

Технология проектирования естественно-языковых интерфейсов вопросно-ответных систем

Житко В.А.; Заболеева-Зотова А.В.

Каф. ИИТ, ФИТиУ

Белорусский государственный университет информатики и радиоэлектроники

Минск, Беларусь

e-mail: zhitko.vladimir@gmail.com

Аннотация — Дано описание технологии проектирования естественно-языковых пользовательских интерфейсов включающих в себя взаимодействие, используя естественные тексты, формальные тексты. Используемая технология развивается в рамках открытого проекта OSTIS. Рассматривается библиотека компонентов проектирования естественно-языкового интерфейса, как её пополнение сторонними компонентами, так и создание новых компонентов.

Ключевые слова: анализ естественно-языковых текстов; естественно-языковой интерфейс; семантическая технология; технология проектирования

I. ВВЕДЕНИЕ

В связи с динамичным развитием и распространением компьютерных систем возникает необходимость в снижении расходов на подготовку новых пользователей. Привлекательно в этом случае выглядит использование привычного для пользователя языка для организации его диалога с компьютерной системой. Такая возможность реализуется средствами естественно-языкового пользовательского интерфейса, обладающего рядом преимуществ: минимальной подготовкой пользователя, необходимой для работы с системой, простотой и высокой скоростью задания произвольных запросов к пользовательскому интерфейсу и высоким уровнем модели предметной области.

Для естественно-языкового пользовательского интерфейса вопросно-ответных систем возможно использование ограниченного набора лексики и грамматики, без серьёзного ущерба функциональности вопросно-ответной системы. Ограниченный естественный язык – это подмножество естественного языка, текст на котором, без каких-либо усилий, воспринимается носителем исходного естественного языка, а также не требует длительного изучения для приобретения навыков составления текстов на этом языке, однако обладает сокращённым набором лексики и грамматики. Это позволит снизить время обработки естественно-языковых конструкций, а также устранить часть лингвистической неоднозначности.

II. ОБЩАЯ МОДЕЛЬ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ИНТЕРФЕЙСА

Естественно-языковой пользовательский интерфейс, в рамках разрабатываемой технологии, рассматривается как специализированная интеллектуальная система, обеспечивающая диалог между прикладной вопросно-ответной системой и пользователем. Являясь интеллектуальной системой, естественно-языковой интерфейс включает в себя базу знаний, машину обработки знаний и пользовательский интерфейс.

Данный функционал естественно-языкового интерфейса интеллектуальной справочной системы определяет структуру интеллектуальной системы естественно-языкового интерфейса. Структура

естественно-языкового интерфейса должна состоять из следующих составляющих:

- пользовательский интерфейс, посредством которого происходит ввод сообщений пользователем и вывод ответа системы пользователю;
- трансляторы естественно-языковых запросов на sc-язык вопросов;
- трансляторы sc-конструкций на естественный язык. Общая структура естественно-языкового пользовательского интерфейса имеет следующий вид:
- база знаний;
 - лингвистическая база знаний;
 - предметная база знаний;
 - база знаний естественно-языкового интерфейса;
- машина обработки знаний;
 - транслятор sc-конструкций в текстовое представление;
 - транслятор вопросов на семантический язык вопросов;
- пользовательский интерфейс;
 - редакторы и просмотрщики текстов.

III. БАЗА ЗНАНИЙ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ИНТЕРФЕЙСА

База знаний естественно-языкового пользовательского интерфейса включает в себя лингвистическую и предметную базу знаний. Предметная база знаний используется вопросно-ответной системой для поиска и генерации ответов по запросам пользователя. Естественно-языковой интерфейс так же использует эту базу знаний для решения ряда задач, связанных с разрешением неоднозначности естественного языка, т.к. предметная база знаний однозначно задаёт контекст диалога пользователя с вопросно-ответной системой [1].

Лингвистическая база знаний включает в себя формальное описание используемого естественного языка, привязку лексики к предметной базе знаний, спецификации семантических языков.

IV. СЕМАНТИЧЕСКИЙ ЯЗЫК ВОПРОСОВ

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.

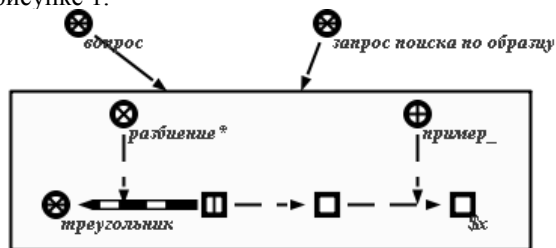


Рис. 1. Запрос изоморфного поиска

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 мощь 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.

Машина обработки знаний естественно-языкового интерфейса включает в себя операции,

обеспечивающие обработку различных лингвистических конструкций, перевод внешних языков на семантические языки интеллектуальной системы и обратно.

Все компоненты машины обработки знаний естественно-языкового интерфейса можно разделить на трансляторы и анализаторы [2]. Задачей трансляторов является перевод знаний из одного языка представления знаний в другой, к примеру, транслятор фактографических знаний по предметной области в текст на естественном языке. Задачей анализаторов является анализ фрагментов знаний и выявления ранее неизвестных фактов, к примеру, анализатор запроса пользователя направленный на поиск цели и задачи вопроса.

На этапе семантического происходит соотношение лингвистической конструкции и конструкций в памяти системы, для выявления отношения соответствия, эквивалентности и т.д.

Результатом семантического анализа является конструкция запроса к системе. Сложность в получении такого результата заключается в семантической неполноте исходного естественно-языкового вопроса.

Вопросы, сгенерированные естественно-языковым интерфейсом, обрабатывает универсальный решатель вопросно-ответных систем, являющийся частью вопросно-ответной системы по соответствующей предметной области. Стоит заметить, что универсальный решатель использует знания и лингвистической базы знаний, таким образом, система может отвечать на вопросы, связанные с используемым естественным языком.

V. ЗАКЛЮЧЕНИЕ

Семантическая технология компонентного проектирования естественно-языковых интерфейсов вопросно-ответных систем предоставляет единую основу для анализа естественно-языковых текстов, начиная с ввода текста, морфологического, лексического, семантического анализа и генерации текста. Так же, позволяет устанавливать отношения между эквивалентными лингвистическими конструкциями и структурой внутренней памяти системы.

Возможность интеграции сторонних разработок и проектов в качестве внешних ip-компонентов позволяет производить интеграцию различных подходов и методов в рамках одного проекта, что позволяет использовать их лучшие стороны.

[1] Open Semantic Technology for Intelligent Systems. [Электронный ресурс]. - 2011. - Режим доступа: <http://www.ostis.net/>. - Дата доступа: 01.04.2011

[2] Byron Long, Natural Language as an Interface Style / Byron Long // Dynamic Graphics Project Department of Computer Science University of Toronto, 1994.