Отзыв

на магистерскую диссертацию Емельяновой Анастасии Вадимовны

«Процедура семантического отображения номинативного компонента текста в терминах WordNet-словарей»

Рецензируемая диссертация посвящена вечно актуальной теме – автоматическому анализу семантики текста. На материале обучающей выборки из 250 отзывов потребителей на цифровые фотоаппараты строится процедура извлечения информации двух видов – тематической и оценочной, которая затем тестируется на 20 отзывах. Задача автора усложняется тем, что отзывы, размещенные в интернете, носят неформальный характер: во-первых, наряду с нейтральной лексикой, в них используются профессиона­лизмы и жаргонизмы, а, во-вторых, – авторская орфография.

Работа состоит из введения, трех глав, заключения, списка литературы и 9 прило­жений.

Во введении формулируются цели, задачи и методы исследования, обосновываются его акту­альность и практическая значимость.

В первой главе «Интерпретация текста при автоматическом анализе» формулируется подход, который будет использоваться автором при решении поставленной задачи. Учитывая описательный характер текстов, на которых строится исследование, автор в качестве основного элемента текста выбирает номинативный компонент исследования, а логико-понятийную схему, используемую в Wordnet, как способ его представления. Основываясь на списке использующихся на практике алгоритмов семантического анализа, автор принимает решение строить свой алгоритм на словаре и правилах.

Вторая глава посвящена подготовке лингвистического обеспечения алгоритма интерпретации двух составляющих текста – тематической и тональной.

Для интерпретации тематической составляющей на основании статей Википедии, словаря терминов «Яндекс.Маркет» и 250 отзывово покупателей были построены логико-понятийные схемы «Цифровые фотоаппараты», в которых термины сгруппированы в синонимические ряды (синсеты), между которыми установлен ряд отношений. В первой из названных ЛПС 440 узлов, связанных 14 отношениями, во второй – 266 узлов и 7 отношений, в третьей – 224 узла и 11 отношений. На основании ЛПС строится структура фрейма описания цифрового фотоаппарата.

Для интерпретации тональной составляющей на основании анализа обучающей вы­борки отзывов выделено более тысячи тональных маркеров и построено параметри­че­ское поле, большинство параметров которого соотносится с узлами ЛПС. Для оценки тональ­ности текста применяется нейронная сеть, учитываются и отрицательные и положи­тельные оценки, причем принимается во внимание значимость того или иного параметра.

В третьей главе описывается собственно процедура интерпретации текста, реализованная в виде программы на языке С++. Процедура протестирована на 20 отзывах. Точность выделения тематических маркеров составила около 91,4%, полнота 86,6%.

В конце каждой главы суммируются основные выводы. В заключении подводятся итоги исследования.

Список литературы содержит 76 наименований, в том числе 10 на английском языке.

В приложениях приводятся вспомогательные материалы и результаты отдельных этапов работы.

Диссертация представляет собой в высшей степени интересное исследование в области прикладной лингвистики и не оставляет никаких сомнений в высокой квалификации автора в данной области. Все поставленные задачи решены, полученные результаты могут быть использованы, например, маркетологами для автоматического подробного анализа мнений покупателей. Работа написана живым языком и производит очень благоприятное впечатление своей обстоятельностью и интересным анализом материала. Некоторые замечания и вопросы, возникшие по ходу чтения работы, не влияют на общую оценку:

1. Не хватает определений таких понятий, как параметрическое поле и нейронная сеть.

2. Мне кажется, было бы удобнее для читателя, если бы работа алгоритма была представлена не текстом, а схемой.

3. Недостаточно объяснены формулы, по которым проводится расчет тональности отзыва.

4. Очень разумно все громоздкие материалы вынесены в приложение, но в ряде случаев по ходу чтения работы не хватало иллюстративных фрагментов этих материалов собственно в тексте.

Магистерская диссертация Анастасии Вадимовны Емельяновой безусловно отвечает всем требованиям, предъявляемым к такого рода работам, и заслуживает высокой оценки.

18.06.2015

Алексеева Е.Л., доц. каф. математической лингвистики