

ВЕБ-ПРИЛОЖЕНИЕ ДЛЯ ОПТИМИЗАЦИИ РАЗМЕТКИ ДОКУМЕНТАЦИИ

WEB APPLICATION FOR OPTIMIZING DOCUMENTATION

MARKUP *К.Д. Малашко¹⁾, О.В. Дубровина²⁾*

K.D. Malashko¹⁾, O. Doubrovina²⁾

Белорусский государственный университет, Минск, Беларусь,

²⁾*doubrovina@bsu.by*

¹⁾ Belarusian State University,

Minsk, Belarus, ²⁾*doubrovina@bsu.by*

В статье описывается веб-приложение в трехуровневой архитектуре, предназначенное для автоматической оптимизации процесса разметки документации с возможностью разделения задач и их координации между разработчиками.

Ключевые слова: веб-ресурс; разметка документов; трехуровневая архитектура.

The article describes a web application in a three-tier architecture, designed to automatically optimize the process of marking up documentation with the ability to separate tasks and coordinate them between developers.

Keywords: web application; three-tier architecture: marking up documentation.

Современный документооборот любого предприятия подразумевает необходимость обработки огромного количества текстовой информации. Для повышения удобочитаемости, как правило, применяется форматирование текста. Однако более эффективным методом является разметка текста, что означает присвоение меток или тегов отдельным ключевым словам или фразам в тексте. Такой подход позволяет улучшить понимание содержания текста, а также, при необходимости, провести его классификацию.

Предлагаемое веб-приложение предназначено для автоматизации процесса разметки документов с возможностью последующей конвертации документов в формат .pdf. Такое приложение позволит улучшить координацию работы между командами разработчиков и предоставит возможность ролевого распределения задач с отслеживанием статистики производительности сотрудников.

Процесс обработки документов представлен с помощью диаграммы на рисунке 1 и состоит из трех этапов.

Первый этап – предварительная обработка документа (preprocessing). Документы, не содержащие разметки, загружаются в хранилище S3 и конвертируются в формат pdf с последующим разделением на страницы.

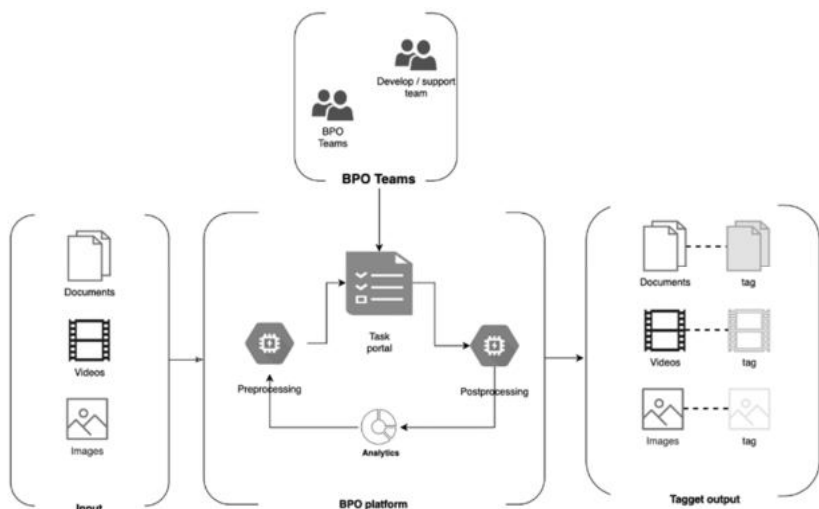


Рис. 1. Диаграмма работы приложения

Второй этап – разметка документа. Для выполнения разметки документа необходимо выполнить следующие шаги:

1. Назначение ролей. Руководителем проекта в системе Keycloak назначаются роли сотрудникам, например, роль agent, которая будет отвечать на начальную разметку, и роль validator, которая отвечает за проверку разметки, сделанной сотрудником с ролью agent.

2. Создание задачи на платформе пользователей Task Portal. На данном этапе руководитель создает задачу и соотносит ее с ролью, которая будет за это отвечать.

3. Разметка документов. После распределения ролей и задач, сотрудник с ролью agent может приступить к разметке назначенных на него документов.

4. Проверка размеченных документов. После первоначальной разметки текущий документ автоматически переназначается для проверки. На данном этапе сотрудникам с ролью validator отображается ранее выполненная разметка, в которую они могут внести изменения и добавить новые элементы. При завершении разметки документа изменяется его статус.

Третий этап – отправка размеченного документа (postprocessing). После валидации, текущий документ вместе с JSON-файлом, содержащий всю разметку, автоматически загружается в новую папку в S3.

При программной реализации использована трехуровневая архитектура клиент-серверного приложения, представленная на рисунке 2. Такая модель разделена на три слоя: клиентский уровень, слой бизнес-логики и слой доступа к данным и обеспечивает более эффективное взаимодействие между клиентом и сервером и является более гибкой в сравнении с другими.

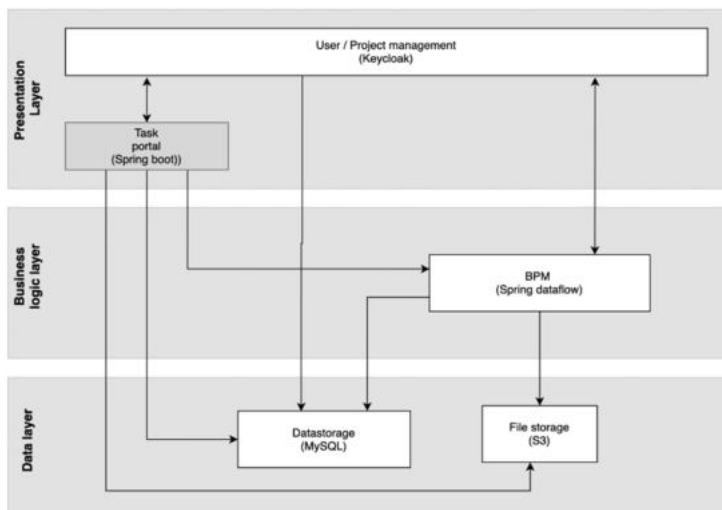


Рис. 2. Архитектура веб-приложения

На клиентском уровне осуществляется работа сервиса Keycloak, в котором руководитель проекта распределяет пользователей по ролям и группам, задает им данные для входа на Task Portal и ограничения доступов. Для администратора в Task Portal доступны все основные страницы, такие как просмотр всех задач, управление пользователями в случае переназначения задачи на другую роль, страница управления задачами, где отображается статистика всех документов, количество выполненных документов, страницы для создания шаблона задания, ручной задачи, страница для распределения задач по категориям. Для пользователя доступна только страница со всеми задачами, которые отображаются согласно его роли.

Уровень бизнес-логики представлен потоком Spring Data Flow и описывается запущенным потоком, зависящим от вида его разметки.

На уровне доступа данных имеется две составляющие: база данных и сервис S3. База данных представлена 9 таблицами, взаимосвязанными

друг с другом, записи в базу данных добавляются только после того, как руководитель проекта создал задачу на Task Portal. На данном уровне также осуществляются операции по загрузке, выгрузке и конвертированию документов в сервисе S3.

Для обеспечения изоляции всех отдельных компонентов (Keycloak, Task Portal, Spring DataFlow) и их зависимостей друг от друга и системных ресурсов было принято решение о разделении компонентов на контейнеры с помощью платформы Docker.

В процессе разработки веб-приложения были использованы современные технологии и инструменты, такие как сервис Amazon S3 в качестве хранилища документов, язык гипертекстовой разметки HTML, каскадные таблицы стилей CSS, язык программирования JavaScript, библиотеки Bootstrap, Material UI для клиентской части приложения и язык программирования Java, инструмент Spring Data Flow, сервис для аутентификации и авторизации пользователей Keycloak, система управления базами данных MySQL для реализации серверной части. Для упаковки веб-приложения с целью распределения сервиса Keycloak, платформы пользователей Task Portal и инструменты Spring Data Flow, предназначенный для создания независимых друг от друга потоков, описывающих процесс обработки, на отдельные контейнеры, в которых указываются все необходимые зависимости и настройки для запуска приложения, были созданы docker-компоненты.

Разработанный программный продукт может использоваться для разметки документации различного типа и формата, что делает его удобным инструментом для различных отраслей. Пользователю предоставляются следующие функциональные возможности: интуитивно понятный и удобный в использовании интерфейс пользователя; распределение ролей сотрудников в приложении и их задач согласно заранее установленным ролям; отслеживание статистики работы сотрудника; ограничение доступа для неавторизованных пользователей; одновременную работу нескольких команд; поддержку большого количества пользователей и большого объема данных; несложное сопровождение и обновление.

Данное веб-приложение предназначено для использования компанией «EPAM Systems» командой ВРО отдела, которая занимается разметкой текстовой и визуальной информации. В дальнейшем планируется расширить способности платформы, внедрив возможность разметки изображений, аудио- и видеофайлов, а также внедрить данную платформу в качестве основного инструмента для разметки информации компании.