

УДК 004+81.13(32)

**АЛГОРИТМЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ
СИСТЕМ ОБРАБОТКИ ТЕКСТОВОЙ ИНФОРМАЦИИ ДЛЯ
ЗАДАЧИ АНАЛИЗА МНЕНИЙ**

Попов О. Р.

к.т.н., доцент, доцент каф. ИТиПМ

Южный университет (ИУБИП)

e-mail: cs41825@aaanet.ru

Гребенюк Е. В.

Сургутский государственный университет

e-mail: pev_86@mail.ru

Аннотация: Несмотря на разноплановые эксперименты в построении интеллектуальных систем обработки текстовой информации, задача анализа мнений остается актуальной. В статье представлены теоретические аспекты автоматической обработки текстов для реализации задачи анализа мнений и перечислены подходы для выбора машинного алгоритма. Показано, что система, которая проводит данный анализ будет интеллектуальной. Рассмотрены понятийный аппарат и основные признаки интеллектуальных систем.

В статье также рассматриваются алгоритмы автоматической обработки текстов для задачи анализа мнений. Автоматическое определение тональности текста предполагает выделение тех частей текста, которые имеют позитивную или негативную оценку по отношению к объекту тональности. Существуют разные подходы к машинной обработке по определению тональности текста. Представлены алгоритмы построения интеллектуальной обработки, получившие научное обоснование, наибольшее распространение и реализацию в программных продуктах. Подобные алгоритмы можно реализовать с помощью специального программного обеспечения: математических пакетов символьных или численных вычислений.

Ключевые слова: анализ тональности текста, анализ мнений, интеллектуальные системы, алгоритмы машинной обработки.

**ALGORITHMS FOR CONSTRUCTING INTELLIGENT SYSTEMS FOR
PROCESSING TEXT INFORMATION FOR THE TASK OF
ANALYZING OPINIONS**

Popov O.R.

Abstract: Despite the diverse experiments in the construction of intelligent systems for processing text information, the task of analyzing opinions remains relevant. The article presents the theoretical aspects of automatic text processing for the implementation of the problem of opinion analysis and lists the approaches for choosing a machine algorithm. It is shown that the system that performs this analysis will be intelligent. The conceptual framework and the main features of intelligent systems are considered.

The article also discusses algorithms for automatic text processing for the task of analyzing opinions. Automatic detection of the tonality of the text involves the selection of those parts of the text that have a positive or negative assessment in relation to the object of the tonality. There are different approaches to machine processing to determine the tonality of a text. The algorithms for constructing intelligent processing, which have received scientific justification, the most widespread and implemented in software products, are presented. Such algorithms can be implemented using special software: mathematical packages for symbolic or numerical calculations.

Keywords: text sentiment analysis, opinion analysis, intelligent systems, machine processing algorithms.

В современном мире анализ текстовой информации применяется в разных предметных областях: социальные сети, маркетинговые исследования, автороведческие исследования, определение эмоциональной характеристики текста и т.д. Несмотря на разноплановые эксперименты в этом направлении, задача анализа мнений остается нетривиальной.

Стоит отметить существенный вклад в разработку методов и алгоритмов анализа текстовой информации таких ученых как, HuM., LiuB., Васильев В. Г., Вежбицкая А., Ионова С.В., Клековкина М. В., Колосова (Солодовникова) Н.Г., Котельников Е. В., Лукашевич Н.В., Рубцова Ю.В., Четверкин И. И., Шаховский В. И. и др.

Анализ тональности текста подразумевает, что система, которая проводит этот самый анализ будет интеллектуальной [1,3,11,14]. «Интеллектуальная система — совокупность взаимодействующих между собой относительно элементарных структур и процессов, объединенных в целое выполнением функции интеллекта (целенаправленного, опосредованного и обобщенного познания, активного отражения

объективной реальности, логического и творческого мышления), несводимой к функции ее компонентов»[5].

«Тональность — это эмоциональное отношение автора высказывания к некоторому объекту (объекту реального мира, событию, процессу или их свойствам/атрибутам), выраженное в тексте»[9].

Тональность текста — выраженная в тексте эмоциональная оценка (сентимент текста) [7,8,10,15].

Обобщенное определение, выглядит следующим образом, анализ тональности (sentiment analysis, sentiment detection) или анализ мнений (opinion mining) или эмоциональная окраска текста — обработка естественного языка (NLP, natural language processing), целью которого является извлечение эмоционального содержания из текста; «представляет собой класс задач в компьютерной лингвистике по автоматизированному выявлению в тексте эмоционально окрашенной лексики и эмоциональной оценки текста (автором мнений)»[9].

«Признаки интеллектуальной системы включают в себя взаимодействие со средой и другими системами как единое целое, а также состоит из иерархии подсистем более низкого уровня»[5].

У истоков описания эволюции интеллектуальных систем стоит теория ноогенеза интеллектуальных систем Еремина А. Л. В том числе, в его работах, дается определение «экология интеллектуальных систем - логическое обобщение опыта и закономерностей взаимодействия интеллектуальных систем с окружающей средой»[5]. Это определение относится к специальному разделу знаний – «информационной экологии» или «информационной гигиене».

Автоматическое определение тональности текста предполагает выделение тех частей текста, которые имеют позитивную или негативную

оценку по отношению к объекту тональности. Объект оценки может быть как один для всего текста, так их может быть и несколько [4,6].

При этом отношение автора может быть выражено: положительно, отрицательно, нейтрально[7, 12].

Существуют разные подходы к определению тональности текста. Наиболее распространенными являются:

Для автоматического определения тональности текста можно выделить следующие подходы:

- 1) на основе правил с использованием шаблонов;
- 2) машинное обучение без учителя;
- 3) машинное обучение с учителем;
- 4) гибридный метод [7];
- 5) подходы, основанные на правилах;
- 6) подходы, основанные на словарях [2].

«Соотношение информации и её языкового выражения предполагает следующие их виды:

- позитивная информация и её позитивная языковая репрезентация;
- негативная информация и её негативная языковая репрезентация;
- позитивная информация и её негативная языковая репрезентация;
- негативная информация и её позитивная языковая репрезентация»[13].

Для решения классификации эмоций используют такие способы, как свёрточные нейронные сети (ConvolutionalNeuralNetworks), в которых векторные представления слов формируются на основе обученной модели [1].

Автоматическое определение тональности текста могут представлять собой облака слов или тегов, т.е. визуальное представление слов и словосочетаний, с выделением их частотности. Визуализация может быть

представлена как точечная текстовая диаграмма, которая показывает относительную частоту слов, ранжируя их по размеру и цвету. Кроме этого, возможна визуализация кластеров в точечной диаграмме текста, посредством векторного представления слов.

Для построения облаков слов или точечных диаграмм текста алгоритм будет выглядеть следующим образом:

- импорт и визуализация текстовых данных;
- извлечение текстовых данных;
- визуализация текста.
- Для получения векторного представления слов алгоритм выглядит:
- преобразование текста к численному формату;
- подсчет слов и N-грамм;
- векторное представление и кодирование словарно.

Подобные алгоритмы можно реализовать с помощью специального программного обеспечения. Например, математических пакетов символьных (Mathematica) или численных (MatLab) вычислений.

В результате проведенной работы представлены теоретические аспекты автоматической обработки текстов для реализации задачи анализа мнений и перечислены подходы для выбора машинного алгоритма.

Библиографический список

1. Hu M., Liu B. Mining and Summarizing Customer Reviews // Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2004. P. 168–177. [Электронный ресурс]. – URL: https://www.researchgate.net/publication/221654582_Mining_and_summarizing_customer_reviews(дата обращения: 18.02.2021).
2. Landowska A., Towards New Mappings between Emotion Representation Models Department of Software Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland Appl. Sci.2018,8(2), 274; [Электронный ресурс]. – URL: <https://doi.org/10.3390/app8020274> (дата обращения: 15.12.2020).
3. Liu B. Sentiment Analysis and Opinion Mining // Synthesis Lectures on Human Language Technologies. – 2012. – Vol. 5(1). – P. 16 [Электронный ресурс]. – URL:

<https://www.morganclaypool.com/doi/abs/10.2200/S00416ED1V01Y201204HLT016>(дата обращения: 20.02.2021).

4. Вежбицкая А. Семантика, культура и познание: Общечеловеческие понятия в культуроспецифических контекстах // Thesis. — Вып. 3. — М., 1993. — С. 185—206. [Электронный ресурс]. — URL: <http://www.philology.ru/linguistics1/wierzbicka-93.htm>(дата обращения: 28.11.2020).

5. Еремин А.Л. Ноогенез и теория интеллекта. — Краснодар: СовКуб, 2005. — 356 с. [Электронный ресурс]. — URL: <http://a-eremin.ru/rus/themes/physiology/ecophis/?id=16> (дата обращения: 27.10.2020).

6. Ионова С.В., Шаховский В.И. Проспекция лингвокультурологической теории эмоций Анны Вежбицкой // Вестник Российского университета дружбы народов. Сер. Лингвистика. - 2018. - Т. 22, № 4. - С. 966-987. [Электронный ресурс]. URL: https://openresearch-repository.anu.edu.au/bitstream/1885/164313/2/01_Gladkova_Anna_Wierzbicka%252C_language%252C_2019.pdf(дата обращения: 27.10.2020).

7. Клековкина М.В., Котельников Е.В. Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики / Вятский государственный гуманитарный университет, Киров // Труды 14-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2012, Переяславль-Залесский, Россия, 15-18 октября 2012 г. [Электронный ресурс]. — URL: <http://ceur-ws.org/Vol-934/paper15.pdf>(дата обращения: 27.10.2020).

8. Колосова (Солодовникова) Н.Г. Эмотивный анализ как метод экологического мониторинга текстов // Изв. Волгогр. гос. пед. ун-та. — 2015. — № 5(100). — С. 103-110. [Электронный ресурс]. — URL: <https://cyberleninka.ru/article/n/emotivnyy-analiz-kak-metod-ekologicheskogo-monitoringa-tekstov>(дата обращения: 02.11.2020).

9. Лукашевич Н.В., Четверкин И.И., Комбинирование тезаурусных и корпусных знаний для извлечения оценочных слов // Системы и средства информ., — 2015. —Т. 25, Вып. 1. — С. 20–33 [Электронный ресурс]. — URL: DOI: <https://doi.org/10.14357/08696527150102> (дата обращения: 15.12.2020).

10. Лукашевич Н.В. Модели и методы автоматической обработки неструктурированной информации на основе базы знаний онтологического типа: дис. ... докт. тех. наук. — М., 2014. — 312 с.: ил. [Электронный ресурс]. — URL: <https://www.dissercat.com/content/modeli-i-metody-avtomaticheskoi-obrabotki-nestrukturirovannoi-informatsii-na-osnove-bazy-zna>(дата обращения: 15.12.2020).

11. Рубцова Ю.В. Методы и алгоритмы построения информационных систем для классификации текстов социальных сетей по тональности: дисс. ... канд. тех. наук. — Новосибирск, 2019.

12. Шаховский В. И. Обоснование лингвистической теории эмоций / В. И. Шаховский // Вопросы психолингвистики. — 2019. — № 1 (39). С. 22-37. [Электронный ресурс]. — URL: <https://iling-ran.ru/library/voprosy/39/VPL-1-2019-all-20-37.pdf>(дата обращения: 15.12.2020).

13. Шаховский В.И. Эмотивная лингвоэкология: комплексный подход к изучению языка, речевой деятельности и человека. — Волгоград, 2013. — 450 с.

14. Попов, О.Р. Выделение и анализ векторов самоорганизации интеллектуальных систем / О.Р. Попов, Г.И. Акперов // Интеллектуальные ресурсы — региональному развитию. — 2020. — № 2. — С. 48-54. — URL: <https://www.elibrary.ru/item.asp?id=19492269> (дата обращения: 28.03.2021).

15. Галоян, Я.Э. Стилистические средства и неоднозначность синтаксических конструкций / Я.Э. Галоян, М.Ю. Масалова // Интеллектуальные ресурсы – региональному развитию. – 2018. – Т. 4, № 1. – С. 268-271. – URL: <https://www.elibrary.ru/item.asp?id=36739546> (дата обращения: 10.04.2021).