

УДК 004.912

РАЗРАБОТКА БИБЛИОТЕКИ PYTHON ДЛЯ РЕДАКТИРОВАНИЯ ОБЪЕКТОВ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Шафиков М.А. (ИТМО), Крылов М.М. (ИТМО), Марцинкевич В.И. (ИТМО),
Терещенко В.В. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Федоров Д.А.
(ИТМО)

Введение. В рамках проекта, нацеленного на проверку соответствия электронных документов определенным стандартам, таким как ГОСТ, была выявлена проблема: существующие библиотеки, предоставляющие функции создания комментариев в файлах формата DOCX и ODT не подходят для нашего проекта, так как являются закрытыми и разработаны иностранными компаниями. Эта ситуация создает необходимость в разработке собственной библиотеки на языке программирования Python, способной взаимодействовать с объектами электронных документов указанных форматов, при этом являющейся Open-Source решением.

Основная часть. Разработка библиотеки Python для редактирования объектов электронных документов сосредоточена в первую очередь на создании функционала для поиска и комментирования конкретных объектов, по заданным параметрам в файлах формата DOCX и ODT, но она также будет предоставлять методы для изменения их свойств. Учитывая, что существующий инструмент Aspose, является закрытым, лицензируемым и иностранным, мы ставим перед собой задачу разработать Open-Source библиотеку. Это обеспечит не только более широкий доступ к инструментам редактирования электронных документов, но и позволит легко адаптировать решение под конкретные требования проекта [1].

Основная идея библиотеки заключается в том, что xml-файлы, из которых состоит документ с расширением DOCX или ODT, представляются в виде набора классов и их свойств, при изменении которых происходит изменение содержимого xml-файла оригинального документа. Таким образом мы можем, например, изменить размер шрифта объекта или добавить комментарий [2].

Выводы. Разработка библиотеки на Python представляет собой значимый шаг в автоматизации процесса проверки файлов на наличие несоответствий со стандартами. Она позволит быстро и удобно редактировать объекты электронных документов, позволяя избежать использования закрытых и иностранных библиотек. Также за счет открытого исходного кода библиотеки каждый желающий сможет доработать библиотеку в соответствии со своими нуждами.

Список использованных источников:

1. Марцинкевич В. И. и др. АНАЛИЗ ВОЗМОЖНОСТЕЙ ПАРСИНГА ЭЛЕКТРОННЫХ ТЕКСТОВЫХ ДОКУМЕНТОВ ДЛЯ АВТОМАТИЗАЦИИ НОРМОКОНТРОЛЯ // Экономика. Право. Инновации. – 2022. – №. 3. – С. 39-49.
2. Pajankar A. Python Unit Test Automation: Practical Techniques for Python Developers and Testers. – Apress, 2017.