

УДК 004.428

АРХИТЕКТУРА СЕРВИСА АВТОМАТИЗИРОВАННОГО НОРМОКОНТРОЛЯ

Ларионова Г. С. (Университет ИТМО), Третьякова В. С. (Университет ИТМО),

Бережков А. В. (Университет ИТМО)

Научный руководитель – доцент, кандидат технических наук, Горлушкина Н.Н.

(Университет ИТМО)

Введение. Нормоконтроль отчетной документации помогает обеспечить точность и непротиворечивость информации, предоставляемой обучающимися, предотвратить недоразумения и ошибки, которые могут возникать при сдаче работ и обеспечить соответствие законодательным и нормативным требованиям.

Вопрос автоматизации нормоконтроля рассмотрен многими авторами, которые нашли частичное решение этой проблемы [1,2]. Основным преимуществом автоматизации процесса нормоконтроля является уменьшение времени на проверку документов [3] [4].

Основная часть. Для автоматизации нормоконтроля авторами были разработаны сервис и библиотека для проверки документов на соответствие правилам оформления. Проект включает в себя несколько частей: модули парсинга документов для форматов ODT, DOCX, PDF; набор классов для хранения свойств и содержимого структурных элементов документа в едином формате; классификатор ошибок элементов; RESTful API работы с модулями; модуль NLP (Natural Language Processing) для анализа текста; веб-интерфейс для загрузки документов и формирования отчетов; база данных для хранения правил проверки.

Набор классов – это структуры для каждого вида элементов документа, хранящая содержимое элементов (текст, таблица, рисунок и т.д.) и их свойства. Так как различные форматы документа хранят информацию о свойствах в разном формате, то для дальнейшего анализа и использования классификатора элементов требуется приведение свойств к стандартизированному виду.

Модули парсинга решает следующие задачи: извлечение изображений, текста, таблиц, формул и их свойств. После извлечения контента и свойств модули приводят их к требованиям соответствующего класса из набора классов.

Классификатор элементов - обученная модель (CatBoost), которая принимает на вход один из объектов документа и возвращает вероятность принадлежности к одному из классов структурных элементов.

Модуль NLP для анализа текста позволяет реализовать следующие проверки: определить есть ли введение и оглавление; определить корреляцию текста введению; предсказать какому факультету соответствует работа; предсказать наличие целей, задач.

База данных хранит эталонные свойства оформления структурных элементов отчета, что позволяет проводить проверку по различным стандартам оформления.

Для интеграции и работы всех сервисов разработан RESTful API на фреймворке Flask. Он является связующим звеном между модулями парсинга, классификатором, модулем NLP, базой данных и веб-интерфейсом.

В дальнейшем предполагается усовершенствование методов парсинга, а также добавление в архитектуру проекта нейронной сети для определения корректности оформления элементов, которые сложно проверить с использованием классификатор, к примеру формул.

Выводы. Чтобы автоматизировать процесс нормоконтроля, авторы разработали комплексный проект, который включает модули синтаксического анализа документов, набор классов для хранения свойств документа, классификатор элементов, модуль NLP для анализа текста, веб-интерфейс и базу данных для хранения правил проверки. Проект интегрирован с помощью RESTful API, построенного на платформе Flask framework, который связывает все сервисы вместе.

Список использованных источников:

1. Соколов, А. А. Разработка метода автоматизации процесса нормоконтроля технической документации / А. А. Соколов, А. М. Дворянкин, А. Ю. Ужва // Известия Волгоградского государственного технического университета. – 2013. – № 22(125). – С. 113-118. – EDN RPERYJ.

2. Назарцева, А. А. Совершенствование процесса нормоконтроля технической документации / А. А. Назарцева, А. В. Губарев // Качество продукции: контроль, управление, повышение, планирование : сборник научных трудов 5-й Международной молодежной научно-практической конференции, Курск, 14 ноября 2018 года. – Курск: Закрытое акционерное общество "Университетская книга", 2018. – С. 84-87

3. Трифанов, Е. А. Оптимизация нормоконтроля технической документации предприятия / Е. А. Трифанов, А. И. Соглаева, О. В. Аникеева // Качество в производственных и социально-экономических системах : сборник научных статей 10-й Международной научно-технической конференции, Курск, 15 апреля 2022 года / Юго-Западный государственный университет. – Курск: Юго-Западный государственный университет, 2022. – С. 393-396. – EDN VHOEDB.

4. Блинова, С. Д. Автоматизация нормоконтроля выпускных квалификационных работ / С. Д. Блинова, А. А. Пермякова // Общество, наука, инновации (НПК - 2014) : Всероссийская ежегодная научно-практическая конференция: сборник материалов: общеуниверситетская секция, БФ, ГФ, ФЭМ, ФАВТ, ФАМ, ФПМТ, ФСА, ХФ, ЭТФ, Киров, 15–26 апреля 2014 года / Вятский государственный университет. – Киров: Вятский государственный университет, 2014. – С. 938-942.

Бережков А. В (автор)

Третьякова В. С (автор)

Ларионова Г. С. (автор)

Горлушкина Н. Н. (научный руководитель)