

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ РАСПОЗНАВАНИЯ ТЕКСТА НА ЕСТЕСТВЕННОМ ЯЗЫКЕ ПО СМЫСЛУ

Богатов Н.М., Родоманов Р.Р.

Кубанский государственный университет

Краснодар, Россия

bogatov@phys.kubsu.ru

Системы обработки текстов на естественном языке создаются для различных целей: коррекции, машинного перевода, извлечения знаний, вопросно-ответного диалога и др. Компьютерные вопросно-ответные системы контроля знаний для конкретных предметных областей реализуют современные тенденции развития образовательного процесса. Существующие тестовые компьютерные программы проверки знаний основаны на двух принципах:

1. В диалоге человек-машина компьютер предлагает экзаменуемому человеку вопрос и несколько вариантов ответа. Человек должен выбрать правильный ответ. В процессе работы такая система может вместо укрепления знаний внести в память человека текст с неправильным ответом из выбираемых вариантов.
2. Машина задает вопрос, человек должен ответить на него, изложив свои мысли на естественном языке. Ответ должен быть сформирован ограниченным количеством лексем. В существующих анализаторах ответов грамматика естественного языка либо полностью игнорирована, либо используется жесткая “грамматика” в рамках строго формализованного обучающего курса. Системы обработки текста на естественном языке, предназначенные для анализа ответов и их оценки по смыслу, ещё не созданы. Компьютерное моделирование смыслового распознавания текста является важным этапом разработки таких систем.

Целью работы является создание вопросно-ответной системы (ВОС), способной оценивать знание, обрабатывая ответы в текстовой форме на естественном языке по смыслу. Для этого создана программа лингвистического анализа русскоязычных текстов [1].

В процессе диалога ВОС задает вопрос в виде предложения, а экзаменуемый должен дать ответ в виде текста. После ввода ответа, происходит поиск заданного вопроса в базе знаний и моделирование всех возможных смысловых вариантов, построенных из слов в предложении ответа. Результаты моделирования смыслового распознавания ответа рассмотрим на примере вопроса: «Что называется электрическим током?». Ответы на данный вопрос могут быть отображены различными словами, при сохранении смысла. Приведем несколько возможных ответов.

1. Упорядоченное движение заряженных частиц.

2. Направленное движение электронов.
3. Направленное движение электрических зарядов.

При распознавании используется семантическая сеть, фрагмент которой показан на рис.1. Первый вариант ответа полностью совпадает с определением в семантической сети и поэтому оценивается машиной как полный. Во втором варианте ответа понятие «упорядоченный» и «заряженная частица» отображаются соответственно другими словами «направленный», «электрон». В семантической сети понятие «электрон» входит в подгруппу «заряженная частица», а «направленный» – в подгруппу «упорядоченный». Поскольку кроме слова «электрон» можно употребить другие понятия, входящие в группу «заряженная частица», то ответ расценивается как правильный, но не полный. В третьем варианте ответа слова «упорядоченное» и «направленное», а также «заряженные частицы» и «электрические заряды» взаимозаменяемы соответственно. Поэтому третий вариант будет расценен как правильный. Если в тексте ответа обнаружены слова, не входящие в подгруппы ответа, то ответ расценивается как неправильный.

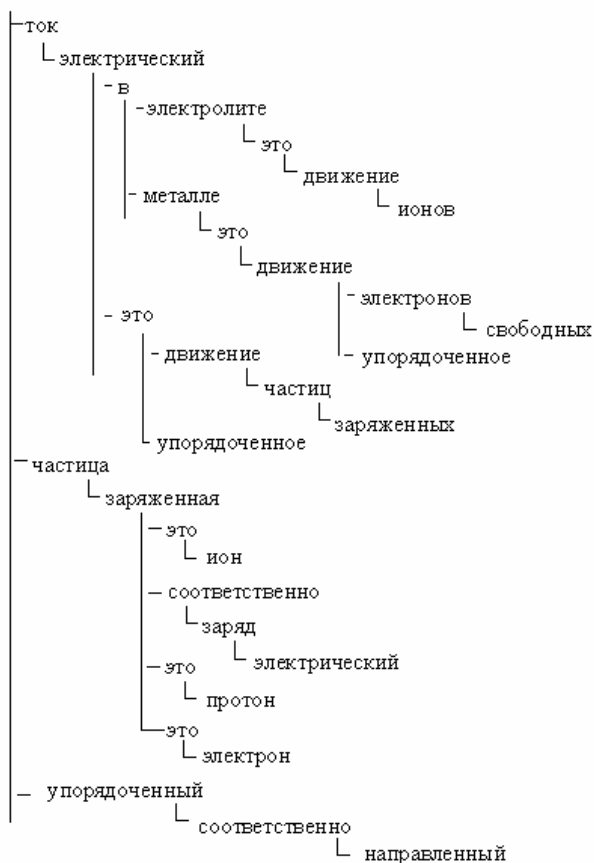


Рис.1. Фрагмент семантической сети ВОС

Используя фрагмент семантической сети, ВОС может задать в режиме диалога несколько вопросов экзаменуемому человеку и проанализировать ответы. Вопросы определяются понятием, которое отображено в семантической сети до слова «это».

Например: Что представляет собой электрический ток в электролитах? Что представляет собой электрический ток в металле?

Моделирование смыслового распознавания текста позволяет оценивать ответы, сформулированные различными терминами. Система способна по смыслу заменять слова и их сочетания, поэтому наряду с ответами на конкретные вопросы появляется возможность приводить разъяснения различных терминов.

Основным достоинством предлагаемой системы является смысловая обработка ответа, позволяющая вводить различные фрагменты знаний, используемых в разных вопросах. Вводимый текст ответа может быть построен в различных комбинациях, отражающих смысл заданного вопроса. Использование предлагаемой системы при самообучении человека даёт возможность осуществлять самоконтроль уровня знаний, научиться описывать образ, а на экзамене позволяет уменьшить влияние человеческого фактора.

Вопросно-ответная система применяется в учебном процессе для тестирования студентов физико-технического факультета Кубанского государственного университета.

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.