

УДК 004.056.53

**РАЗРАБОТКА МЕТОДА ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ ОТПРАВИТЕЛЯ
ЭЛЕКТРОННОГО СООБЩЕНИЯ НА ВЬЕТНАМСКОМ ЯЗЫКЕ НА ОСНОВЕ
МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ И АНАЛИЗА ДАННЫХ**

Магистр: Буй Нам Хань

**Научный руководитель: Воробьева Алиса Андреевна – кандидат технических наук,
доцент ФБИТ, университет ИТМО**

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики. Российская Федерация, 197101, г. Санкт-Петербург, Кронверкский пр., д.49, E-mail: org@mail.ifmo.ru

Аннотация: В процессе работы проводилась разработка метода идентификации пользователя отправителя электронного сообщения на вьетнамском языке на основе методов машинного обучения и анализа данных. Используемые алгоритмы включают Наивный Байесовский, SVM, случайный лес и логистическая регрессия.

Введение: С быстрым развитием Интернета и сопутствующих технологий, общение между людьми стало проще, чем когда-либо. Электронная почта, новостные сайты, приложения для социальных сетей становятся незаменимым инструментом связи. К сожалению, из-за того, что общая функция является анонимной, пользователи могут использовать онлайн-сообщения в незаконных целях или причинять вред другим. Следовательно, необходимо иметь способ идентификации пользователя онлайн-сообщения. Было проведено много исследований с разными корпусами и разными языками.

Цель работы заключается в исследовании подход к выявлению авторов онлайн-сообщений на вьетнамском языке на основе алгоритмов машинного обучения. Промежуточные результаты:

- Выявление характеристик онлайн-сообщений;
- Выявление методов сбора данных
- Определение подходов и расчетов признаков сообщения.

Основной результат: Применение алгоритмов машинного обучения для классификации пользователей. Проанализировать результаты и сделать выводы.

Литературы:

1. Воробьева А.А. Методика идентификации интернет-пользователя на основе стилистических и лингвистических характеристик коротких электронных сообщений – Информация и космос. 2017. № 1. С. 127-130.
2. Воробьева А.А., Гвоздев А.В. Идентификация анонимных пользователей Интернет-порталов на основании технических и лингвистических характеристик пользователя. Научно-технический вестник информационных технологий, механики и оптики. 2014. № 1(89). С. 139-144.
3. Стивен Берд, Эван Кляйна, Эдвард Лопер. «Обработка естественного языка с Python». О'Рейли Медиа, 2009, 502 стр.

Автор _____ Буй Н.Х.

Научный руководитель _____ Воробьева А.А.