

УДК 004.415 : 004.6

ОПТИМИЗАЦИЯ ПАРСИНГА СЕРВИСА АВТОМАТИЗИРОВАННОГО НОРМОКОНТРОЛЯ ДОКУМЕНТОВ

Ларионова Г. С. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Лушников Д. М. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Терещенко В. В. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Научный руководитель - преподаватель, Насыров Н. Ф. (Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»)

Аннотация: В данном докладе проводится анализ существующего парсера системы автоматизированного нормоконтроля документов, а также описаны планируемые изменения и их влияние на работу системы.

Введение.

В процессе написания научных и выпускных квалификационных работ студенты, а иногда и научные сотрудники, допускают ошибки при оформлении. В результате текст работы может не соответствовать ГОСТам или другим стандартам. Вместе с тем процесс проверки работ научными руководителями и иными уполномоченными лицами проводится преимущественно вручную, что требует больших человеческих ресурсов. Решением этих проблем может стать разрабатываемая на базе Университета ИТМО система автоматизированного нормоконтроля документов, одной из основных частей которого является корректор документов формата DOCX. Целью данной работы является выявление недостатков существующей системы парсинга документов и формирование списка требуемых изменений.

Основная часть.

Предлагается модернизировать либо весь парсер, либо отдельные его части с использованием высокоуровневого языка программирования Python. Это связано с тем, что существенная часть необходимых функций для работы с электронными документами различных форматов на данном языке уже реализована. Также сравнительный анализ возможностей языков программирования Java и Python показал, что при работе с электронными документами Python уступает только при извлечении данных из файлов формата PDF. Поскольку существующий корректор реализован в виде консольного приложения при помощи языка C#, его модернизация на текущем этапе разработки сервиса позволит обеспечить упрощенную поддержку и внесению изменений в будущем.

Выводы.

Внедрение описанных изменений позволит улучшить масштабируемость разрабатываемой системы автоматизированного нормоконтроля документов, а также сделает возможным контейнеризацию отдельных модулей приложения. Данные модификации, в свою очередь, позволят снять нагрузку с серверной части системы, что положительно скажется на временных показателях и развертывании системы в виде веб-сервиса.

Ларионова Г. С. (автор)

Подпись

Лушников Д. М. (автор)

Подпись

Терещенко В. В. (автор)

Подпись

Насыров Н. Ф. (научный руководитель)

Подпись