

Приложение выполняет обработку загруженных PDF-файлов и изображений, извлекает из них текст, отправляет его в API GigaChat для получения краткого содержания и предоставляет возможность скачать результаты в формате DOCX. Давайте подробнее рассмотрим, как работает каждый компонент приложения:

## 1. Импорт необходимых модулей и установление параметров доступа:

Приложение начинается с импорта необходимых библиотек, включая `pytesseract` для OCR, `PyPDF2` для работы с PDF, `camelot` для извлечения таблиц из PDF, и другие. Также установлены переменные `CLIENT_ID` и `CLIENT_SECRET`, которые используются для получения токена доступа от GigaChat.

## 2. Функции для работы с файлами:

- **`extract_text_from_pdf(pdf_path)`**: Эта функция извлекает текст из PDF-файла с помощью библиотеки `PyPDF2`.
- **`extract_tables_from_pdf(pdf_path)`**: Эта функция использует библиотеку `camelot`, чтобы извлечь таблицы из PDF-файла.
- **`extract_text_from_image(image_path)`**: Эта функция извлекает текст из изображения с помощью `pytesseract`.
- **`save_text_to_docx(text, docx_filename)`**: Эта функция сохраняет текст в DOCX-файл с помощью библиотеки `python-docx`.

## 3. Функции для работы с GigaChat:

- **`get_gigachat_token()`**: Функция отправляет запрос на получение токена доступа к API GigaChat. Она использует предоставленные `CLIENT_ID` и `CLIENT_SECRET` для аутентификации.
- **`send_to_ollama(auth_token, user_message)`**: Эта функция отправляет текстовое сообщение в GigaChat API и получает ответ. Перед отправкой текст предваряется промптом (в данном случае запросом на краткое содержание).

## 4. Основная функция `upload_file(request)`:

- Эта функция обрабатывает HTTP POST запросы, в которых передается файл (PDF или изображение).
- Файл сохраняется на сервере с помощью `FileSystemStorage`.
- В зависимости от типа файла (PDF или изображение) извлекается текст с помощью соответствующих функций.
- После извлечения текста приложение запрашивает токен доступа у GigaChat и отправляет текст на анализ.
- Полученное от GigaChat краткое содержание сохраняется в DOCX-файл, а также отображается в окне Summary в веб-интерфейсе.
- Все сгенерированные файлы (полный текст и краткое содержание) сохраняются на сервере, и пользователю предоставляются ссылки для их скачивания.

## 5. Возврат результатов:

- Результаты работы (краткое содержание в формате DOCX и полный текст с таблицами) возвращаются в виде JSON-ответа. Краткое содержание также отображается в окне Summary на веб-странице.

## **6. Шаблон `upload.html`:**

- Это HTML-шаблон, который используется для загрузки файлов и отображения результатов на веб-странице.

Таким образом, приложение позволяет загружать PDF-файлы или изображения, извлекать из них текст, получать краткое содержание от GigaChat и сохранять результаты в удобном для пользователя формате DOCX.