

Авторы:

М.М. Калугина, А.С. Яркеев, А.Д. Пашнин, Университет ИТМО, Санкт-Петербург

Научный руководитель:

Цопа Евгений Алексеевич, Университет ИТМО, Санкт-Петербург

Тезис доклада:

Оценка пертинентности результатов работы алгоритма поиска экземпляров в тексте на естественном языке

На сегодняшний день актуальной задачей при решении целого ряда проблем является извлечение информации из текстов на естественном языке. Для решения такого рода задач разрабатывается множество алгоритмов, отсюда возникает необходимость создания инструментов оценки пертинентности существующих решений.

В работе авторами предложен подход к оценке пертинентности результатов работы алгоритма извлечения экземпляров понятий из текстов на естественном языке. Подход базируется на использовании семантической сети, содержащей информацию о смысловых значениях, способе их выражения в тексте, а также экземплярах этих смысловых значений. Данная информация используется для генерации тестового набора входных значений, для проверки результата работы алгоритма и определения его пертинентности.

Авторами был разработан инструмент, реализующий упомянутый ранее подход. Инструмент представляет собой программный модуль на языке C++, интегрированный с модулем семантической сети и модулем поиска экземпляров. Программный модуль генерирует исходные данные, моделируя различные ситуации такие, как опечатки, зашумление исходной строки, пропуск словоформ, определяющих смысловое значение и т. д.

Разработанный инструмент был успешно опробован при тестировании алгоритма поиска экземпляров адресной информации в текстах на естественном языке с использованием подготовленной семантической сети.