

Е. С. Василенко, Н. В. Шалимо

СЕМАНТИЧЕСКИЕ СЕТИ В ЗАДАЧАХ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Обработка естественного языка – общее направление искусственного интеллекта и математической лингвистики, изучающее проблемы компьютерного анализа и синтеза естественных языков. Методы обработки естественного языка лежат в основе технологий распознавания речи, информационного поиска, средств проверки правописания, систем общения, и др.

Основные трудности в обработке естественного языка вызваны проблемой его многозначности, выражающейся на всех ее стадиях обработки: от фонетического до семантического уровня с точки зрения лингвистической теории «Смысл \Leftrightarrow Текст». Таким образом, методы обработки естественного

языка направлены на разрешение многозначности в различных ее проявлениях. Например, смысл предложения *Я натисся из ключа* зависит от того, в каком значении употреблено многозначное слово *ключ*.

Сегодня технологии обработки естественного языка прочно вошли в повседневную жизнь и помогают людям лучше понимать друг друга и быстрее ориентироваться по поступающей информации. Этому способствуют технологии машинного перевода, анализа эмоциональной окраски текстов, автоматического реферирования документов, распознавания и синтеза речи и т.д. Несмотря на высокую популярность статистических методов обработки естественного языка, существуют задачи, для решения которых требуются знания об окружающем мире. Так, разрешение лексической многозначности, построение вопросно-ответных систем, автоматическая рубрикация документов и др. Решение таких задач производится при помощи систем, основанных на знаниях.

На заре инженерии знаний и обработки естественного языка под семантической сетью понимался размеченный ориентированный граф, вершины которого соответствуют некоторым сущностям (понятиям, событиям, характеристикам или значениям), а дуги выражают связи между этими сущностями. Замечено, что люди воспринимают окружающий мир как иерархию понятий, связанных отношениями общего и частного. Например, если человек знает, что воробей – это птица, то он сможет предположить, что у него есть крылья.

Семантическая сеть – это ориентированный граф, вершины которого – понятия, а дуги – связи между понятиями.

Семантические сети являются одной из форм представления знаний. В них представлены различные семантические отношения между понятиями, которые бывают симметричными и асимметричными. Примерами симметричных семантических отношений являются синонимия и антонимия, асимметричных – ‘часть–целое’, ‘причина–следствие’, ‘противоположность’, и др.

Сегодня наиболее известной семантической сетью в области обработки естественного языка является **WordNet**, построенная на основе формализации человеческого восприятия окружающего мира. Основная задача, в которой широко используются семантические сети, – разрешение лексической многозначности.

Семантические сети используются в современных методах машинного перевода. Поскольку в задаче машинного перевода необходимо не только правильно определить лексическое значение слова на исходном языке, но и корректно выбрать лексическое значение слова на целевом языке, семантические сети применяются в качестве инвентаря смыслов. Кроме того, поскольку в таких семантических сетях, как WordNet, к синсетам представлены толкования и примеры употребления, то эти сведения также используются для построения векторных представлений значений слов в машинном переводе.

Семантические сети применяются для решения задачи автоматической рубрикации текстовых документов. В этом случае используются знания о

синонимах и лексических отношениях, описанных в семантических сетях. На основе этих знаний производится построение классификатора при помощи машинного обучения с учителем. Классификатор используется для назначения класса ранее неизвестным документам. Как правило, высококачественные семантические сети, применяемые для решения задач обработки естественного языка, создаются коллективами экспертов-лексикографов.

Процесс работы экспертов трудноформализуем, однако отмечаются его общие этапы:

- отбираются разные примеры употребления слова в текстах;
- значения слов разделяются и группируются в понятия;
- строятся семантические отношения между понятиями.