



OSTIS-2014

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

СЕМАНТИЧЕСКОЕ КОНЦЕПТУАЛЬНОЕ ПРОЕКТИРОВАНИЕ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ИНТЕРФЕЙСА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ

Елисеева О.Е.*

* *Белорусский государственный университет,
г. Минск, Республика Беларусь*

olae@open.by

В работе осуществляется проектирование концепции естественно-языкового интерфейса интеллектуальных систем на основе открытой семантической технологии проектирования интеллектуальных систем OSTIS. Изложение основных результатов предлагается непосредственно на графовом семантическом языке SC, в результате чего формируется фрагмент БЗ подсистемы естественно-языкового интерфейса. Визуализация знаний о процессе общения и о естественном языке в виде семантической сети является инструментом сформирования более глубокого понимания соответствующих предметных областей.

Ключевые слова: естественно-языковой интерфейс; интеллектуальная система; лингвистическая база знаний; семантическая сеть; семантический язык.

Введение

Актуальность создания естественно-языкового интерфейса (ЕЯИ) для любой интеллектуальной системы является вполне очевидной. Тем не менее, до настоящего времени открытым остаётся вопрос массового использования ЕЯИ. Главными причинами такого положения дел являются:

- Широкое использование ставшего стандартом графического интерфейса с разнообразными панелями инструментов в виде кнопок, переключателей, форм и пр., к которым все пользователи уже привыкли, в связи с чем возникает проблема грамотного сочетания стандартных интерфейсных средств с естественно-языковыми;
- Значительное количество проблем, связанных со сложностью автоматического анализа и синтеза естественного языка. Предметная ориентированность прикладных систем, а также отсутствие универсальной технологии создания ЕЯИ фактически приводят к тому, что для каждой отдельной системы приходится создавать ЕЯИ практически с нуля, что слишком затратно и ресурсоёмко.

В данной работе осуществляется очередная попытка (наряду с уже представленными в работах [Гецевич и др., 2012], [Елисеева, 2009], [Корончик, 2013] и др.) проектирования концепции ЕЯИ интеллектуальных систем, которая ориентирована

на массовое универсализированное применение. В качестве основы для проектирования взята открытая семантическая технология проектирования интеллектуальных систем (Open Semantic Technology for Intelligent Systems - OSTIS) [Ostis, 2012] и соответствующие ей графовые языки представления знаний на основе семантических сетей [Голенков и др., 2013].

Изложение основных результатов в статье предлагается непосредственно на графовом семантическом языке SC (Semantic Code), в его графической нотации SCg-коде. Таким образом, в данной работе ставится своеобразный эксперимент, в результате которого должна быть доказана либо опровергнута гипотеза о том, что представление знаний в виде семантической сети может способствовать более эффективному пониманию соответствующей информации.

1. Основные понятия

Для начала уточним понимание термина «семантическое проектирование». В данной работе осуществляется анализ смысла (семантики) каждого понятия (термина), используемого в процессе проектирования, с целью определения семантических связей между компонентами ЕЯИ, выделения основных (ключевых) технологических моментов, а также формирования базовой онтологии понятий и отношений предметной области, соответствующей понятию ЕЯИ. Процесс такого переосмысления и представления его

результатов в виде sc-текстов будем называть **семантическим проектированием**.

1.1. Моделирование процесса общения на естественном языке

Фактически ЕЯИ является компьютерной (программной) **моделью процесса общения**. Для краткости в данной работе мы не рассматриваем все возможные определения понятия «общение» (такая попытка осуществлялась автором в [Елисеева, 2009]), а будем опираться на формальную модель общения, предложенную Э.В.Поповым в книге [Попов, 1982]. Следуя рассуждениям в [Попов, 1982], представим модель общения в виде фрагмента семантической сети на языке SC (sc-текста, рис. 1). С целью облегчения восприятия в данном sc-тексте не уточнена семантика связей между указанными понятиями.

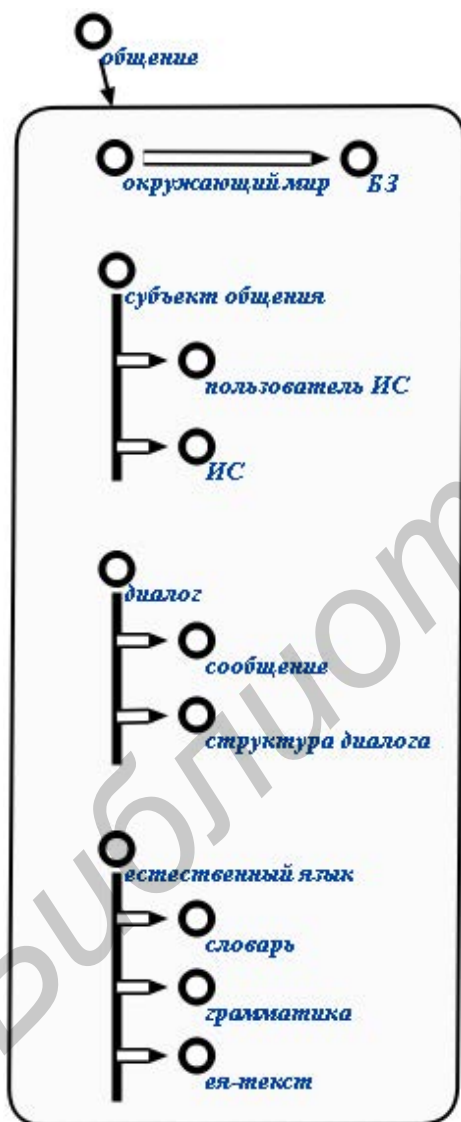


Рисунок 1 – Модель общения

Согласно базовой денотационной семантике языка SC, узлы с идентификаторами являются знаками соответствующих понятий, а дуги обозначают наличие связей между ними. На рис.1 задействован также специальный тип sc-дуги,

которая изображена в виде двойной линии со стрелкой и является знаком ориентированной пары. Для уточнения семантики связи между понятиями в указанные дуги должны быть проведены дуги из знаков соответствующих отношений. Достаточно часто для этого используются теоретико-множественные отношения.

Рассмотрим более подробно каждый из перечисленных в указанной модели компонентов. База знаний (БЗ) предметной области в каждой конкретной ИС, как правило, является некоторой частью модели окружающего мира. Рассмотрение технологии создания указанной БЗ не является целью данной работы, т.к. этому вопросу посвящено достаточное количество других работ. В следующей главе данной работы БЗ будет проанализирована на предмет наличия в ней «естественно-языковых объектов».

1.2. Субъекты общения и обмен сообщениями

Основой данной модели является наличие двух субъектов общения: пользователя и интеллектуальной системы. Начнём с рассмотрения примера описания знаний о субъектах общения в рамках базы знаний ЕЯИ. Предположим, мы имеем некоторую прикладную интеллектуальную систему и некоторого её пользователя *X*. Опишем данное утверждение в линейной нотации языка SC:

пользователь ИС -> *Пользователь X*;
ИС -> *Прикладная ИС-X*;

Из данного фрагмента sc-текста видно, что мы ввели два знака множеств:

- всех пользователей интеллектуальных систем (ИС), которое задали sc-узлом с идентификатором *пользователь ИС*. При этом узел с идентификатором *Пользователь X* обозначает некоторого конкретного пользователя, и в общем случае из узла с идентификатором *пользователь ИС* будет выходить ровно столько дуг, сколько пользователей у системы. Все узлы, в которые указанные дуги входят, относятся к константному типу и являются некоторыми «виртуальными представителями» (моделями) конкретных пользователей, каждый из которых относится к объектам окружающего мира (см. рис.1). Для описания этого в языке SC имеются соответствующие средства;

- всех интеллектуальных систем, которое задали sc-узлом с идентификатором *ИС*. Аналогично сказанному выше, узел *Прикладная ИС-X* является знаком конкретной прикладной интеллектуальной системы. Если мы будем рассматривать ЕЯИ конкретной ИС, то, по всей видимости, данное описание может оказаться избыточным. Но если следовать логике рассуждений Э.В.Попова ([Попов, 1982]), то в ИС должна быть представлена некоторая «модель себя». Следовательно, такое описание может оказаться вполне востребованным, например, в

рамках компонента ИС, отвечающего за компонент «Помощь (руководство пользователя)»).

В процессе общения субъекты обмениваются сообщениями (рис. 2). Из sc-текста на рис. 2 мы видим, что для описания сообщений в рамках ЕЯИ необходимо ввести новое тернарное отношение *сообщение-автор-адресат**, описывающее семантические связи между автором (обозначено атрибутом *автор_* рассматриваемого отношения), адресата (*адресат_*) и собственно сообщения (*сообщение_*).

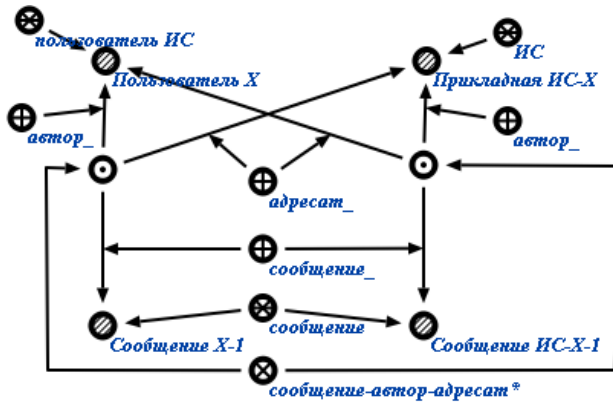


Рисунок 2 – Сообщения субъектов общения

Уточним понимание использованных на рис. 2 символов и обозначений. Для описания тернарного (как и любого другого) отношения используются следующие изображения sc-узлов ([Ostis, 2012]), задающие семантику (роли) каждого компонента отношения:

- ⊗ - собственно знак отношения;
- ⊕ - знаки атрибутов отношения;
- ⊙ - знак связки отношения.

Очевидно, что в прикладной ИС сообщения пользователя и ИС взаимосвязаны по принципу вопрос – ответ (импульс – реакция). Данное бинарное асимметричное отношение опишем с использованием ключевого узла *вопрос-ответ*.

Помимо базовой взаимосвязи типа «вопрос – ответ» очевидно, что при проектировании и создании интерфейса ИС, как и любой другой прикладной системы, все сообщения соответствующим образом классифицируются, представляются в различной форме и реализуются различными способами. В рамках ЕЯИ ставится задача автоматического анализа поступающих от пользователя сообщений, в результате которого выполняется (рис. 3):

- распознавание типа сообщения;
- определение соответствующего запроса к базе знаний (БЗ), его параметров и операндов либо обработка введённых пользователем данных/знаний в ответ на запрос ИС.

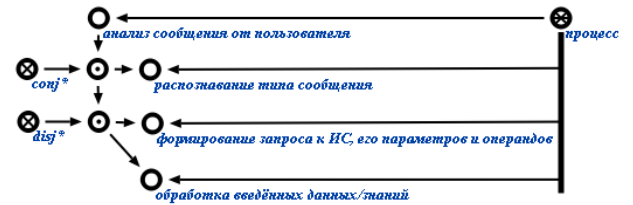


Рисунок 3 – Процесс анализа сообщений от пользователя

Из сказанного выше следует, что в процессе проектирования ЕЯИ необходимо определить типологию сообщений от пользователя. В первом приближении все сообщения можно разбить на 2 группы: запросы и ответы. При этом в качестве запросов можно выделить: поиск информации в БЗ; преобразование информации; добавление информации в БЗ; уточняющий вопрос и др.

Исходя из приведённой типологии, на данном этапе проектирования можно сформулировать первые правила анализа входных сообщений в форме следующих имплицитивных высказываний:

1. «Если входное сообщение от пользователя относится к типу «запрос», то в результате его анализа необходимо сформировать соответствующий запрос к БЗ и определить перечень его параметров и операндов» (рис. 3);
2. «Если входное сообщение от пользователя относится к типу «ответ», то в результате его анализа необходимо обработать введённые пользователем данные или знания и сформировать запрос на добавление этих данных/знаний в соответствующую БЗ и/или базу данных (БД)», если таковая используется в ИС (рис. 4).

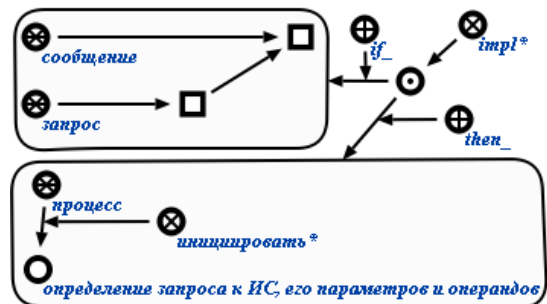


Рисунок 3 – Правило № 1 анализа сообщения от пользователя

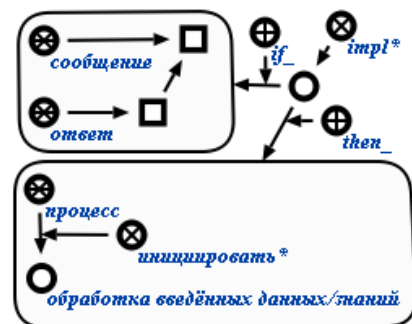


Рисунок 4 – Правило № 2 анализа сообщения от пользователя

Следует отметить, что из указанных формулировок также следует, что первым этапом анализа сообщения от пользователя является

именно определение его типа. Таким образом, мы можем начать описывать упорядоченный сценарий (последовательность операций) обработки сообщений (рис.5).



Рисунок 5 – Описание процесса анализа сообщений от пользователя

После обработки сообщения от пользователя ИС в результате обработки БЗ должна сгенерировать ответное сообщение. Ответные сообщения также классифицируются и разбиваются на две группы: ответы и запросы. Среди ответов, в частности, можно выделить:

- прямой ответ, т.е. ответ, полностью соответствующий (релевантный) запросу пользователя;
- косвенный ответ – ответ, в котором перехватывается инициатива и уточняется понимание [Попов, 1982];
- инструкция – ответ, в котором ИС сообщает (объясняет) пользователю, что и как он должен делать и др.

Очевидно, что рассмотренные классификации сообщений пользователя и ИС весьма приблизительные и требуют значительных уточнений, но на данном этапе концептуального проектирования мы такую задачу не ставим. Распознавание типа сообщения при этом представляет собой достаточно нетривиальную технологическую-лингвистическую задачу, в постановке которой следует учесть два момента:

- какие запросы и ответы могут быть реализованы ИС и как;
- каким образом и на основании каких лексических и грамматических составляющих тот или иной тип сообщения можно распознать в результате автоматической обработки соответствующего естественно-языкового текста (ЕЯ-текста).

Указанную задачу в данной работе мы также лишь обозначаем, выделив на концептуальном уровне набор соответствующих операций. Отметим, что при реализации «традиционного» графического интерфейса типология сообщений (которые в

сущности представляют собой команды или операции) осуществляется в процессе проектирования и представляет собой заранее заданную совокупность, которая поддерживается соответствующими элементами интерфейса: кнопками, переключателями и пр.

1.3. Моделирование диалога

В модели общения (рис.1) показано, что процесс взаимного обмена сообщениями пользователя и ИС реализуется в форме диалога. Таким образом, следующим этапом моделирования процесса общения и его семантического описания является уточнение всех понятий, связанных с **диалогом**. Сообщения при этом являются основными единицами, из которых строится диалог. При создании диалоговых систем говорят о проектировании общей и конкретной структуры диалога. Рассмотрим, каким образом это можно описать на языке SC.

На рис.6 представлен пример конкретной структуры диалога, которая, очевидно, является линейной последовательностью сообщений.

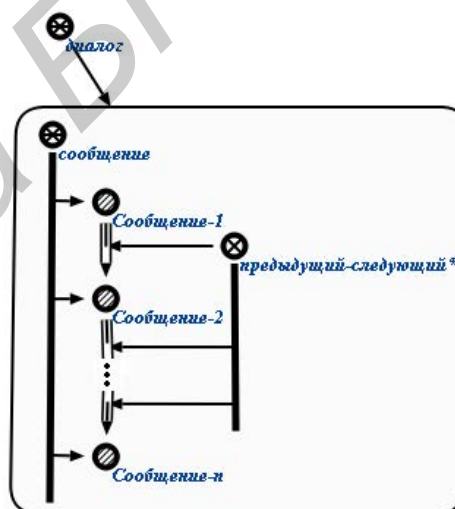


Рисунок 6 – Линейная последовательность сообщений в конкретной структуре диалога

Не менее очевидно также, что общая структура диалога представляет собой направленную, разветвленную структуру. Для некоторых конкретных ИС на этапе их разработки можно попытаться описать первоначальные конструкции, отражающие возможные варианты развития диалога пользователя с данной системой. В реальности же общение на естественном языке является недетерминированным процессом, описать который заранее практически невозможно. Поэтому в данной работе будем исходить из следующих положений:

- при проектировании ЕЯИ описывается некоторая базовая совокупность общих структур диалога, соответствующих различным типам общения (например, диалог-приветствие (рис.7), диалог-справка о системе, диалог-запрос на решение задачи по предметной области, диалог-

информационно-поисковый запрос по предметной области);

- в процессе создания конкретной ИС по конкретной предметной области инженер знаний может описать дополнительные общие структуры аналогичным образом;

- в рамках концептуального проектирования ЕЯИ ставится задача разработки методов и алгоритмов автоматического формирования общих структур диалогов различного типа на основе анализа и обобщения конкретных взаимодействий с пользователями (блок самообучения ЕЯИ).

В sc-тексте на рис.7 мы ввели ряд семантических связей, отражающих сценарий работы с пользователем на этапе входа в ИС. Помимо сообщений, которыми обмениваются пользователь и ИС, целесообразно говорить о происходящих в процессе рассматриваемого взаимодействия **событиях**, для первичного описания необходимости обработки которых вводится бинарное ориентированное отношение *обработка события** (рис.8).

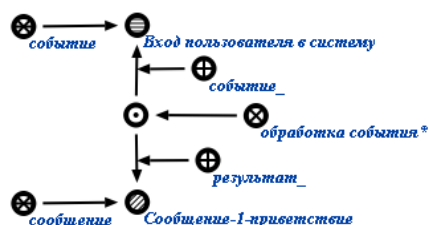


Рисунок 7 – Пример использования отношения *обработка события*

Использованные на рис.6 и 7 идентификаторы sc-констант весьма условны и введены лишь для лучшего понимания читателем данного изложения. В процессе реального функционирования ЕЯИ эти идентификаторы не нужны, т.к. семантика соответствующих узлов описывается с помощью соответствующих введённых в БЗ ЕЯИ знаков множеств и отношений.

1.4. Естественный язык

Центральным элементом рассматриваемой модели общения является **естественный язык** как основное средство взаимодействия пользователя и ИС в рамках ЕЯИ. Следуя логике размышлений данной работы, для начала необходимо уточнить понимание понятия «естественный язык». Уточним высказывание, представленное на рис.1, указав семантику представленных на нём узлов и дуг (рис.8).

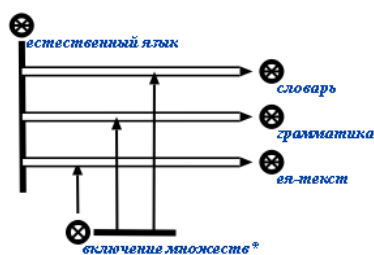


Рисунок 8 – Состав естественного языка

Заметим, что в данной работе мы не претендуем на полную представленность описания, а лишь обозначаем основные моменты. Кроме того, некоторые примеры и подходы к формализации знаний о естественном языке были представлены в [Гецевич и др., 2012], [Елисеева, 2009], [Елисеева и др., 2013].

Очевидно, что в процессе создания ЕЯИ необходимо обеспечить эффективное решение двух основных задач:

- автоматический анализ ея-текстов;
- автоматическая генерация ея-текстов.

Таким образом, в рамках представляемого концептуального моделирования ЕЯИ можно говорить о соответствующих указанным задачам двух наборах (множествах) процедур и/или операций, а также исходных данных для них. На данном этапе размышлений не достаточно ясно, каким образом соотносятся между собой указанные два множества. Поэтому здесь мы пока используем теоретико-множественное отношение *включение множеств**. Множества, заданные sc-узлами *автоматический анализ ея-текстов* и *автоматическая генерация ея-текстов* могут включать как декларативные, так и процедурные знания, относящиеся к любому из компонентов языка, перечисленных на рис.8.

В процессе проведённого анализа основных понятий был выполнен первый шаг формализации предметной области «Общение» и создан фрагмент соответствующей базы знаний, а точнее – определён набор ключевых sc-узлов. Все знания в процессе создания конкретной ИС хранятся в общей БЗ системы и представляются единым образом. Благодаря этому, мы получаем возможность использовать одни и те же инструменты как для обработки БЗ ЕЯИС, так и для обработки БЗ предметной области.

2. Лингвистическая база знаний и база знаний предметной области: формы представления языка

Отметим следующее:

- очевидно, что при создании ЕЯИ необходимо говорить о создании лингвистической БЗ. Однако на уровне БЗ предметной области также можно обнаружить достаточно много «лингвистических конструкций»;

- в процессе формализации знаний об окружающем мире и представления их в БЗ ИС на языке SC мы используем слова естественного языка;

- связи между объектами, явлениями и пр. окружающего мира, а также их базовые семантические характеристики, мы задаём либо с использованием специально вводимых обозначений и символов языка представления знаний SC, либо с использованием соответствующих ключевых

sc-узлов с идентификаторами-словами естественного языка;

- ввиду сказанного становится очевидно, что sc-тексты содержат элементы естественного языка, которые могут быть использованы в реализации естественно-языкового общения с пользователем.

Итак, в БЗ ИС информацию о ЕЯ мы имеем:

- в составе идентификаторов ключевых sc-узлов – это основные понятия соответствующей предметной области;

- в формальных высказываниях, описывающих определения понятий, аксиомы, теоремы и их доказательства, утверждения, задачи и т.п. (см., например, [Давыденко и др., 2011]);

- в виде протоколов решения задач из предметной области;

- в виде формальных описаний конкретных объектов предметной области. Например, для ИС решения задач по геометрии [Давыденко и др., 2011] это описания геометрических фигур и их соотношений (рис.9);

- в текстовых содержимых sc-узлов;

- в нетекстовых содержимых sc-узлов, которые можно описать на ЕЯ. Например, для ИС по геометрии это чертежи геометрических фигур. Для таких содержимых в предметной области создаются мета-описания и/или подписи на ЕЯ (рис.10).

Исходя из сказанного выше, можем сформировать первоначальную совокупность операций и исходных данных для них, необходимых для автоматической обработки естественного языка:

автоматический анализ ея-текстов

⇒ формирование sc-текста на основе ея-текста

{⇒ формирование sc-идентификатора на основе ея-слов

{⇒ автоматическое извлечение из ея-текста ключевых понятий

⇒ лемматизация

⇒ модель управления слова }

⇒ формирование sc-высказывания на основе фрагмента ея-текста

⇒ формирование связей между sc-высказываниями }

⇒ обработка текстового содержимого sc-узла

автоматическая генерация ея-текстов

{⇒ перевод sc-текста на ея

{⇒ преобразование идентификатора в ея-словоформу

{⇒ правила словоизменения

⇒ словарь словоформ }

⇒ формирование ея-словосочетания

{⇒ правила согласования словоформ

⇒ типы сочетаемости слов

⇒ шаблоны словосочетаний

⇒ лексическая функция }

⇒ формирование ея-предложения

{⇒ типология предложений

⇒ правила построения предложений

⇒ модель управления слова

⇒ шаблоны предложений }

⇒ правила трансляции sc-высказываний на ея

{⇒ правила трансляции теретико-множественных высказываний

⇒ правила трансляции логических высказываний

{⇒ правила трансляции имплицативных высказываний

⇒ правила трансляции высказываний о существовании }

⇒ формирование ея-текста

{⇒ стиль текста

⇒ структура текста

⇒ определение последовательности предложений в тексте }

⇒ обработка текстового содержимого sc-узла }

На рис. 11 представлен пример описания в БЗ синонимичных высказываний о составе естественного языка, которые являются трансляцией (переводом) sc-текста, представленного на рис.8, на русский язык.

Из представленного на рис.11 примера видно, что одно и одно и то же sc-выражение может быть транслировано на естественный язык различными способами. В дополнение к характерному для естественного языка явлению синонимии мы имеем синонимию, связанную с особенностями представления базы знаний на основе теоретико-множественных рассуждений. Очевидно, что конечным (неподготовленным) пользователем ИС наиболее понятно и естественно будет воспринято первое предложение: «Естественный язык состоит из словаря, грамматики и текстов.». Отметим, что из исходного sc-текста указанное предложение «напрямую» не следует. Для его получения необходимо выполнить следующие операции:

- транслировать идентификаторы в соответствующие словоформы и словосочетания (результат см., например, на рис.12);

- выбрать наиболее подходящий (естественный) эквивалент из списка синонимичных слов и словосочетаний;

- определить синтаксические роли и связи отобранных словоформ и словосочетаний;

- построить синтаксическую структуру результирующего предложения;

- синтезировать результирующее предложение.

На рис.13 приведён пример описания словообразовательной парадигмы, которая либо должна быть занесена в лингвистическую БЗ экспертом-лингвистом, либо сгенерирована

автоматически на основе правил словоизменения. На основе указанной парадигмы выбирается необходимая для синтеза ея-текста словоформа.

На рис.14 показана связь синонимии, благодаря которой в результирующем ея-тексте вместо слова «ея-тексты» можно получить слово «тексты».

На рис.15-17 представлены фрагменты лингвистической БЗ, на основе которого генерируется рассматриваемый нами пример ея-предложения. В частности, на рис.15 указано следующее:

- слово «включение», являющееся частью словосочетания «включение множеств», синонимичного идентификатору «множеств*», является производным отглагольным существительным от слова «включать»;
- слово «включать» имеет синонимы «содержать», «состоять из»;
- для предиката «состоять из» описана модель управления слова (МУС) [Мельчук, 1974], в которой задана соответствующая синтаксическая, морфологическая и семантическая информация.

На рис.16 описано преобразование перевода (трансляции) словосочетания «включение множеств» в эквивалентное его представление в виде sc-текста, в котором уточнён тип связи слов в словосочетании.

Наконец, на рис. 17 приведён пример эквивалентного соответствия (перевода) ея-предложения и его представления в виде синтаксической структуры на языке SC.

В завершение данного рассмотрения следует отметить, что здесь в некотором приближении был рассмотрен процесс генерации ея-текста на основе sc-текста, что фактически представляет собой перевод с одного языка (языка представления знаний) на другой (естественный язык). Предлагаемый подход продиктован тем, что фактически обмен сообщениями пользователя и ИС в процессе естественно-языкового диалога представляет собой ни что иное, как перевод ея-текстов на язык SC и наоборот. Важным преимуществом языка SC при этом является его изначальная ориентация на семантическое описание различных предметных областей, что в значительной степени облегчает интерпретацию многих высказываний.

Следовательно, из сказанного выше, можно предположить, что в процессе реализации предлагаемой концепции возможно получение вполне надёжных универсальных правил, методов, алгоритмов и процедур автоматического перевода текстов.

3. От базы знаний – к пониманию. Когнитивно-семиотический аспект

Как сказано выше, важнейшее особенностью языка SC и его подязыков является изначальная

ориентация на семантическое представление знаний различного вида. Особенно важно, что на этом же языке реализуются и программы обработки рассмотренных sc-текстов. Таким образом, мы имеем интегрированную среду, в которой знания и процедуры их обработки представляются единым образом.

Решение задач обработки естественного языка, опираясь на sc-представления, таким образом, даёт нам возможность описывать единым образом все уровни языка: лексику, морфологию, синтаксис, семантику, - а также процедуры обработки знаний каждого уровня. При этом естественным образом мы поддерживаем взаимосвязь всех уровней языка и фактически не сталкиваемся с проблемой специального выделения каждого уровня и их разграничения. Заметим, что аналогичная ситуация характерна и для самого естественного языка, в котором, например, в процессе синтаксического анализа/синтеза так или иначе приходится апеллировать к семантике. Таким образом, мы сохраняем эту особенность в БЗ ЕЯИ.

В данной работе для иллюстрирования основных моментов проектировании сознательно не использовались никакие иные средства, кроме изобразительных возможностей графовой нотации языка SC – SCg-кода. Автор надеется, что уважаемый читатель не испытывал особых трудностей в понимании приведённых в тексте sc-конструкций. Если это действительно так, то это, с одной стороны, является доказательством семантической мощи языка SC. С другой стороны, следует ещё раз отметить, что фактически в процессе рассуждений мы создавали фрагменты реальной БЗ. Очевидно, что многие моменты в статье, учитывая объективные ограничения, опущены, но если всё изложенное представить на SC, мы получим полноценную БЗ естественно-языковой подсистемы ИС.

Использование технологии формирования и уточнения понятийной системы предметной области путём введения ключевых sc-узлов в качестве знаков понятий и явного описания теоретико-множественных и логических взаимосвязей между ними в виде семантической сети даёт возможность более чётко осознать семантику соответствующей предметной области. При этом эта семантика вполне естественным образом визуализируется в виде sc-текстов. Таким образом, можно утверждать, что язык SC является мощным инструментом понимания. Эксперт, создающий БЗ на языке SC, получает возможность глубже понять соответствующую предметную область. Чтение же и понимание смысла sc-текстов – не намного более сложная, а порой даже более простая задача, чем чтение и понимание естественно-языкового текста, т.к., несмотря на первоначальное ощущение некоторой громоздкости, sc-тексты более структурированы по сравнению с ея-текстами. Так же, как в памяти человека информация представляется нелинейным образом, sc-тексты

тоже являются нелинейными. Тем самым сохраняется, если так можно выразиться, преемственность представления семантики на когнитивном «человеческом» уровне и на уровне внутренних представлений интеллектуальной системы.

Учитывая сказанное, мы приближаемся к моделированию процессов понимания в БЗ ИС.

Заключение

В данной работе представлено семантическое концептуальное моделирование ЕЯИ ИС, в результате которого были созданы фрагменты БЗ соответствующей ея-подсистемы в виде семантической сети. К сожалению, формат статьи

не даёт возможности представить полную онтологию понятий ЕЯИ. Некоторые поднятые в работе вопросы требуют более тщательного осмысления с привлечением экспертов из области лингвистики.

Кроме того, одной из задач данной работы явилось иллюстрирование экспертам-лингвистам, а также специалистам любых других предметных областей, поэтапного процесса формализации знаний, представленных на естественном языке. Всё повествование в работе построено по принципу перевода сформулированных на естественном языке размышлений на семантический язык представления знаний.

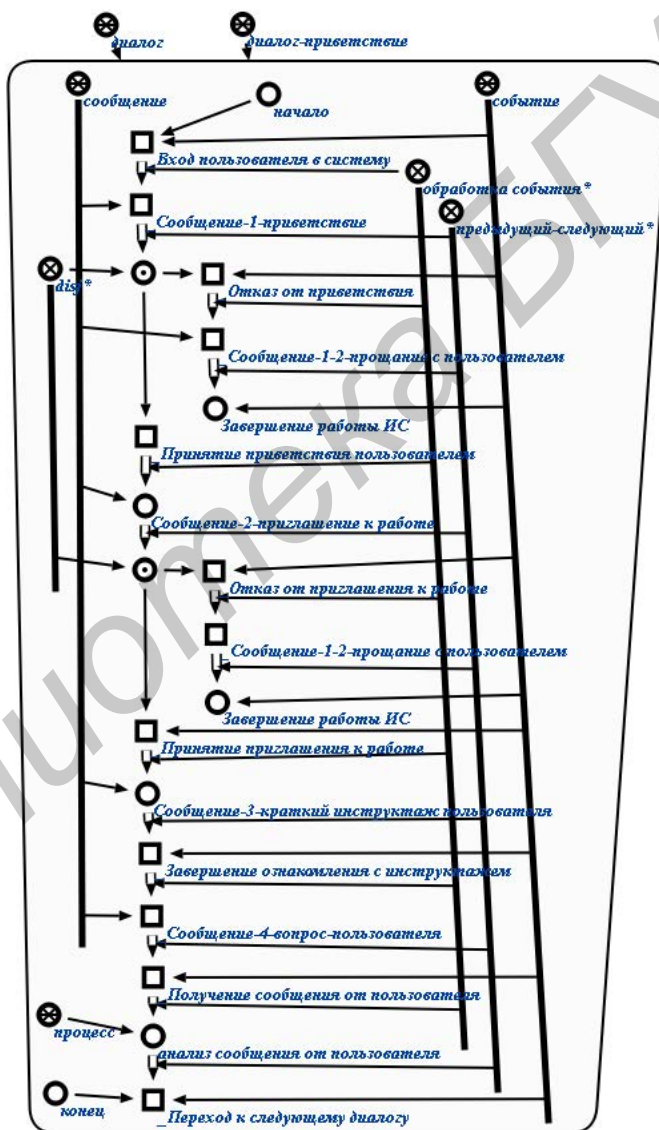


Рисунок 9 – Общая структура диалога-приветствия

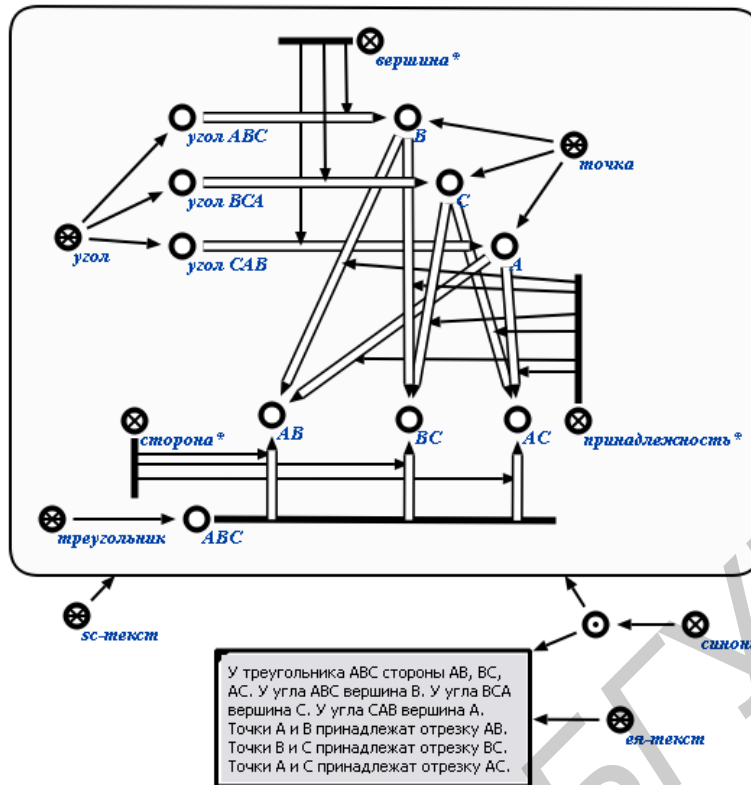


Рисунок 10 – Пример описания геометрических фигур в виде sc-текста и синонимичного ея-текста

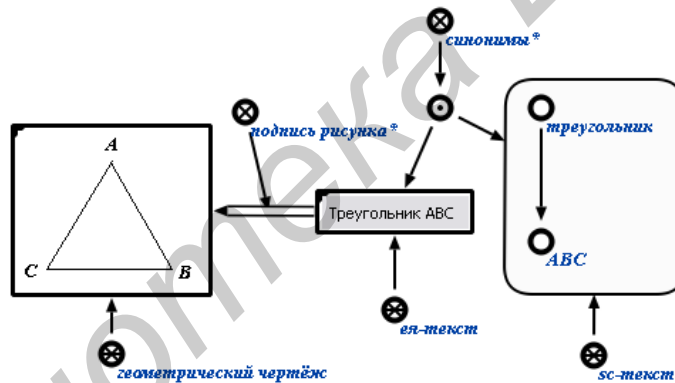


Рисунок 11 – Пример подписи геометрического чертежа

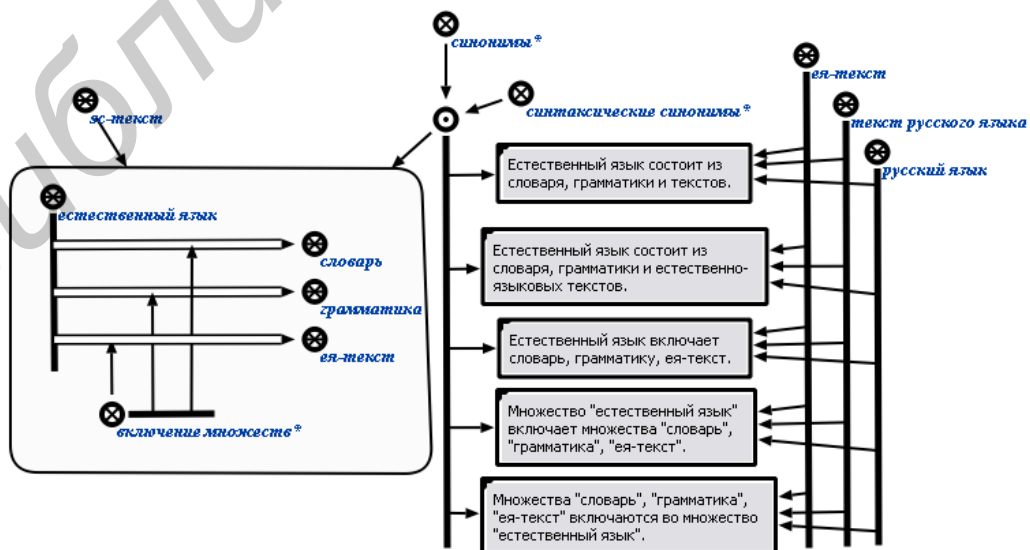


Рисунок 12 – Пример синонимии sc-текстов и ея-текстов



Рисунок 13 – Описание синоними идентификатора и его записи в виде ея-текста

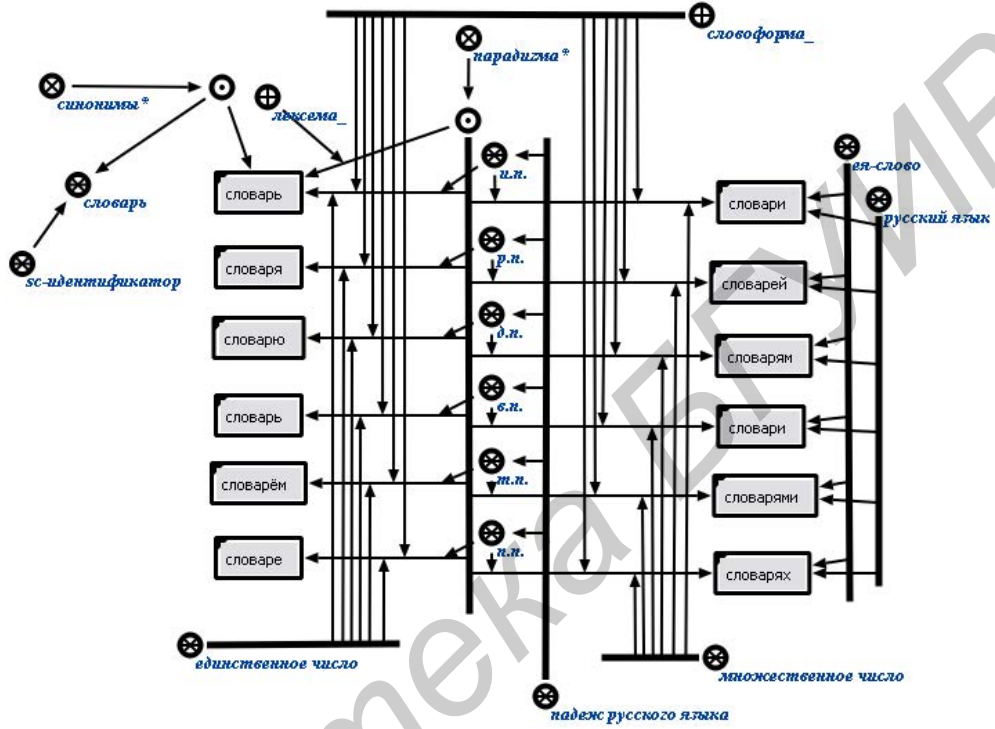


Рисунок 14 – Пример словообразовательной парадигмы

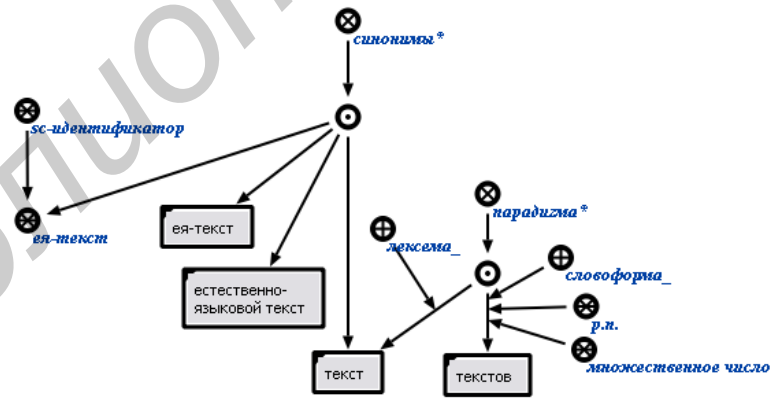


Рисунок 15 – Несколько синонимов для ея-представления идентификатора

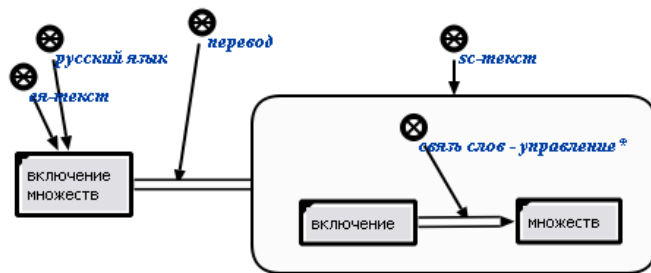


Рисунок 16 – Пример перевода ея-словосочетания в формат сс-текста

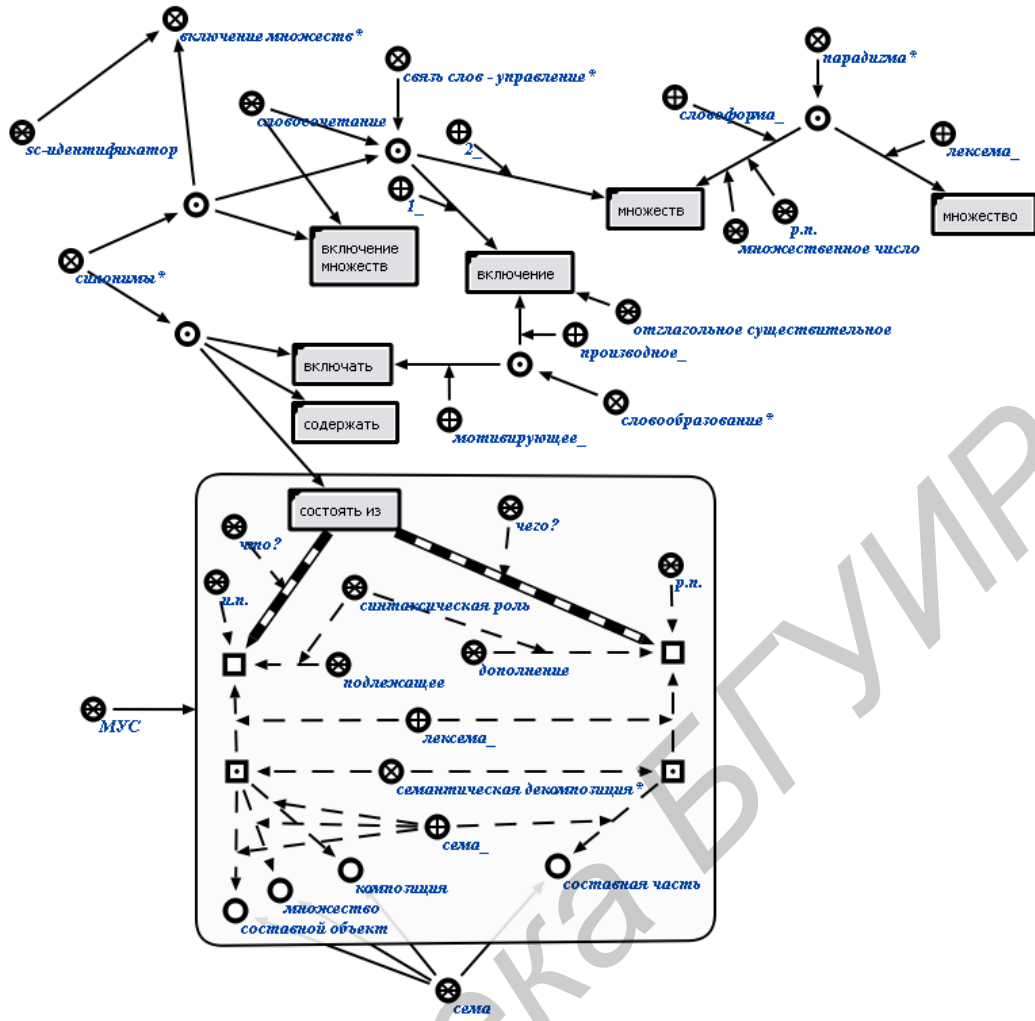


Рисунок 17 – Фрагмент лингвистической БЗ

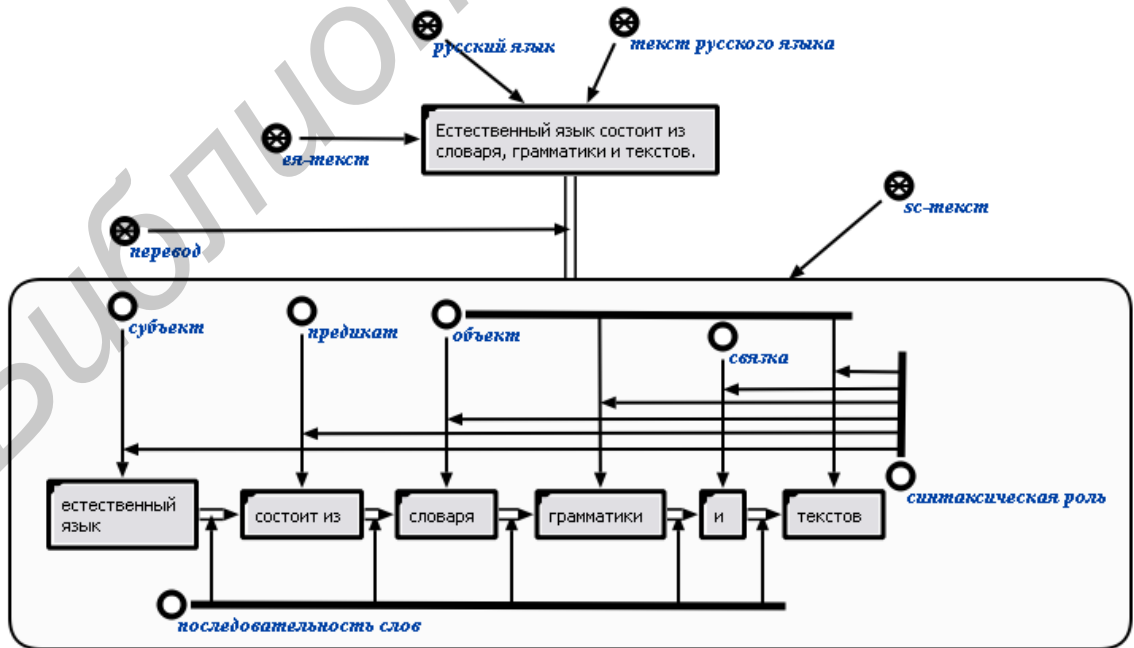


Рисунок 18 – Соответствие ея-предложения и его синтаксической структуры в формате sc-текста

Библиографический список

[Апресян, 1995] Апресян, Ю. Д. Избранные труды, том II. Интегральное описание языка и системная лексикография / Ю. Д. Апресян – Москва: Школа «Языки русской культуры», 1995.

[Гецевич и др., 2012] Семантическая технология проектирования белорусско- и русскоязычных ея-интерфейсов вопросно-ответных систем / Ю.С. Гецевич [и др.] // сборник трудов Международной научно-технической конференции OSTIS-2012, Минск, 2012 / БГУИР – Минск, 2012.

[Голенков и др., 2013] Голенков, В.В., Гулякина, Н.А. Открытый проект, направленный на создание технологии компонентного проектирования интеллектуальных систем / В.В. Голенков, Н.А. Гулякина // сборник трудов Международной научно-технической конференции OSTIS-2013, Минск, 2013 / БГУИР – Минск, 2013.

[Давыденко и др., 2011] Интеллектуальная справочная система по геометрии / Давыденко И.Т. [и др.] // сборник трудов Международной научно-технической конференции OSTIS-2011, Минск, 2011 / БГУИР – Минск, 2011.

[Елисеева, 2009] Елисеева О.Е. Естественно-языковой интерфейс интеллектуальных систем: учеб. пособие / О. Е. Елисеева; под науч. ред. проф. В. В. Голенкова. – Минск: БГУИР, 2009.

[Елисеева и др., 2013] Елисеева О.Е., Русецкий К.В. Подходы к формализации естественного языка на семантическом графовом языке представления знаний // Седьмые Карповские научные чтения. Сборник научных статей, 2013.

[Корончик, 2013] Корончик, Д.Н. Унифицированные семантические модели пользовательских интерфейсов интеллектуальных систем и технология их компонентного проектирования / Д.Н. Корончик // сборник трудов Международной научно-технической конференции OSTIS-2013, Минск, 2013 / БГУИР – Минск, 2013.

[Мельчук, 1974] Мельчук, И.А. Опыт теории лингвистических моделей «Смысл-Текст». Семантика, Синтаксис / И.А. Мельчук. – Москва, 1974.

[Попов, 1982] Попов, Э.В. Общение с ЭВМ на естественном языке / Э.В. Попов. – Москва: Наука, 1982.

[Ostis, 2012] Open Semantic Technology for Intelligent Systems. [Электронный ресурс]. – 2012. – Режим доступа: <http://www.ostis.net/>. – Дата доступа: 20.12.2013.

SEMANTIC CONCEPTUAL DESIGN OF NATURAL LANGUAGE INTERFACES FOR INTELLIGENT SYSTEM

Yeliseyeva O.E.*

**Belarusian State University, Minsk, Republic
of Belarus*

olae@open.by

In the work conceptual design of natural language interfaces of intelligent system based on Open Semantic Technology for Intelligent Systems OSTIS is performed. Exposition of the main results is offered directly on graph semantic language SC, thereby fragment of knowledge base for natural language interface subsystem is formed. Visualization of knowledge about the process of communication and natural language as a semantic network is a tool for a better understanding of the relevant domains.

Introduction

Urgency of creation of natural language interface to any intelligent system is quite obvious. However, until now an open question is mass usage of natural language interface. The main reasons for this situation are:

- The widespread use of the standard graphical interface with multiple toolbars in the form of buttons, switches, forms, etc., to which all users have become accustomed. Therefore there is a problem an intelligent combination of standard interface tools with natural-language;

- A significant number of problems are associated with the complexity of the automatic analysis and synthesis of natural language. Object-oriented application systems, as well as the absence of a universal technology of natural language interface cause the that each system necessary to create natural-language interface virtually from scratch that is too costly and resource intensive.

Main Part

Natural language interface is a computer (software) model of the communication process. The basis of this model is the two subjects of communication: the user and intelligent system. In the process of communicating subjects are exchanging messages. In the design process of natural language interface we must define a typology of messages. Next we have to formulate the rules of analysis of input messages.

A process of messaging user and intelligent system is implemented in the form of a dialogue. When it's creating a dialogue systems, general and specific structure of dialogue is designed.

The central element of the model of communication is the natural language. Relationships between objects, events, etc. of the world, as well as many of their characteristics, we set either using specially introduced signs and symbols of the language knowledge representation SC, or by using appropriate key sc-node identifiers words of natural language. Sc-texts contain elements of natural language, which may well be used in the natural language of communication with the user.

Conclusion

In this paper we presented a semantic conceptual modeling of natural language interface for intelligent system. As a result the fragments of knowledge base of natural language subsystem were specified as a semantic network. Unfortunately, the article format does not allow to provide full ontology of concepts of natural language interface. Some raised in the work questions require a more thorough thinking by involving of experts from the field of linguistics.

In addition, one of the objectives of this work was to illustrate the linguistic experts, as well as professionals of any other domains, a gradual process of formalization of knowledge represented on natural language. All narrative in the work built on the principle of translation of thoughts formulated in natural language to the semantic knowledge representation language.