

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра компьютерных технологий и систем

Аннотация к дипломной работе

Определение тональности текстов на разных языках

Кутас Елизавета Анатольевна

Научные руководители – доцент кафедры КТС, кандидат физ.-мат. наук
Дайняк В.В., ассистент кафедры КТС Усатов А.А.

Минск, 2022

РЕФЕРАТ

Дипломная работа, 57 стр., 9 рис., 3 табл., 47 источников.

Ключевые слова: ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, КЛАССИФИКАЦИЯ ТЕКСТА, АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТА, СЛОВАРИ ТОНАЛЬНОСТИ, МАШИННОЕ ОБУЧЕНИЕ, ПРЕОБРАЗОВАНИЕ ТЕКСТОВЫХ ДАННЫХ, ПРЕДСТАВЛЕНИЕ ТЕКСТОВЫХ ДАННЫХ, ОЦЕНКА КАЧЕСТВА РАБОТЫ АЛГОРИТМА.

Объект исследования: методы и программные средства определения тональности текстов.

Цель работы: изучить задачи классификации и определения тональности текстов, рассмотреть и описать основные этапы классификации, изучить основные методы классификации тональности, рассмотреть существующие программные решения для анализа тональности текстов. Разработать метод определения тональности текстов на разных языках и реализовать его.

Методы проведения работы: теоретический анализ литературы по теме исследования; подробное описание и классификация методов, применяемых в исследуемой области; технический анализ существующих решений, относящихся к теме исследования, и их программной документации; проведение вычислительного эксперимента; описание и сравнение полученных результатов.

Результаты работы: изучены задачи классификации и определения тональности текстов, основные этапы классификации, подходы к классификации и методы классификации тональности. Проанализированы существующие программные решения для анализа тональности текстов. На основе полученных знаний разработан, реализован и протестирован собственный метод автоматического определения тональности текстовых сообщений на разных языках.

Область применения результатов: в управлении, политологии, социальных науках, экономике, в прогнозировании показателей продаж, в изучении торговых стратегий, в исследованиях социальных медиа.

Дипломная работа выполнена автором самостоятельно.

РЭФЕРАТ

Дыпломная праца, 57 ст., 9 мал., 3 табл., 47 крыніц.

Ключавыя словы: АПРАЦОЎКА НАТУРАЛЬнай мовы, КЛАСІФІКАЦЫЯ ТЭКСТУ, АНАЛІЗ ТАНАЛЬНАСЦІ ТЭКСТУ, СЛОЎНІКІ ТАНАЛЬНАСЦІ, МАШЫННАЕ НАВУЧАННЕ, ПЕРАЎТВАРЭННЕ ТЭКСТАВЫХ ДАНЫХ, ПРАДСТАЎЛЕННЕ ТЭКСТАВЫХ ДАНЫХ, АЦЭНКА ЯКАСЦІ ПРАЦЫ АЛГАРЫТМУ.

Аб'ект даследавання: метады і праграмныя сродкі вызначэння танальнасці тэкстаў.

Мэта працы: вывучыць задачы класіфікацыі і вызначэння танальнасці тэкстаў, разгледзець і апісаць асноўныя этапы класіфікацыі, вывучыць асноўныя метады класіфікацыі танальнасці, разгледзець існуючыя праграмныя рашэнні для аналізу танальнасці тэкстаў. Распрацаваць метады вызначэння танальнасці тэкстаў на розных мовах і рэалізаваць яго.

Метады правядзення працы: тэарэтычны аналіз літаратуры па тэме даследавання; падрабязнае апісанне і класіфікацыя метадаў, якія выкарыстоўваюцца ў вобласці даследавання; тэхнічны аналіз існуючых рашэнняў, звязаных з тэмай даследавання, і іх праграмай дакументацыі; правядзенне вылічальнага эксперыменту; апісанне і параўнанне атрыманых вынікаў.

Вынікі працы: вывучаны задачы класіфікацыі і вызначэння танальнасці тэкстаў, асноўныя этапы класіфікацыі, падыходы да класіфікацыі і метады класіфікацыі танальнасці. Прааналізаваны існуючыя праграмныя рашэнні для аналізу танальнасці тэкстаў. На аснове атрыманых ведаў распрацаваны, рэалізаваны і выпрабаваны уласны метады аўтаматычнага вызначэння танальнасці тэкставых паведамленняў на розных мовах.

Вобласць прымянення вынікаў: у кіраванні, паліталогіі, сацыяльных навук, эканоміцы, у прагназаванні паказчыкаў продажаў, у вывучэнні гандлёвых стратэгий, у даследаваннях сацыяльных медыя.

Дыпломная праца выканана аўтарам самастойна.

ABSTRACT

Diploma thesis, 57 pages, 9 pictures, 3 tables, 47 sources.

Keywords: NATURAL LANGUAGE PROCESSING, TEXT CLASSIFICATION, SENTIMENT ANALYSIS, TONALITY DICTIONARIES, MACHINE LEARNING, TEXT DATA CONVERSION, REPRESENTATION OF TEXT DATA, QUALITY EVALUATION OF THE ALGORITHM.

The object of the research: methods and software tools for determining the tonality of texts.

The purpose of the work: to study the tasks of classification and determination of the tonality of texts, to consider and describe the main stages of classification, to study the main methods of classification of tonality, to consider existing software solutions for analyzing the tonality of texts. To develop a method for determining the tonality of texts in different languages and implement it.

Methods of the work: theoretical analysis of the literature on the research topic; detailed description and classification of methods used in the field under research; technical analysis of existing solutions related to the research topic and their software documentation; conducting a computational experiment; description and comparison of the results obtained.

Results of the work: the problems of classification and determination of the tonality of texts, the main stages of classification, approaches to classification and methods of classification of tonality are studied. The existing software solutions for determining the tonality of texts are analyzed. Based on the acquired knowledge, own method for automatically determining the tonality of text messages in different languages has been developed, implemented and tested.

Scope of the results: in management, political science, social sciences, economics, in forecasting sales figures, in the study of trading strategies, in social media research.

The thesis is performed by the author independently.