

РАЗРАБОТКА МЕТОДА ДООБУЧЕНИЯ ВИЗУАЛЬНО-ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ АВТОМАТИЗИРОВАННОГО ФУНКЦИОНАЛЬНОГО ТЕСТИРОВАНИЯ ВЕБ-ПРИЛОЖЕНИЙ

Хабнер Г. Е.

Научный руководитель – канд. техн. наук, доцент Болдырева Е. А.

Университет ИТМО

georgiy.khabner@mail.ru

Введение

Традиционное автоматизированное функциональное тестирование основано на использовании селекторов, при котором тестовые сценарии описываются через указание явных идентификаторов - CSS-селекторов, XPath-выражений, атрибутов. Такой подход оказывается хрупким при изменении верстки и рефакторинге, поскольку появляется необходимость в переписывании тестов. В связи с этим было рассмотрено решение по использованию визуально-языковых моделей (VLM), которые в последние годы продемонстрировали высокую точность в определении элементов веб-интерфейсов по изображению страницы [1]. Однако существующие модели визуального обнаружения, такие как Moondream и Florence-2, не всегда предоставляют желаемую точность при работе с интерфейсами пользовательских приложений, а передовые модели требуют значительных вычислительных ресурсов. В качестве решения в данной работе рассматривается дообучение моделей на примерах пользовательских интерфейсов [4], а также решается задача составления такого датасета.

Основная часть

В данной работе рассматриваются два способа формирования обучающих данных. Первый включает сбор реальных примеров сайтов, из которых собираются: скриншот веб-страницы – текстовая инструкция – координаты элемента. Координаты элементов извлекаются из DOM-структуры страницы. Для расширения датасета используется аугментация: переключение цветовых тем и различные разрешения области просмотра. Второй способ основывается на генерации синтетических веб-страниц для обучения специфическим вариантам дизайна интерфейсов. Генерация осуществляется за счет комбинирования различных компонентов, CSS-стилей и состояний страницы при помощи автоматизированного скрипта. Собранные данные используются для дообучения готовой визуально-языковой модели с применением Low-Rank Adaptation (LoRA) [2], после чего проводится сравнение с предсказаниями базовой модели.

Для оценки влияния дообучения на качество решения практической задачи – автоматизированного QA-тестирования – модель интегрируется в открытый фреймворк автоматизированного тестирования Magnitude [3]. Модель является основой субагента, отвечающего за описание контента и навигацию по странице на основе инструкций на естественном языке. Проведено сравнение качества работы Magnitude с использованием базовой и дообученной моделей.

Выводы

Предложенный метод направлен на ускорение процесса написания сценариев для проведения функционального тестирования. Также за счет исключения использования селекторов, порог вхождения к написанию сценариев значительно снижается. Интеграция в фреймворк Magnitude демонстрирует практическую применимость полученного подхода по сравнению с традиционным автоматизированным функциональным тестированием.

Литература

1. Moondream: tiny vision language model [Электронный ресурс]. – Режим доступа: <https://moondream.ai> (дата обращения 17.02.2026).
2. Hu E. J., Shen Y., Wallis P. et al. LoRA: Low-Rank Adaptation of Large Language Models // Proc. of the 10th International Conference on Learning Representations (ICLR). 2022.
3. Magnitude: open-source AI-native testing framework [Электронный ресурс]. – Режим доступа: <https://magnitude.run> (дата обращения 17.02.2026).
4. Hugging Face: документация по дообучению визуально-языковых моделей [Электронный ресурс]. – Режим доступа: <https://huggingface.co/docs> (дата обращения 17.02.2026).