

УДК 519.688

## **ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ГЕНЕРАЦИИ ОПИСАНИЯ ТОВАРОВ ИНТЕРНЕТ-МАГАЗИНА**

© Дудникова Н.А.

*Самарский национальный исследовательский университет  
имени академика С.П. Королева, г. Самара, Российская Федерация*

e-mail: na.dudnikova25@gmail.com

Научный руководитель: Семенова И.В., канд. физ.-мат. наук, доцент  
*Самарский национальный исследовательский университет  
имени академика С.П. Королева, г. Самара, Российская Федерация*

Ключевые слова: обработка естественного языка, генерация текста, обучение, дообучение.

Ежегодно число онлайн-покупателей возрастает. Интернет-магазины должны предоставлять пользователям актуальную и достоверную информацию о товарах, в том числе – подробное описание товара. Частое изменение ассортимента в магазинах приводит к необходимости постоянного добавления больших объемов уникальной информации о новых товарах. Автоматизировать процесс формирования описаний товаров можно при помощи методов генеративного искусственного интеллекта, который все больше применяется компаниями для решения различных задач.

В рамках работы разработана система по работе с товарами интернет-магазина и подсистема для генерации описаний товаров интернет-магазина с помощью методов машинного обучения.

Для подсистемы генерации описаний с точки зрения обработки естественного языка возникает задача генерации текста. Для неё могут быть использованы различные модели машинного обучения, среди которых рекуррентные нейронные сети и их модификации, трансформеры, генеративные модели. Кроме этого, для решения данной задачи можно использовать предобученную модель [1–3].

Для поиска наиболее эффективного решения задачи генерации описания товаров в работе было реализовано ее решение двумя методами:

1) с использованием модели CopyNet, которая обучалась на подготовленном наборе данных «с нуля». Особенностью модели CopyNet является механизм копирования, позволяющий использовать для формирования выходных данных токены из входных данных в случае их отсутствия в словаре. Энкодер и декодер представлены одной LSTM-моделью, что способствует повышению обучающей способности модели за счёт меньшего числа параметров.

2) путем дообучения для решения рассматриваемой задачи предобученной языковой модели ruGPT-3 Small, основным преимуществом которой является то, что она ориентирована на русский язык. Метод дообучения позволяет сохранить с помощью заморозки нижних слоёв общих языковых паттернов, и, следовательно, сократить время и вычислительные ресурсы, требуемые для получения работоспособной модели. Модель ruGPT-3 Small основана на архитектуре Трансформер, которая зарекомендовала себя как архитектура, позволяющая получить качественные результаты для задач обработки естественного языка.

Для обучения моделей необходим набор данных (датасет), соответствующий решаемой задаче. Для его формирования был разработан скрипт, с помощью которого с различных интернет-источников была собрана информация о товарах интернет-

магазина. Датасет включает в себя название каждого товара, его категорию, список характеристик и описание. Объём датасета – 2000 записей: по 400 записей для каждой из 5 категорий (смартфоны, телевизоры, мониторы, ноутбуки и компьютерные мыши). Важным моментом является подготовка собранного датасета для обучения моделей, включающая очистку данных от пустых строк, от лишних пробелов и так далее.

Для реализации системы по работе с товарами интернет-магазина была выбрана трёхуровневая клиент-серверная архитектура. При этом в качестве СУБД была выбрана PostgreSQL, для реализации серверной части использован язык программирования Java с фреймворком Spring Boot, а для клиентской части – Java Script, HTML, фреймворк Bootstrap [4].

Подсистема генерации описаний товаров представляет собой сервис, работающий по принципам REST, что позволяет обеспечить универсальность подсистемы и её интегрируемость с существующими системами, поддерживающими данные принципы. Для реализации был использован язык Python, среди основных библиотек присутствуют Pandas, PyTorch.

Таким образом, была спроектирована и реализована система по работе с товарами интернет-магазина, а также, в рамках данной системы, разработана подсистема генерации описаний товаров интернет-магазина.

### **Библиографический список**

1. Уоллис, К. Spring в действии / пер. с англ. А.Н. Киселева. – 6-е изд. – М.: ДМКПресс, 2022. – 544 с.: ил/
2. Новиков, Б.А. Основы технологий баз данных: учебное пособие / Б.А. Новиков. – 2-е изд. – М.: ДМК Пресс, 2020. – 582 с.
3. Лейн, Х. Обработка естественного языка в действии / Х. Лейн, Х. Хапке, К. Ховард. – СПб.: Питер, 2020. – 576 с.: ил. – (Серия «Для профессионалов»).
4. Краткий экскурс в ruGPT-3. Инструкция и демонстрация / Хабр. – URL: <https://habr.com/ru/articles/589663> (дата обращения: 20.04.2024).