

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.42:81'371

Куличок  
Никита Сергеевич

Методы и программное средство анализа эмотивности текстов

**АВТОРЕФЕРАТ**

диссертации на соискание степени магистра

по специальности 1-40 80 05 – Программная инженерия

Научный руководитель  
Парамонов Антон Иванович  
кандидат технических наук, доцент

Минск 2023

## КРАТКОЕ ВВЕДЕНИЕ

Ежедневно пользователями соцсетей, новостными порталами, различными блогами генерируется множество данных. Постоянный рост объемов текстовой информации в цифровом виде приводит к возникновению потребности в новых инструментах для ее обработки и анализа. На сегодняшний день достаточно эффективно решаются задачи классификации текстов и автоматического синтаксического анализа. Известны подходы по семантическому и контент-анализу текстов. Однако на современном этапе развития автоматической обработки текстов задача эмотивной оценки или идентификации эмотивности текстов не решена в полной мере. Анализ эмотивности является достаточно сложной задачей, что связано с трудностями при выделении нужной эмоциональной лексики в текстах, а также с определением самого эмотивного пространства, количества и состава его измерений. В этой связи разработка методов и средств анализа эмотивности текстов остается актуальной и востребованной задачей.

Анализ эмотивности требуется для мониторинговых, аналитических и сигнальных систем, для систем документооборота и рекламных платформ, таргетированных по тематике веб-страниц и многих других. В настоящее время популярным становится создание разных чат-ботов с поддержкой искусственного интеллекта, способных в той или иной мере интерпретировать текст на естественном языке и реагировать в ответ на него. Таким образом сфера применения анализа эмотивности текстов ежедневно расширяется и приобретает новые аспекты.

В рамках данной работы рассматриваются вопросы анализа текстовой информации на естественном языке в контексте выявления их эмоциональной окраски. Проводится исследование моделей и методов обработки текстовой цифровой информации, а также современные технологии и подходы в разработке программных средств для решения прикладных задач анализа эмотивности текстов.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

**Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики.**

Тема диссертации соответствует п.1 приоритетных направлений научной, научно-технической и инновационной деятельности Республики Беларусь на 2021–2025 гг., утвержденных Указом Президента Республики Беларусь № 156 от 7 мая 2020 г. «Цифровые информационно-коммуникационные и междисциплинарные технологии, основанные на них производства». Диссертация выполнялась на кафедре программного обеспечения информационных технологий УО «Белорусский государственный университет информатики и радиоэлектроники».

### **Цель и задачи исследования.**

Целью диссертационного исследования является разработка методов и программного средства анализа эмотивности текстов. Объектом исследования выступают тексты на естественном языке небольшого размера, свойственные коротким сообщениям и небольшим записям в интернете. Предметом исследования являются модели и методы контент-анализа текстов на естественном языке.

Задачи, поставленные в рамках исследования:

- 1) Изучить существующие модели и методы анализа текстов;
- 2) Разработать модели эмотивности текста и его автора, адаптировать известные методы для работы с предложенными моделями;
- 3) Провести сравнительный анализ эффективности разных методов анализа эмотивности для текстов в рамках выбранной предметной области;
- 4) Разработать программное средство анализа эмотивности текстов для эксплуатации в рамках конкретной предметной области;
- 5) Провести компьютерный эксперимент.

Основной гипотезой, положенной в основу диссертационной работы, является возможность применения известных методов контент-анализа текстов для выявления их эмотивности в контексте определенной прикладной предметной области.

### **Новизна полученных результатов.**

Предложенные модели и программное средство анализа эмотивности текстов позволят выполнять анализ эмотивности текстов и авторов. Область применения полученной модели и программного средства обширна: для мониторинговых, аналитических и сигнальных систем, для систем документооборота и рекламных платформ, таргетированных по тематике веб-страниц и многих других. Достоверность научных выводов будет подтверждена практической реализацией моделей и методов, а также результатами проведенных компьютерных экспериментов.

### **Положения, выносимые на защиту.**

1. Модели эмотивности текста и его автора, содержащие базовые характеристики текста и известные закономерности построения текстовых форм. Адаптированные для работы с предложенными моделями методы контент-анализа текстов на естественном языке;

2. Программное средство анализа эмотивности текстов, основанное на предложенных моделях и методах.

### **Апробация результатов диссертации.**

Материалы, положенные в основу работы, докладывались и обсуждались на 57-ой научной конференции аспирантов, магистрантов и студентов БГУИР (г. Минск, Республика Беларусь, 2021) и IV международной научно-практической конференции «Непрерывное профессиональное образование лиц с особыми возможностями» (г. Минск, Республика Беларусь, 2021).

### **Опубликованность результатов диссертации.**

По теме диссертации опубликовано 3 печатных работы в научных изданиях, из них 2 статьи в сборнике научных статей VII Международной научно-практической конференции «BIG DATA and Advanced Analytics» (Минск, 2021) [1] и сборнике статей IV Международной научно-практической конференции «Непрерывное профессиональное образование лиц с особыми потребностями» (Минск, 2021) [2], и 1 тезис в сборник докладов 57-й научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, 2021) [3].

### **Личный вклад соискателя**

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя А. И. Парамонова заключается в определении целей и задач исследования, рекомендациях по содержанию диссертации и автореферата.

### **Структура и объём диссертации**

Работа состоит из общей характеристики работы, введения, четырёх глав, заключения, списка использованных источников, списка публикаций автора и приложения.

Общий объём работы составляет 73 страниц, из которых основного текста – 57 страниц, 16 рисунков на 12 страницах, список использованных источников на 4 страницах и 1 приложения на 11 страницах.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении определена область исследования, указаны современные подходы и проблемы анализа эмотивности, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов.

В первой главе представлен анализ предметной области и основные теоретические сведения, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Описаны аналоги программных средств анализа эмотивности и области его применения.

Вторая глава описывает методы, применяемые для анализа текстов и их эмотивности, описан выбранный алгоритм анализа эмотивности с помощью наивного байесовского классификатора и созданные модели эмотивности текста, автора и контекста.

Третья глава посвящена проектированию программного средства, анализу требований к нему. Также в главе описаны основные модули программного средства.

В четвёртой главе описано тестирование программного средства, примеры его работы и результаты. Также дана инструкция по эксплуатации программного средства.

В результате было реализовано программное средство анализа эмотивности с моделями эмотивности текста, автора и контекста и проведено его тестирование.

## ЗАКЛЮЧЕНИЕ

В ходе изучения литературы по теме магистерской диссертации и по причине специфики разрабатываемого проекта, был проведён ряд исследований на основе предыдущих разработок по тематике анализа эмотивности.

Была проанализирована классификация методов анализа эмотивности, а также исследован вопрос их эффективности. Кроме того были изучены аналоги программных средств, использующих в процессе своей работы анализ эмотивности.

В ходе моделирования предметной области были установлены требования к программному средству, сформулирована функциональная спецификация, спроектирована архитектура базы данных и программного средства анализа эмотивности.

Были созданы модели эмотивности для текста, контекста и автора текста.

На основе выбранного алгоритма анализа эмотивности и полученных моделей было разработано программное средство на языке Python. В программном средстве был реализован алгоритм наивного байесовского классификатора, а также разработанные модели эмотивности. Алгоритм наивного байесовского классификатора показал точность и эффективность в рамках решаемых задач.

Результаты проделанной работы показывают эффективность и возможность применения методов анализа эмотивности для анализа текстовых сообщений с реакциями на новости проекта онлайн-игры.

Разработанное программное средство может быть использовано в исследовательских целях для анализа эмотивности текстов в контексте и эмотивности авторов данных текстов.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Куличок, Н. С. Анализ эмотивности сообщений как инструмент адаптивного образования / Куличок Н. С., Парамонов А. И. // Непрерывное профессиональное образование лиц с особыми потребностями : сборник статей IV Международной научно-практической конференции, Минск, 9–10 декабря 2021 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: А. А. Охрименко [и др.]. – Минск, 2021. – С. 126–129.

2. Куличок, Н. С. Анализ эмотивности цифровых текстов / Н. С. Куличок, А. И. Парамонов // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник научных статей VII Международной научно-практической конференции, Минск, 19-20 мая 2021 года / редкол.: В. А. Богуш [и др.]. – Минск : Бестпринт, 2021. – С. 358–360.

3. Куличок, Н. С. О проблеме определения эмотивной оценки текста / Н. С. Куличок // Компьютерные системы и сети: 57-я научная конференция аспирантов, магистрантов и студентов, Минск, 19-23 апреля 2021 г. : сборник тезисов докладов / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2021. – С. 53.