

КОМПЬЮТЕРНЫЙ АНАЛИЗ ВОПРОСНО-ОТВЕТНОГО ДИАЛОГА

Богатов Н.М., Родоманов Р.Р.

Кубанский государственный университет

Краснодар, Россия

bogatov@phys.kubsu.ru

Предлагается система анализа текста на естественном языке (САТ), позволяющая оценить ответы экзаменуемого человека. Для этого используется программа лингвистического анализа русскоязычных текстов [1], предназначенная для определения смысла предложения по значению слов.

Основой базы данных слов САТ являются 15 классов. Каждый класс имеет определенный код, состоит из определенной части речи, подразделяется на группы и подгруппы. Группы каждого класса имеют свой код. Код группы каждого класса определяется определенной категорией слова, грамматической категорией, родом, временем, числом и т.д.

Код каждого слова состоит из 8 чисел, первое число определяет класс, второе группу. Остальные шесть, определяют индивидуальный номер слова данного класса. Индивидуальный номер присваивается автоматически в порядке ввода данного слова в базу данных.

Основным свойством каждого класса является сочетаемость, т.е. способность связываться с другими классами. Связь между классами определяется вопросами. На рис.1 отобразено сочетание класса 0 с другими классами.



Рис.1. Связь класса 0 с другими классами

САТ определяет связь между классами с помощью словаря сочетания слов, в котором указан порядок расположения слов в различных сочетаниях. Связь между классами отображена на рисунке 2.

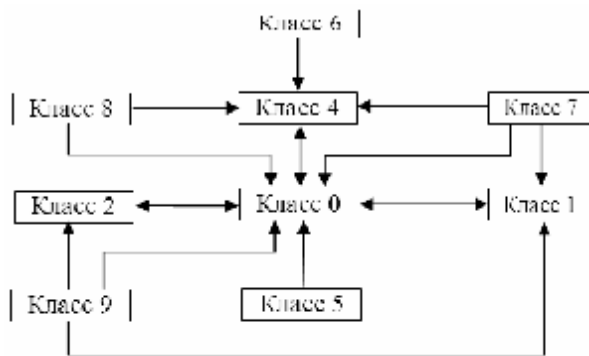


Рис. 2. Связь между классами

Обработка текста происходит в три этапа: морфологический, синтаксический и семантический.

Морфологический анализатор текста делит текст естественного языка на фрагменты и присваивает каждому слову определенный код. Первым шагом разложения текста на фрагменты является таблица в каждой строке, которой отображается номер предложения, номер фрагмента предложения и текст. Далее определяется код класса слова, группа и в зависимости от окончания подгруппа, в которой в данный момент находится слово. Рассмотрим предложение «Железо, введенное внутрь катушки, значительно усиливает магнитное действие катушки». Результатом анализа является таблица, в которой отображается слово, код слова, номера предложения, фрагмента предложения и слова (табл.1).

Таблица 1

Таблица морфологического анализа

№ предл.	№фр.предл.	№слова	Слово	Код слова
0001	00	0001	Железо	021C0001h 0Ah 07h 04h 01h
0001	00	0002	,	
0001	01	0003	Введенное	10000116h 0Bh 08h
0001	01	0004	Внутрь	71000019h 01h
0001	01	0005	Катушки	0113001Ch 0Ah 07h 02h
0001	01	0006	,	
0001	02	0007	Значительно	7400000Dh 01h
0001	02	0008	Усиливает	41210231h 07h
0001	02	0009	Магнитное	100002E2h 0Bh 08h
0001	02	000A	Действие	02010022h 04h 01h
0001	02	000B	Катушки	0113001Ch 0Ah 07h 02h
0001	02	000C	.	

В зависимости от окончания определяются все возможные значения классов слова, групп и их подгрупп. Множество подгрупп слова «железо», объясняется одним и тем же окончанием при различных падежах. Выделить единственно правильное окончание входит в задачу следующего этапа – синтаксического анализа.

В процессе синтаксического анализа происходит анализ возможности сочетания всех слов предложения. Информация о синтаксическом строении предложения представляет собой набор сведений о «главенствовании» одних слов над другими. Словосочетание возникает на основе подчинительной связи. Одним из способов изображения синтаксической структуры предложения является дерево подчинения. Дерево подчинения создается на основе таблицы классов слов.

Во время проверки текста ответа, синтаксический анализатор учитывает данные базы знаний, т.е. при построении дерева подчинения учитываются смысловые отношения между словами с помощью семантической сети. Если сочетания слов обнаружено, создается дерево подчинения, все слова в котором связаны между собой с помощью вопросов. На рис.3 приведен пример дерева подчинения для предложения, расположенного в табл.1.

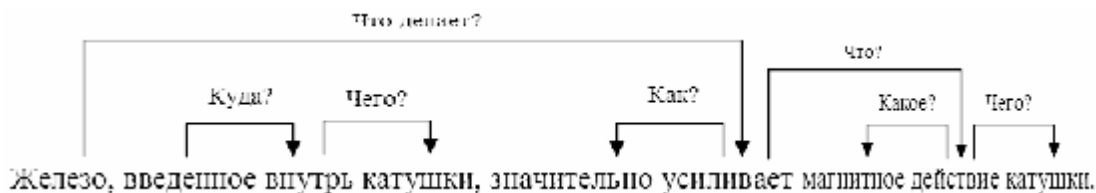


Рис.3. Дерево подчинения предложения

При семантическом анализе происходит поиск всех понятий, отображенных в тексте ответа, в семантической сети, а также возможность замены одних слов на другие при сохранении смысла предложения.

Оценка знаний происходит по смыслу предложения, поэтому ответы на задаваемые вопросы, возможно, вводить в базу знаний компьютера, используя различные термины. Система способна по смыслу заменять слова и их сочетания, поэтому наряду с ответами на конкретные вопросы появляется возможность приводить разъяснения различных терминов.

В большинстве систем обработки текста на естественном языке, используемых в анализаторах ответов, не происходит оценка ответов по смыслу. Предлагаемая система выполняет смысловую обработку ответа, позволяющую вводить различные фрагменты знаний используемых в разных вопросах. Вводимый текст ответа может быть построен в различных комбинациях, отражающих смысл вопроса. Использование предлагаемой системы при самообучении человека, позволяет производить самоконтроль уровня знаний, научиться описывать образ, уменьшить влияние человеческого фактора при тестировании.

ЛИТЕРАТУРА

1. Р.Р. Родоманов, Н.М. Богатов. Программа лингвистического анализа русскоязычных текстов «ПЛАРТ». Свидетельство об официальной регистрации программы для ЭВМ №2005612382 12.09.2005.