

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ПОСТРОЕНИЯ ЗАПРОСОВ К БАЗЕ ЗНАНИЙ

Рассматривается разработка модуля построения запросов к базе знаний на естественном языке.

ВВЕДЕНИЕ

Естественный язык в данном контексте – предпочтительный способ общения человека с компьютером. Пользователя информационной системы чаще всего интересует не функционал и возможности интерфейса, а информация необходимая ему. Актуальность разработки обусловлена большим спросом на такие системы, т.к. в индустрии информационных технологий сформировался класс программных систем, в которых человеку проще сформулировать свою цель на естественном человеческом языке.

I. ЦЕЛЬ И ЗАДАЧИ ПРОЕКТА

Цель проекта состоит в разработке автоматизированной системы построения запросов к базе знаний.

Достижение указанной цели связывается с решением следующих задач:

- анализ существующих моделей для обработки предложений естественного языка;
- разработка подхода к семантическому анализу естественного языкового запроса;
- разработка системы перевода естественных языковых запросов к базе знаний в SQL-запросы.

II. АНАЛОГИ

Существует несколько коммерческих и промышленных систем естественного языка для общения с базой знаний, которые успешно применяются на практике. Одним из таких коммерческих продуктов является система InBASE. Система основана на семантически-ориентированном анализе.

В полном соответствии с особенностями семантически-ориентированной парадигмы InBASE позволяет строить естественно-языковые интерфейсы ко многим языкам. Для русского и для английского используется один и тот же JI-процессор. Интересной особенностью InBASE является возможность моделирования предметной области на естественном языке: с помощью класса словарных статей «Толкование»

смысл слова можно описать простой фразой. Это позволяет настраивать естественно-языковые интерфейсы людям, не обладающим навыками инженеров знаний. Основным недостатком данной системы является то, что кортежи базы данных продублированы в словарях отдельных файлов. В базах данных больших объемов этот недостаток может стать проблемой.

Также заслуживает внимания система English Query. Система English Query от Microsoft основана на синтаксически-ориентированных шаблонах, связываемых с моделью предметной области, и через нее - со схемой базы данных. Этот продукт позволяет строить естественно-языковые интерфейсы только для английского языка и работает только с Microsoft SQL Server, в этом смысле это лишь утилита, поставляемая с SQL-сервером, именно так она и позиционируется. К сожалению, она не совмещена с пониманием запросов к базе данных.

III. ПОДХОДЫ К РЕШЕНИЮ ЗАДАЧИ ПОНИМАНИЯ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ЗАПРОСОВ

В основе модуля лежит модель знаний, основанная на семантической сети, которая укладывается в базу данных. Любой естественно-языковой запрос к такой базе можно разбить на синтаксические конструкции. Синтаксическое представление запроса строится на основе подлежащего, сказуемого, прямого дополнения и т.п., которые определяются с помощью морфологических характеристик. Каждое слово запроса разбивается по частям речи (аналогично разбиению предложения на триады, предложенные Мартыновым В.В.). Далее каждое слово запроса заносится в базу, с которой мы работаем с помощью обычного языка SQL, извлекая нужный ответ.

Таким образом, данный модуль позволяет обрабатывать естественно-языковые запросы в пределах определенной темы и извлекать релевантные по смыслу ответы. Данную систему можно применять в качестве поисковика.

Тоболич Виктория Николаевна, студентка кафедры информационных технологий автоматизированных систем БГУИР, victoriatobolich@gmail.com.

Руководитель дипломного проекта: Герман Юлия Олеговна, ассистент кафедры информационных технологий автоматизированных систем БГУИР, jgerman@bsuir.by.