

УДК 621.396.6

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ КЛАССИФИКАЦИИ И ФОРМ  
ПРЕДСТАВЛЕНИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧИ АВТОМАТИЧЕСКОГО АНАЛИЗА  
ТОНАЛЬНОСТИ КОРОТКИХ ТЕКСТОВ**

**А.А.Нугманова**

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, г. Санкт-Петербург)

**Научный руководитель – А.П.Затворницкий**

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, г. Санкт-Петербург)

Работа проводится в рамках темы НИР № 616029 «Разработка и исследование методов и алгоритмов распознавания эмоционального и психофизического состояния человека по многомодальным данным».

Анализ тональности – область компьютерной лингвистики, которая занимается изучением мнений и эмоций в текстовых документах. Основной задачей в анализе тональности является классификация полярности документа или предложения, то есть определение, является ли выраженное мнение в тексте позитивным, негативным или нейтральным. В наши дни компьютеры могут выполнять автоматический анализ цифровых текстов, используя элементы машинного обучения.

Алгоритмы, решающие задачу классификации, работают с математической моделью представления данных (в данном случае – текстового документа). Наиболее распространенная модель для данной задачи – это представление текста в виде некоторого признакового описания. Определение признаков и сопоставление им весов является нетривиальной задачей и во многом влияет на результат классификации. По этой причине было выделено 2 основных этапа работы:

1. Исследование различных форм представлений данных
2. Исследование различных алгоритмов классификации

Проведенные исследования показали, что результаты автоматической обработки текста во многом зависят от выбранного пространства признаков, а также от методов классификации. Сравнение работы алгоритмов проводилось на основании вычисленных характеристик, таких как точность классификации (Accuracy) и F-мера.