

УДК 004.057.3

## МОДЕРНИЗАЦИЯ СЕРВИСА АВТОМАТИЗИРОВАННОГО НОРМОКОНТРОЛЯ ДОКУМЕНТОВ

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

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

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

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

Аннотация: В данном докладе проведен анализ существующей системы автоматизированного нормоконтроля документов, проведено исследование различных форматов документов и смоделирован универсальный класс для хранения и обработки необходимых системе параметров текста.

### Введение.

Согласно данным статистики примерно четверть ошибок при создании отчетности о научно-исследовательской работе приходится на несоблюдение правил в оформлении. В итоге выявление этих ошибок возлагается на научных руководителей, которые помимо анализа качества изложенного исследования, вынуждены контролировать еще и выполнение требований стандартов. Для того, чтобы ускорить процесс проверки отчетности и избежать возможные ошибки человеческого фактора, на базе Университета ИТМО ведется разработка сервиса автоматизированного нормоконтроля документов, основной задачей которого становится поиск ошибок в оформлении отчета (в частности, связанных с отступами, кеглем, шрифтами, выравниванием текста и пр.). Однако, ключевой особенностью данной системы является ориентированность на работу с файлами исключительно формата DOCX. В данной работе будет определена необходимость отойти от подобного подхода и предложено возможное решение по модернизации системы.

### Основная часть.

Согласно проведенному исследованию, можно смело заявить, что работа сервиса исключительно с форматом DOCX имеет существенные ограничения. Во-первых, они обоснованы тем, что при обращении к статистике видно, что в научном сообществе DOCX является не единственным популярным форматом и даже не занимает первое место в рейтинге по частоте использования. Во-вторых, политика РФ в последние годы активно направлена на импортозамещение и использование открытых, бесплатных разработок, что идет вразрез с текущими возможностями сервиса, поддерживающего исключительно расширение, созданное под работу с документами Word. Таким образом, нами была обозначена необходимость осуществить масштабируемость системы под работу с различными форматами. На текущий момент система напрямую работает с элементами текста DOCX документа, извлекая из них все необходимые параметры для обработки и храня их значения в CSV-файлах. Нами же предлагается создание универсального класса, который сможет хранить описание параметров элементов отчета. Для достижения данной цели был проведен предварительный сравнительный анализ форматов и определены их различия и сходства.

**Выводы.**

Модернизация системы под взаимодействие с элементами отчета через универсальный тип данных позволит обеспечить легкую масштабируемость сервиса под работу с разными форматами. Более того, это снизит расход трафика на передачу данных, поскольку взаимодействие сервиса с серверной частью будет совершаться через JSON-запросы, а не отправку CSV-файлов.