, нейросетевой подход, эволюционные алгоритмы и др.), но при этом должны быть интегрированы в одну систему и использовать один источник данных. Все это приводит к необходимости создания общей технологии проектирования подобных средств.

Целью данной работы является создание семантической технологии компонентного проектирования средств навигации и поиска в семантических сетях. Такая технология основывается на семантической технологии компонентного проектирования машин обработки знаний и баз знаний [1], разрабатываемой в рамках открытого проекта OSTIS.

Одна из главных проблем интеллектуальных технологий связана с крайне длительными сроками проектирования интеллектуальных систем. Это, в первую очередь, связано с отсутствием всесторонней технологии проектирования интеллектуальных систем, охватывающую пользовательский интерфейс, базы знаний и средства обработки знаний. Такой технологией является открытая семантическая технология проектирования интеллектуальных систем [1]. Одной из составных её частей является семантическая технология компонентного проектирования средств навигации и поиска в семантических сетях. Т.к. технология предполагаем компонентных подход в проектировании, то это позволит значительно сократить общее время проектирования интеллектуальных систем, за счет использования готовых модулей. Предметно независимые, совместимые модули будем называть ip-компонентами. К ip-компонентам могут относиться фрагменты баз знаний, операции поиска и навигации, элементы пользовательского интерфейса и даже наборы некоторых компонентов взаимосвязанных между собой. Все подобные компоненты хранятся в библиотеке совместимых ip-компонентов. Такая библиотека осуществляет хранение ip-компонентов, предоставляет разработчику доступ к информации о хранимых компонентах, а также средства поиска нужных компонентов.

Семантический язык вопросов является одним из важнейших компонентов любой интеллектуальной системы основанной на технологии OSTIS. Язык вопросов является семантическим языком представления знаний, текстами которого являются формальные записи вопросов и ответов. В естественно-языковом интерфейсе семантический язык вопросов выступает посредником между естественно-языковым пользовательским интерфейсом и интеллектуальной предметной системой.

На сегодняшний день существует множество различных языков запросов (языки запросов к базам данных, к данным, представленным по модели RDF, информационным системам, особенно к информационно-поисковым системам и др.):

- язык запросов XQuery для обработки данных в формате XML;
- язык запросов SQL к реляционным базам данных;
- язык запросов SPARQL к данным, представленным по модели RDF;
- и др.

Для примера рассмотрим пример на языке SPARQL: поиск всех примеров каждого класса понятия треугольника.

Запрос на SPARQL будет выглядеть следующим образом:

–*SELECT \$x WHERE { <треугольник> rel:decomp [rel:example \$x] }*

Запрос, используя изоморфный поиск, представлен на рисунке 1.

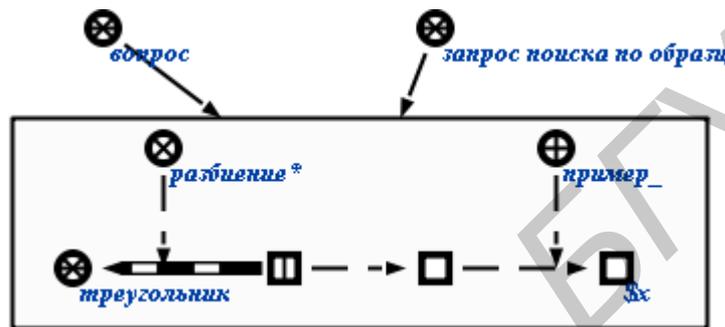


Рисунок 1 - Запрос изоморфного поиска

Из приведённого примера видно, что семантический язык вопросов обладает всеми функциональными возможностями языка запросов SPARQL, но в то же время обладает следующими достоинствами:

- позволяет в «краткой» форме формулировать известные системе вопросы (вопросы, описанные в базе знаний системы);
- расширять список возможных вопросов (семантический язык вопросов включает в себя средства описания и классификации вопросов)
- указывать дополнительную информацию о вопросе (автор вопроса, предметная область вопроса и пр.).

В то время как использование языка запросов SPARQL требует от пользователя знания внутренней организации интеллектуальной системы, что требует более высокого начального уровня подготовки пользователя, наличие упрощенной формы задания вопросов в семантическом языке вопросов позволяет пользователю общаться с интеллектуальной системой без знания внутренней структуры базы знаний.

Семантическая мощность языка вопросов в первую очередь зависит от многообразия типов вопросов. Т.к. для каждой предметной области могут существовать специфические только для неё типы вопросов, то в языке вопросов существует возможность расширять типологию и классификацию вопросов. Это позволяет создавать отдельные вопросы для специфических задач в предметных областях. Для этого достаточно описать новый вопрос и включить его в общую классификацию вопросов.

Классификация вопросов позволит ускорить выбор подходящей операции поиска ответа, т.к. операции достаточно проверить входит ли заданный вопрос во множество тех вопросов, на которые операция может найти ответ. Классификация вопросов может осуществляться по различным признакам.

Классификация вопросов по структуре:

- атомарные – вопросы, представленные в системе как единый запрос;
- неатомарные – вопросы, представленные в системе как множество взаимосвязанных атомарных вопросов.

Классификация по полноте ответа:

- открытые – вопросы, предполагающие возможность неограниченного количества верных ответов, примером может служить вопрос «привести пример»;
- закрытые – вопросы, предполагающие ограниченное количество ответов, примером может служить вопрос «верно ли».

По отношению к предметной области:

- предметно-независимые – вопросы, не зависящие от предметной области;
- предметно-зависимые – вопросы, определенные только в конкретной предметной области.

В системе также существуют операции, реагирующие на неуспешную попытку найти ответ на вопрос. Такие операции осуществляют обработку и преобразование вопроса:

- декомпозиция вопроса на подвопросы;
- объединение нескольких подвопросов в один вопрос;
- упрощение вопроса;
- корректировка и уточнение вопроса;
- и др.

После этого вопрос вновь актуализируется и происходит поиск ответа на него.

Вопросы можно классифицировать по структуре и элементному составу ремы вопроса [2]:

- однородные однообъектные – вопросы, рема которых связана с одним из объектов определенного множества;
 - Пример: «Какой из приведенных треугольников имеет наибольшую площадь?»
- однородные многообъектные – вопросы, рема которых связана с несколькими объектами определенного множества;
 - Пример: «Какие из приведенных треугольников имеют прямоугольный угол?»
- многородовые однообъектные – вопросы, рема которых связана множеством объектов, имеющих различную структуру;
 - Пример: «Какова высота и площадь наибольшего из приведенных треугольников?»
- многородовые многообъектные – вопросы, рема которых связана с некоторым числом объектов, принадлежащих различным множествам.
 - Пример: «Каковы площади и высоты тех, из приведенных, треугольников имеющих прямой угол?»

Приведенная классификация может быть полезна при анализе качества полученных ответов, а так же их правильности. Например, если в ответе присутствует множество однотипных объектов на однородный однотипный вопрос, то можно с уверенностью сказать что ответ не является корректным.

Литература

1. Open Semantic Technology for Intelligent Systems. [Электронный ресурс]. - 2011. - Режим доступа: <http://www.ostis.net/>. - Дата доступа: 01.04.2011
2. Федоров Б.И. Проблема классификации вопросов в эротетической логике / Федоров Б.И. – Вестник Санкт-Петербургского университета. С.-Петербург: Типография Издательства СПбГУ, 2008