

СЕМАНТИЧЕСКАЯ МОДЕЛЬ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ОБЩЕНИЯ В ИНТЕЛЛЕКТУАЛЬНЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМАХ

В.А. Житко, С.Г. Барбук

Белорусский государственный университет информатики и радиоэлектроники,
кафедра интеллектуальных информационных технологий

П.Бровки, 6, г. Минск, Республика Беларусь
+375-017-293-80-92; e-mail: minsk.nes@gmail.com

Рассмотрена семантическая модель естественно-языкового общения в интеллектуальных вопросно-ответных системах, построенных на базе универсального семантического кода

Вопросно-ответная система, естественно-языковой интерфейс, интеллектуальная система, семантическая технология.

1 ВВЕДЕНИЕ

Целью данной работы является разработка формальной модели естественно-языкового интерфейса интеллектуальных систем, построенных на семантических сетях (sc-технология) [1].

В основе предлагаемой модели лежит модель пользовательских интерфейсов, основу которой составляет унифицированное кодирование информации при помощи SC-кода (Semantic Code) [2]. Естественно-языковой интерфейс рассматривается как специализированная интеллектуальная система, основанная на SC-коде (интеллектуальные системы, основанные на SC-коде, называются sc-системами).

Естественно-языковой интерфейс включает в себя транслятор естественно-языковых запросов в SC-код, транслятор sc-конструкций в естественно-языковые сообщения, синтез речевых сообщений по тексту, параллельная их визуализация посредством «говорящей головы», а также инициализация отдельных команд посредством голосового ввода.

Как и любая sc-система, естественно-языковой интерфейс включает в себя базу знаний и машину обработки знаний.

База знаний естественно-языкового интерфейса включает в себя полное либо частичное описание используемого естественного языка, что может включать в себя:

- фонетические правила. Описывают звуковое строение языка (слоги, звуко сочетания, закономерности соединения звуков в речевую цепочку);
- графические правила. Описывают графическую систему знаков языка;
- семантические правила. Описывают систему

значений единиц языка;

- лексические правила. Описывают индивидуальные свойства отдельных языковых знаков;
- грамматические правила. Описывают общие правила комбинации отдельных языковых знаков, их употребление и понимание. В них включены морфологические (грамматические свойства слов) и синтаксические (грамматические свойства словосочетаний и предложений) правила;
- различные лингвистические словари, представленные в семантических сетях.

Использование семантических сетей для «хранения» словарей естественных языков позволяет избавиться от избыточности данных, а также объединять словари в одну базу знаний. Такая база знаний представляет собой отдельный ip-компонент, который будет востребован в различных интеллектуальных вопросно-ответных sc-системах.

База знаний естественно-языкового интерфейса содержит спецификации используемых sc-языков описания морфологической, синтаксической и семантической составляющей естественно-языкового текста, а также sc-язык описания связи естественно-языкового текста и внутренних sc-конструкций системы.

Машина обработки знаний естественно-языкового интерфейса включает в себя операции морфологического, синтаксического и семантического разбора естественно-языковых текстов, а также операции по генерации естественно-языкового текста из sc-конструкций. Каждая такая операция представляет собой отдельную ip-компоненту, что позволяет с минимальными затратами адаптировать систему естественно-языкового интерфейса под определенные критерии, такие как быстродействие, качество анализа или генерации текстов, используемые стили текстов и т.д.

2 ОПИСАНИЕ МОДЕЛИ

Введенный в систему текст разбивается по предложениям. Для каждого предложения строится sc-конструкция, отображающая данное предложение в память системы. Все введенные предложения анализируются в порядке следования, предшествующие результаты

анализа используются для анализа текущего предложения.

Анализируемое предложение представляет собой следующую sc-конструкцию:

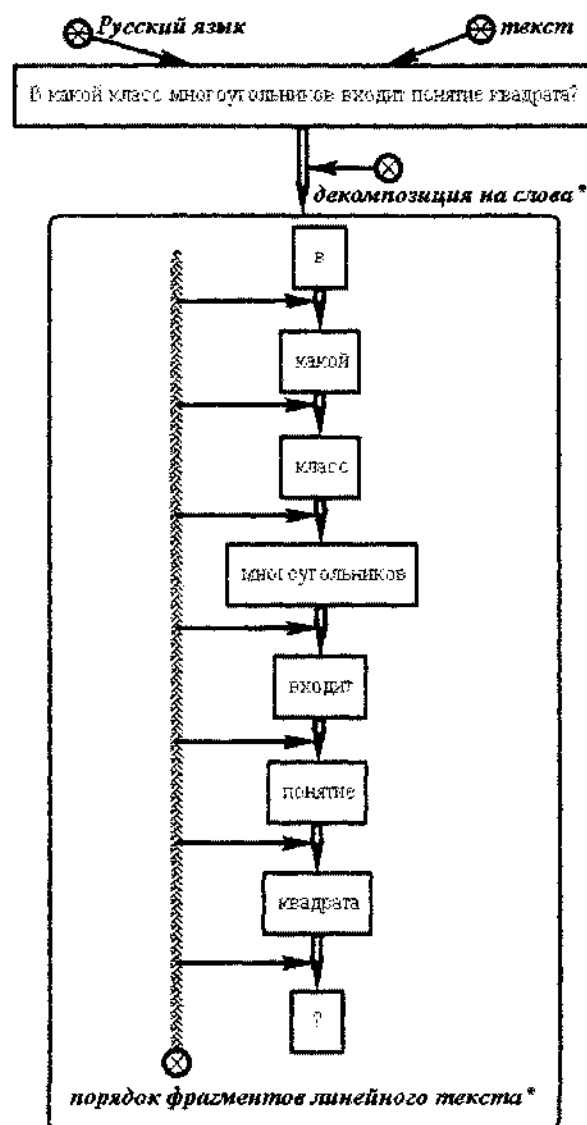


Рис.1. Пример введенного предложения

На этапе морфологического анализа для каждой словоформы в sc-конструкции предложения строятся отношения в соответствии значений ее грамматических категорий, т.е. совокупности грамматических значений (род, падеж, склонение и т.д.) (рис.2). Грамматические категории могут присутствовать в базе знаний естественно-языкового интерфейса явно (словоформа присутствует в словаре с описанием ее грамматических категорий) либо выявляться на основании морфологических правил используемого естественного языка. Такой подход позволяет разработчику при проектировании естественно-языкового интерфейса выбирать между быстродействием и объемом проектируемой базы знаний.

На этапе лексического анализа строятся отношения синтаксических связей внутри предложения, выделяются

главные и второстепенные члены предложения, выявляется тип предложения и т.д. Такой анализ протекает поэтапно, "наращивая" sc-конструкцию предложения, используя информацию, полученную на этапе морфологического анализа.

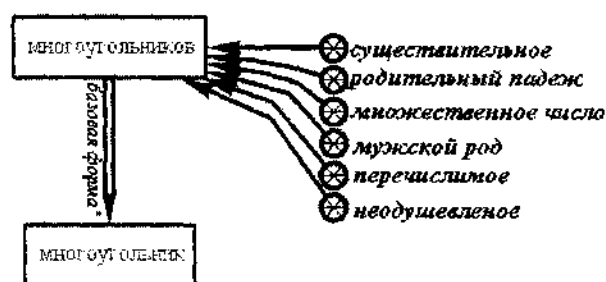


Рис.2. Пример морфологического разбора

Результатом этапа лексического анализа является sc-конструкция лексического разбора предложения в памяти естественно-языкового интерфейса.

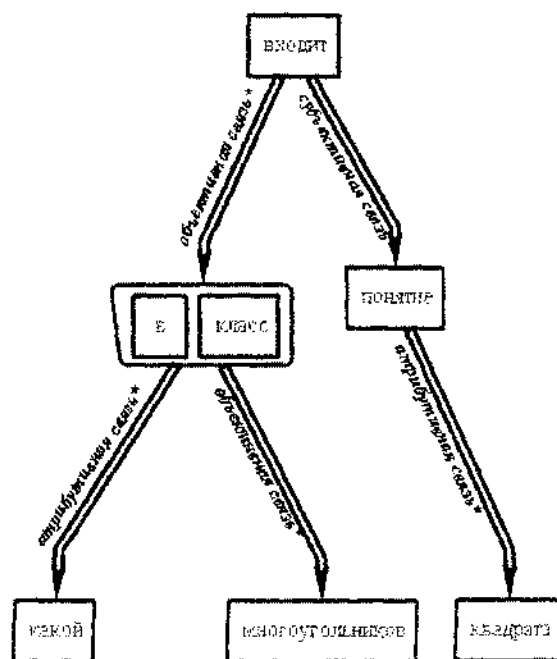


Рис.3. Пример sc-конструкции дерева лексического разбора

На этом этапе используются лексические и синтаксические правила используемого языка. В общем случае такие правила описывают критерии существования синтаксических связей в предложении. При анализе предложения проверяется соответствие таким критериям и, в зависимости от результата, строятся соответствующие связи в предложении. К примеру, правило определения подлежащего будет содержать условие: именительный падеж имя существительное или местоимение, либо неопределенная форма глагола. Данное правило входит в состав правила выделения грамматической основы предложения. Все подобные правила записываются на SC-языке в унифицированном виде, как показано на рис.4.

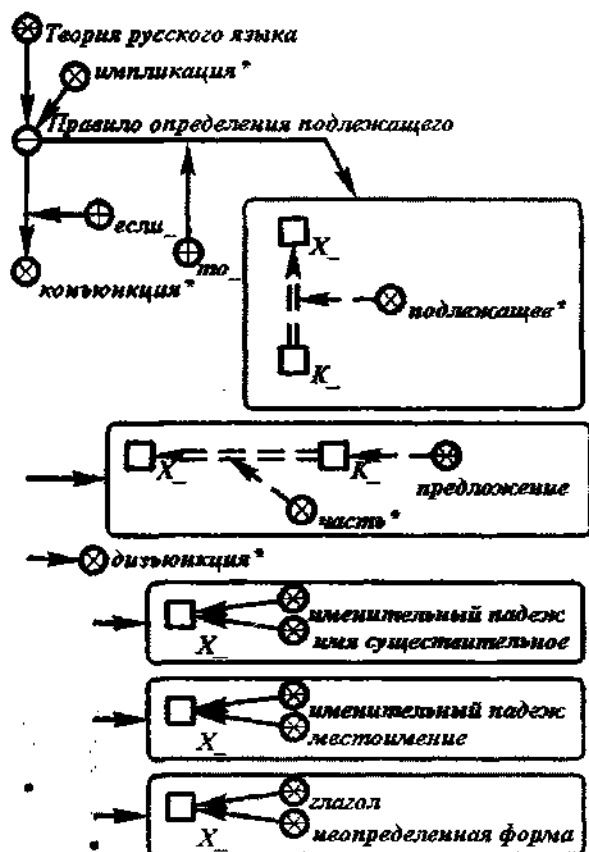


Рис.4. Пример записи правила

На этапе семантического анализа строятся отношения эквивалентности элементов предложения и узлов в памяти системы. На основании, выявленного на этапе лексического анализа, тип предложения влияет на тип запроса к системе (ввод данных, запрос на выполнение, вопрос к системе и т.д.). Например, если в конце предложения стоит пунктуатор вопросительный знак, то будет генерироваться вопрос к системе.

Данный этап является наиболее сложным этапом обработки естественно-языкового текста. В рамках всетехнологии, семантический анализ представляет собой обработку семантической сети, отражающую результат анализа естественно-языкового текста на всех предыдущих этапах, а также присутствующих в системе знаний о языке и предметной области основной системы. На этом этапе происходит соотношение лингвистической конструкции (совокупность всех вс-конструкций, построенных на предыдущих этапах обработки естественно-языкового текста) и конструкций в памяти системы (знания о используемом языке, а также знания системы о предметной области), для выявления отношения соответствия, эквивалентности и т.д.

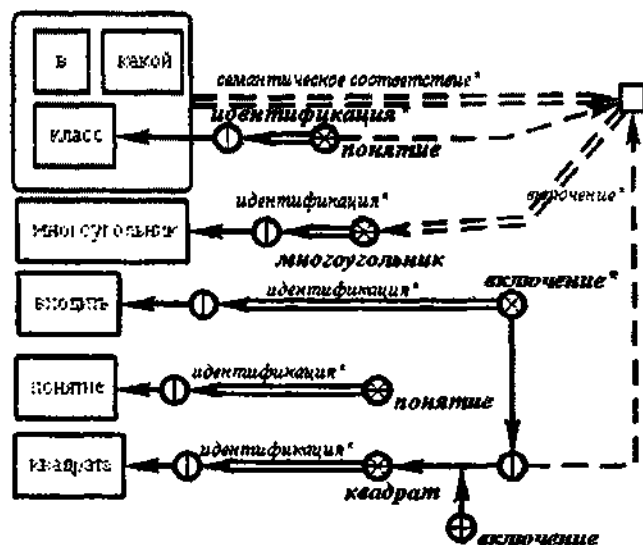


Рис.5. Пример результата семантического анализа

Как видно из рис.5, на этапе семантического анализа происходит не только построение связей лингвистических конструкций и внутренней памяти системы, но и генерация новых для системы конструкций и построение отношений между элементами памяти системы.

Результатом семантического анализа является конструкция запроса к системе. Сложность в получении такого результата заключается в семантической неполноте исходного естественно-языкового вопроса. Например, вопрос «В какой класс многоугольников входит понятие квадрата?» в развернутом виде будет выглядеть следующим образом: «Какое понятие является подмножеством понятия многоугольника и надмножеством понятия квадрата?». Из примера видно, что в процессе развертки произошли следующие подстановки: высказывание «входит» было развернуто в «является подмножеством», «квадрата» - «понятие квадрата», «класс многоугольников» - «понятие, являющееся подмножеством понятия многоугольника», «многоугольник» - «понятие многоугольника». Такие преобразования протекают на этапе семантического анализа и кодируются на SC-языке, как показано на рис.5.

Таким образом, вопрос, сгенерированный естественно-языковым интерфейсом, будет выглядеть, как на рис.6. Такие вопросы обрабатывает универсальный решатель вопросно-ответных систем, являющийся частью основной интеллектуальной системы по соответствующей предметной области. Стоит заметить, что аналогичный универсальный решатель есть и в системе естественно-языкового интерфейса. Таким образом, система естественно-языкового интерфейса может отвечать на вопросы, связанные с используемым естественным языком, в рамках его описания в базе знаний естественно-языкового интерфейса.

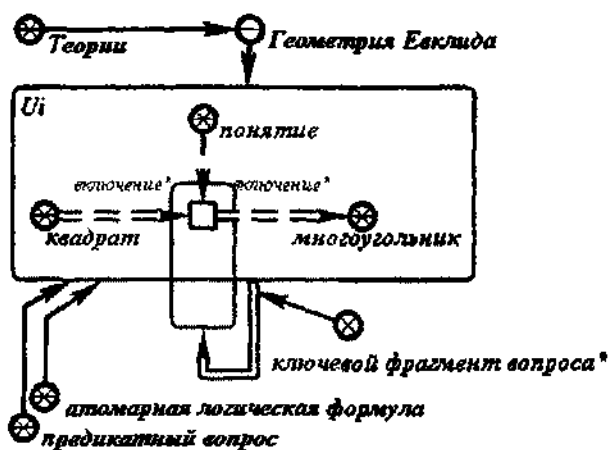


Рис.6.Пример сгенерированного вопроса к системе

Далее происходит обработка сгенерированного вопроса основной вопросно-ответной системой. Результат её работы будет сгенерированная и переданная системе естественно-языкового интерфейса sc-конструкция ответа на поставленный вопрос. Пример такого ответа приведен на рис.7.

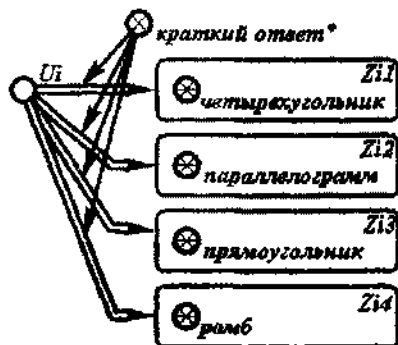


Рис.7. Пример краткого ответа от системы

После получения ответа от основной системы, задача естественно-языкового интерфейса заключается в генерации текста ответа пользователю. При генерации ответа используются sc-конструкции, полученные при анализе естественно-языкового текста запроса к вопросно-ответной системе. Такой подход позволяет более качественно и быстро получить текст ответа. Результат «склейки» ответа от вопросно-ответной системы и sc-конструкций обработки вопроса приведен на рис.8.

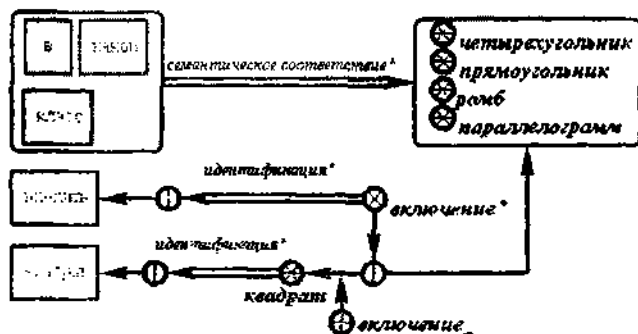


Рис.8. Пример сгенерированной sc-конструкции ответа

Результатом обработки полученной sc-конструкции с учетом информации, полученной на этапе синтаксического анализа вопроса, будет дерево лексического разбора генерируемого ответа (рис.9). На основании полученного дерева будет сгенерирован естественно-языковой текст ответа.

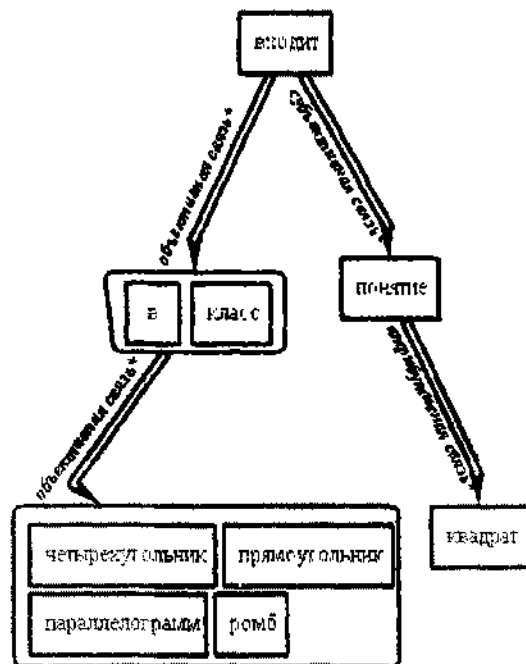


Рис.9. Пример sc-конструкции дерева лексического разбора ответа

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

ЛИТЕРАТУРА

- [1] Представление и обработка знаний в графодинамических ассоциативных машинах /В.В. Голенков [и др.]; под ред. В.В. Голенкова. – Минск : БГУИР, 2001.
- [2] Программирование в ассоциативных машинах / В.В. Голенков [и др.]; под ред. В. В. Голенкова. – Минск БГУИР, 2001
- [3] Елисеева, О.Е. Естественно-языковой интерфейс интеллектуальных систем : учебное пособие / О.Е. Елисеева. – Минск, БГУИР, 2009.
- [4] Лобанов, Б.М. Речевой интерфейс интеллектуальных систем : учебное пособие / Б.М. Лобанов, О.Е. Елисеева. – Минск, БГУИР, 2006.