

УДК 004.89

ПРИМЕНЕНИЕ ИНСТРУМЕНТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ФУНКЦИОНИРОВАНИИ ПОРТАЛА «НАШ САНКТ-ПЕТЕРБУРГ»

Беген П.Н.

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к.полит.н, Чугунов А.В.

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В работе представлены основные результаты разработки прикладного сервиса для системы портала «Наш Санкт-Петербург» с применением инструментов искусственного интеллекта. Сервис позволяет оптимизировать процесс подачи сообщения на портал и его последующую обработку.

Применение информационно-коммуникационных технологий и средств в сферах цифрового государственного управления, электронного правительства, электронной демократии, электронного участия граждан в настоящее время направлено на повышение эффективности функционирования существующих государственных информационных систем. Согласно программе «Цифровая экономика Российской Федерации» и «Стратегии развития информационного общества в Российской Федерации на 2017–2030 гг.» применение новейших цифровых технологий в органах государственной власти способствует повышению качества госуправления, развитию социальной сферы, повышению эффективности системы государственного управления, улучшения взаимодействия граждан и государства и др. В качестве одного из видов цифровых технологий на данный момент активно используется применение искусственного интеллекта и его инструментов (машинное обучение, прогнозирование, обработка естественного языка, речевые технологии и т. д.).

В работе в качестве системы, реализующей принципы электронного участия граждан, был рассмотрен портал «Наш Санкт-Петербург», позволяющий оперативно взаимодействовать гражданам с исполнительными органами государственной власти по решению насущных городских проблем.

Целью работы являлось повышение эффективности работы с сообщениями на портале «Наш Санкт-Петербург» посредством внедрения современных цифровых технологий.

В ходе анализа основных бизнес-процессов системы портала обнаружены следующие пункты, влияющие на эффективность работы портала с сообщениями граждан:

- сложность и постоянные ошибки при самостоятельном выборе гражданином одной из 200 доступных категорий проблем на портале;
- отклонение 20–25% поступающих сообщений по причине неверно выбранной категории проблемы или несоответствия текстового описания проблемы одной из категорий;
- комплексная схема бизнес-процесса отработки сообщений, не рассчитанная на дни и часы повышенной нагрузки на портал, приводящая к увеличению сроков отработки и несоблюдению регламента.

Совместно с ГК «Нетрика», являющейся разработчиком и участником рабочей группы по созданию портала «Наш Санкт-Петербург», предложен подход по оптимизации процесса подачи сообщения на портал и его последующей обработки за счет использования и внедрения инструментов искусственного интеллекта. В качестве инструментов использованы методы машинного обучения для решения задачи текстовой классификации и методы обработки естественного языка.

Основой предложенного подхода стала разработка алгоритма автоматической классификации сообщений граждан на основе текстового описания проблемы. Решение задачи основывалось на принципах обучения с учителем (Supervised learning), в частности решалась задача

классификации, т. е. сопоставление множества объектов некоторым типам классов (категорий).

Для формирования обучающей выборки была использована копия базы данных портала, содержащая 1.5 млн сообщений с указанными категориями проблем. Для проверки релевантности и качества последующего обучения сформирована тестовая выборка из 300 тыс. сообщений (20% от общего числа), не участвующая в обучении моделей алгоритма.

Для реализации методов по обработке текстовой информации и преобразованию ее в числовой вид в алгоритме использованы методы обработки естественного языка (токенизация, лемматизация, стемминг, очистка данных, TF-IDF и т. д.).

В качестве реализации задачи классификации и обучения моделей использованы основные методы машинного обучения: наивный байесовский классификатор, деревья решений, метод опорных векторов (SVM) и искусственные нейронные сети с разным типом архитектур (сети прямого распространения, сверточные, рекуррентные). В результате обучения и соответствующей валидации данных наилучшими показателями обладали сверточные нейронные сети (около 83% точности определения по 200 категориям, F-мера).

Реализованный алгоритм автоматической классификации использован при разработке интеллектуального классификатора, представляющего собой веб-модуль и реализующего API-механизмы для автоматической классификации сообщений граждан по категориям проблем на портале. Модуль позволяет упростить процесс подачи сообщения на портал для пользователя, а для службы модерации предоставить автоматическую проверку сообщения на соответствие категории.

Вся техническая разработка велась на языке программирования Python версии 3.6. Библиотеки scikit-learn, Keras, TensorFlow, pandas, pymorphy2, nltk были использованы для реализации методов машинного обучения и методов обработки естественного языка. Для разработки веб-модуля и API был использован фреймворк Django.

Разработанные прикладные решения позволяют оптимизировать процесс работы с сообщением на портале за счет снижения процента отклонения сообщений по причине неверно выбранной категории или несоответствия ей с 20–25% до 5-10%, сокращения времени на проверку модератором сообщения примерно на 30% и, как следствие, снижения риска увеличения сроков отработки сообщения в пиковые дни нагрузки, а также повышения удобства при подаче сообщения на портал для гражданина.

Перспективы исследования заключены в проведении тестирования работы прикладных решений на основе реальных данных с проведением статистического анализа, а также последующего внедрения решений в существующую структуру системы портала.