

Тема выпускной квалификационной работы: «Моделирование интеллектуальной системы классификации коротких текстов (на материале русскоязычных текстов)»

Автор ВКР: Дряева Анастасия Александровна

Научный руководитель ВКР: канд. экон. наук, доц. кафедры информационно-коммуникационных технологий, математики и информационной безопасности Тимченко Ольга Викторовна

Сведения об организации-заказчике: НОИК ИКИМТ ПГУ

Актуальность темы исследования: автоматизированная классификация документов или просто текстов на сегодняшний день позволяет сэкономить финансовые ресурсы и отказаться от человеческого участия в данном процессе. Однако стандартные подходы не всегда могут дать хороший результат, когда речь пойдет об операции над короткими текстами. Прежде всего, это обусловлено малым количеством характеристик, которые можно «добыть» из таких текстов. Целесообразность такой классификации обоснована прикладной задачей идентификации образа в коротких текстах, где в роли последних выступают объявления, размещенные на различных веб-ресурсах.

Цель работы: построение и исследование аналитической модели для классификации коротких текстов.

Задачи:

1. анализ проблемы автоматической классификации текстов;
2. сравнительный анализ и обзор существующих методов машинного обучения;
3. анализ и обзор подходов для решения задачи бинарной классификации, а также выбор средств реализации моделей;
4. разработка аналитической модели классификации коротких текстов;
5. разработка программного продукта классификации коротких текстов;

б. выбор оценки и оценка эффективности разработанной интеллектуальной системы.

Теоретическая и практическая значимость исследования: в ходе выполнения ВКР была предложена и апробирована модифицированная модель классификации коротких текстов, которая улучшила точность классификации в сравнении с классическими методами машинного обучения.

Результаты исследования: интеллектуальная система классификации коротких текстов на примере русского языка, обязательным элементом которой является программный продукт.

Рекомендации: полученные результаты и разработанный программный продукт рекомендуются ко внедрению в компании при формировании ее маркетинговой стратегии в интернет – пространстве, а также для повышения эффективности функционирования компании.