



ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ OPINION MINING

Терещенко В.И.

*Национальный технический университет
«Харьковский политехнический институт»,
г. Харьков, ул. Фрунзе 21,
тел. (057) 700-15-64,
e-mail: omsroot@kpi.kharkov.ua*

Сегодня все большую популярность в развитии семантических сетей и автоматическом анализе текстовой информации приобретает такое направление как Opinion Mining. Не смотря на то что этим вопросом занимается очень большое количество лингвистов и программистов, задача оценки мнения текста остается до конца нерешенной. В связи с тем, что каждый день в интернете накапливается все больше информации в текстовом виде относительно того или иного явления или продукта, появилась потребность в средствах быстрого анализа информации и получения основных данных, представляющих наиболее сжатую и в то же время, информативную выжимку из текста. Например при покупке телефона, поиск информации в интернете об определенной модели и существующих о ней мнений может занять от нескольких часов до нескольких дней. Однако даже с такими затратами времени и сил, пользователь не может быть до конца уверенным в своем решении поскольку полученная информация является не полной и ограничивается несколькими десятками сайтов или форумов. Методы и технологии Opinion Mining являются более эффективными, поскольку позволяют проанализировать большие объемы информации из множества различных источников (от нескольких сотен до нескольких тысяч) за короткий срок времени [1].

Для релевантного определения мнения текста, технологии Opinion Mining наряду с задачей определения мнения текстов, должны включать в себя задачу обработки семантических эквивалентов. Поиск семантических эквивалентов неразрывно связан с задачами Opinion Mining, поскольку он позволяет находить в тексте слова с эквивалентным значением [2]. Хотелось бы отметить, что частично эти задачи решаются с помощью онтологий и тезаурусов [3].

Предметом данного исследования являются онтологии и тезаурусы с помощью которых в дальнейшем стало бы возможным приблизиться к решению проблем Opinion mining.

В ходе исследования были проанализированы такие online-библиотеки онтологий как "DAML ontology library", "Protégé ontologies library", "Code4Lib library ontology", содержащие большое количество онтологий разных направлений. После проведенного анализа, для дальнейшего исследования были выбраны следующие онтологии: WordNet, Ruthes, Gemet, SentiWordNet.

В рамках исследования проведен сравнительный анализ этих онтологий, связей используемых в них а также проанализирована возможность их



использования для решения задачи Opinion Mining и поиска семантических эквивалентов. В результате исследования получены следующие выводы:

- в онтологии SentiWordNet, наилучшим образом определены связи используемые для решения задачи Opinion Mining;
- в онтологии WordNet определены связи для поиска семантических эквивалентов.

Однако эти онтологии не решают поставленные задачи полностью. SentiWordNet как и WordNet можно использовать только для слов и некоторых словосочетаний, однако остаются почти бесполезными для анализа текстовой информации, как то новостные рассылки или другие её источники.

В связи с этим, был сделан вывод о том что для решения задач Opinion Mining и поиска семантических эквивалентов наиболее действенным будет разработка собственного тезауруса или онтологии на основе существующих и с использованием основных принципов и методов их построения.

Список литературы

1. *Большакова Е.И., Баева Н.В., Васильева Н.Э.* Структурирование и извлечение знаний, представленных в научных текстах / Девятая Национальная конференция по искусственному интеллекту с международным участием КИИ-2004. Труды конференции в 3-х томах. Т. 2. - Москва: Физматлит, 2004. – 480-488 с.
2. *Пэнг Б., Ли Л.* Извлечение мнений и смысловой анализ. Т.2. / Б.Пэнг, Л. Ли. - Foundations and Trends in Information Retrieval., 2008.-135 с.
3. *Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В.* Онтологии и тезаурусы: Учебное пособие / В.Д. Соловьев, Б.В. Добров, В.В. Иванов, Н.В. Лукашевич. – Казань: Москва, 2006. – 157 с.