

# ЯЗЫКОВЫЕ СРЕДСТВА СИНТЕЗА ТЕКСТОВ ЗАДАНИЙ ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМ

К. В. Русецкий, А. Ф. Хусаинов

Кафедра интеллектуальных информационных технологий, Белорусский государственный университет информатики и радиоэлектроники

НИИ «Прикладная семиотика» Академии наук Республики Татарстан

Минск, Республика Беларусь; Казань, Российская Федерация

E-mail: {rusetski.k, khusainov.aidar}@gmail.com

*Рассмотрено семейство подязыков языка SC, ориентированных на представление лингвистической информации и обеспечивающих синтез текстов заданий в интеллектуальных обучающих системах.*

## ВВЕДЕНИЕ

Для реализации синтеза текстов заданий в интеллектуальных системах, разрабатываемых по технологии OSTIS, необходимо разработать и дополнить ряд подязыков языка Semantic Code для представления лингвистических знаний в семантических сетях, а именно: языки представления синтаксической структуры и морфологических характеристик, расширенный язык вопросов для поддержки вопросов от системы к пользователю, соответствующим этапам перехода от формального представления вопроса в памяти интеллектуальной системы к естественно-языковому тексту задания.

### I. РАСШИРЕНИЕ ЯЗЫКА ВОПРОСОВ

SC-язык вопросов [1] необходимо расширить для представления вопросов системы к пользователю. Такое расширение языка вопросов состоит во введении ключевого узла отношения *адресат\**, которое специфицирует конкретного адресата вопроса (в зависимости от автора вопроса может быть программным агентом, платформенным агентом, либо текущим пользователем), а также узлов *Система* (синонимичен знаку конкретной обучающей системы, с которой взаимодействует пользователь в данный момент) и *Текущий пользователь* (синонимичен знаку пользователя, взаимодействующего с системой в текущем сеансе и выступающего в качестве адресата учебных заданий).

### II. ФОРМАЛЬНАЯ СТРУКТУРА УЧЕБНОГО ЗАДАНИЯ

Формально структура учебного задания представляет собой ориентированную пару  $LT = \langle D, Q \rangle$ , где  $LT$  – учебное задание (от “learning task”),  $D$  – множество исходных данных (data),  $Q$  – множество вопросов об исходных данных (questions). С точки зрения теории актуального членения, это примерно соответствует теме и реме высказывания [3].

Для формализации учебного задания в памяти интеллектуальной системы вводятся ключевые SC-узлы для следующих понятий (их можно увидеть на рисунке 2):

- *учебное задание* – множество всевозможных учебных заданий, известных обучающей системе или сгенерированных ей;
- *исходные данные*’ – ролевое отношение, связывающее конкретный экземпляр учебного задания со множеством его исходных данных;
- *вопросы*’ – ролевое отношение, связывающее конкретный экземпляр учебного задания со множеством вопросов относительно его исходных данных.

Следует отметить, что множество возможных вопросов пользователю обучающей системы совпадает со множеством возможных запросов к системе, так как обучающая система должна знать и уметь то же, что в конечном итоге должен знать обучающийся.

Частью формализации всякого запроса к системе является указание его обобщенной формулировки на естественном языке, для чего было введено специализированное бинарное отношение *обобщенная формулировка на ЕЯ\**, связывающее вопрос с его обобщенной синтаксической структурой, в которой используются нормализованные словоформы и пропущена формулировка темы.

### III. СИНТАКСИЧЕСКАЯ СТРУКТУРА ПРЕДЛОЖЕНИЯ

Для описания синтаксической структуры предложения необходимо ввести специализированный предметно-ориентированный подязык языка SC – SC-язык представления синтаксических структур. Среди ключевых узлов данного подязыка выделяются узлы, соответствующие классификационным признакам предложений, разновидностям связей в предложениях и словосочетаниях, ролям слов и словосочетаний в предложении. Перечислим некоторые из них:

- Классификационным признакам соответствуют такие узлы, как: *односоставное предложение, утвердительное предложение, сложносочиненное предложение* и др.;
- Разновидностям связей в предложениях соответствуют такие узлы, как: *связь согла-*

сования\*, связь управления\*, связь примыкания\* и их частные разновидности;

- Ролям слов и словосочетаний в предложении соответствуют такие узлы, как: *подлежащее*’, *сказуемое*’, *определение*’, *дополнение*’, *обстоятельство*’ и их частные разновидности.

Приведем описание одного из ключевых узлов на языке SCn[2]:

**связь слабого управления\***

=> пояснение\*:

[Слабое управление имеет место в тех случаях, когда связь между компонентами словосочетания не обусловлена лексико-грамматическими особенностями главного компонента, для реализации значения которого не требуется его распространение.]

∈ Предметная область текстов Русского языка

∈ понятие

∈ отношение

∈ бинарное отношение

∈ ориентированное отношение

С связь управления\*

<= область определения\*:

слово ∪ словосочетание

<= схема отношения\*:

- {
- *главное*’
- *зависимое*’
- }

<= домены\*:

- {
- (слово ∪ словосочетание)
- (слово ∪ словосочетание)
- }

=> библиографический источник\*:

РозентальДЭ\_СовремРЯ

#### IV. РАСШИРЕННАЯ СИНТАКСИЧЕСКАЯ СТРУКТУРА

Расширение и уточнение (в частности, выбор корректной грамматической формы) синтаксической структуры осуществляется на основе структуры связей в предложении, морфологической БЗ, содержащей основы слов и правила словообразования и синтактико-морфологических зависимостей, записанных в базе знаний в виде продукций. Для представления морфологической БЗ в памяти интеллектуальной системы, а также для наращивания синтаксической структуры морфологической информацией необходим еще один специализированный предметно-ориентированный подязык языка SC – SC-язык представления морфологической информации. Ключевые узлы данного подязыка подразделяются на следующие категории:

- узлы, соответствующие классификационной иерархии лексем русского языка, ина-

че говоря, частям речи: *имя существительное*, *имя прилагательное*, *глагол* и др. (включая их частные разновидности и обобщающие категории, например, самостоятельная часть речи);

- узлы ролевых отношений, показывающие, какое место в парадигме лексемы занимает та или иная словоформа: *именительный падеж*’, *мужской род*’ и др.
- узлы, обозначающие классы конкретных реализаций словоформ в виде слов в предложениях: *именительный падеж*, *мужской род* и др.

Приведем формальное описание одного из ключевых узлов на языке SCn:

**pluralia tantum**

= *существительное*, *имеющее только форму множественного числа*

=> пояснение\*:

[Имена существительные, не имеющие формы единственного числа. Некоторые из таких существительных обозначают считае-мые предметы, но их единичность и множественность формой числа не выражается.]

∈ понятие

∈ Предметная область текстов Русского языка

С часть речи

С самостоятельная часть речи

С имя существительное

=> библиографический источник\*:

РозентальДЭ\_СовремРЯ

#### ЗАКЛЮЧЕНИЕ

Преобразование расширенной синтаксической структуры в текст сводится к пословному обходу каждого из предложений со сцеплением слов через пробел, расстановке больших букв в начале предложений и знаков препинания, соответствующих интонационным и смысловым типам предложений, в их конце.

Данная работа выполнялась при поддержке гранта БРФФИ №Ф15PM-073.

1. Шункевич, Д. В. Модели и средства компонентно-го проектирования машин обработки знаний на основе семантических сетей. – В кн.: Междунар. научн.-техн. конф. «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2013). Материалы конф. [Минск, 2013 г.]. – Минск: БГУИР, 2013.
2. Голенков, В. В., Гулякина, Н. А. Графодинамические модели параллельной обработки знаний: принципы построения, реализации и проектирования. – В кн.: Междунар. научн.-техн. конф. «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2012). Материалы конф. [Минск, 16-18 февр. 2012 г.]. – Минск: БГУИР, 2012.
3. Актуальное членение предложения | Энциклопедия Кругосвет [Электронный ресурс]. – 2015. – Режим доступа: [http://www.krugosvet.ru/enc/gumanitarnye\\_nauki/lingvistika/AKTUALNOE\\_CHLENENIE\\_PREDLOZHENIYA.html](http://www.krugosvet.ru/enc/gumanitarnye_nauki/lingvistika/AKTUALNOE_CHLENENIE_PREDLOZHENIYA.html) – Дата доступа: 14.09.2015 г.