

ОТ СОСТАВИТЕЛЯ

Настоящий выпуск журнала «Электронные библиотеки» представляет собой тематический сборник статей, посвященный проблеме автоматического анализа тональности текстов на русском языке.

Задача анализ тональности состоит в автоматическом определении отношения автора текста (позитивном, негативном или нейтральном) к объектам и ситуациям, о которых говорится в анализируемом тексте. В настоящее время автоматический анализ тональности используется в самых различных приложениях, включая мониторинг репутации компаний и публичных персон, анализ общественных настроений в том или ином регионе, анализ сообществ в социальных сетях и многое другое.

В данном тематическом выпуске представлены статьи участников открытого тестирования систем анализа тональности на русском языке SentiRuEval, проведенном в 2014–2015 годах. В данном тестировании участникам были предложены для решения две основные задачи.

Первая задача состояла в автоматическом анализе отзывов пользователей в двух предметных областях (рестораны и автомобили) с целью определить основные характеристики обсуждаемых объектов (так называемые аспекты, например, салат, интерьер для ресторанов) и их оценку пользователем – автором отзыва.

Вторая задача заключалась в анализе постов Твиттера (твитов) для мониторинга репутации организаций в заданной сфере деятельности (банки и телекоммуникационные компании). Данная задача включает как выявление положительного или отрицательного отношения авторов твитов к заданным организациям, так и оценку распространяемых в Твиттере позитивных или негативных новостей об этих организациях.

В статье Н.В. Лукашевич (НИВЦ МГУ им. М.В. Ломоносова) «Автоматический анализ текстов по отношению к заданному объекту и его характеристикам» представлен обзор задач, возникающих в рамках анализа тональности текстов по аспектам. Представлены особенности предлагаемых подходов и достигаемые ими характеристики качества.

Статья П.Д. Блинова и Е.В. Котельникова (Вятский государственный гуманитарный университет) «Семантическое сходство в задаче аспектно-эмоционального анализа» описывает совокупность подходов к анализу тональности текстов по аспектам, начиная с извлечения аспектов, их дальнейшей классификации и

определению тональности. Подход к извлечению аспектов сущности основан на выявлении контекстов употребления слов, представления их в виде векторов и дальнейшем группировании этих слов в аспектные категории.

В статье группы авторов из Казанского федерального университета (Е.В. Тутубалина, В.В. Иванов, М.А. Загулова, Н.Р. Мингазов, И.С. Алимова, В.А. Малых) представлены подходы на основе методов машинного обучения к обоим задачам SentiRuEval: анализ отзывов и анализ твитов. Подробно описаны признаки, используемые в применяемых методах машинного обучения, их модификации в конкретных задачах, а также проведен анализ ошибок.

В статье Ю.В. Адаскиной, П.В. Паничевой и А.М. Попова (ООО «InfoQubes», Санкт-Петербургский государственный университет) исследуется вклад синтаксического анализа в задаче анализа тональности твитов. Для этого проводится синтаксический анализ твитов, получившаяся синтаксическая структура преобразуется в тройки вида (отношение, слово1, слово2), и затем эти тройки используются как дополнительные признаки для системы классификации.

Статья П.Ю. Полякова, М.В. Калининой, В.В. Плешко (ООО «ЭР СИ О») посвящена рассмотрению лингвистико-инженерного подхода к анализу тональности твитов, включающего использование словаря оценочных слов, синтаксического анализатора, а также правил вычисления тональности на основе проведенного анализа.

В статье Ю.В. Рубцовой и С.А. Кошельникова (Институт систем информатики им. А.П. Ершова СО РАН) рассматриваются особенности применения известного метода машинного обучения CRF для анализа тональности твитов, анализируются ошибки полученного классификатора.

Нужно отметить, что мировая практика научных исследований в области автоматической обработки текстов свидетельствует о важности открытых тестирований типа SentiRuEval, в результате которых выявляются и получают большее распространение лучшие подходы, в целом ускоряется развитие автоматических систем. Поэтому практика проведения открытых тестирований становится все более распространенной в мире, в России также будут продолжаться такого рода тестирования автоматического анализа текстов на русском языке.

Н.В. Лукашевич