

Документация, содержащая описание функциональных характеристик экземпляра программного обеспечения, предоставленного для проведения экспертной проверки

Оглавление

1. Общие сведения	3
2. Назначение и решаемые задачи	4
3. Функциональные характеристики	5
3.1. Управление и конфигурирование	5
3.2. Обработка и анализ документов	5
3.3. Хранение и безопасность данных	6
4. Входные и выходные данные	7
4.1 Входные данные	7
4.2. Выходные данные	7
5. Затрачиваемые ресурсы и требования к среде функционирования	8

1. Общие сведения

Наименование программного обеспечения — «Сканикс».

Программное обеспечение «Сканикс» (далее — ПО, программа) предназначено для автоматизированного анализа, распознавания и извлечения структурированных данных из электронных документов в формате PDF.

ПО имеет клиент–серверную архитектуру.

Язык программирования, используемый при разработке: Python.

2. Назначение и решаемые задачи

Основное назначение ПО — автоматизация рутинных процессов обработки больших массивов документов, снижение трудозатрат на ручной ввод данных и минимизация ошибок, связанных с человеческим фактором.

Программа решает следующие задачи:

- Автоматическое определение типа загруженного документа на основе визуальных и структурных признаков.
- Извлечение структурированных данных из документов в соответствии с заданной конфигурацией распознавания.
- Исключение дублирования данных при обработке идентичных файлов.
- Формирование сводных аналитических отчетов по результатам обработки в табличном формате.
- Обеспечение безопасной передачи файлов между клиентом и сервером обработки по защищенному протоколу.

3. Функциональные характеристики

3.1. Управление и конфигурирование

Управление работой ПО осуществляется через интерфейс телеграм–бота. Функциональные возможности управления включают:

- **Авторизация и идентификация:** связка ИНН организации, контактного номера телефона и лицензионного ключа для доступа к функциям.
- **Создание конфигураций распознавания:** пользователь формирует параметры обработки, указывая корневую папку для документов, тип хранилища (SFTP) и наименование конфигурации.
- **Настройка полей отчета:** для каждого типа документа пользователь может редактировать перечень полей в итоговом отчете.
- **Управление конфигурациями:** просмотр списка созданных конфигураций, их архивация, редактирование и выбор конфигурации по умолчанию.
- **Запуск обработки:** инициация процесса распознавания документов на основании выбранной конфигурации по умолчанию.

3.2. Обработка и анализ документов

Серверная часть ПО выполняет следующие функции:

- **Анализ и классификация:** определение типа документа из перечня, согласованного с заказчиком.
- **Распознавание данных:** извлечение значений из полей документа с использованием технологий оптического распознавания символов (OCR) и машинного обучения.
- **Дедупликация:** автоматическое выявление файлов с идентичным содержимым и исключение их повторной обработки для предотвращения дублирования данных в отчетах.
- **Обработка ошибок:** фиксация документов, которые не удалось обработать и логирование причин ошибок в отчете.

3.3. Хранение и безопасность данных

- Передача файлов осуществляется исключительно по протоколу SFTP с использованием аутентификации по SSH–ключам.
- При использовании хранилища на стороне владельца ПО обеспечивается автоматическое удаление загруженных документов и сформированных отчетов.

4. Входные и выходные данные

4.1 Входные данные

Входными данными для ПО являются электронные документы в формате PDF, загружаемые в SFTP-хранилище.

Требования к документам:

- Максимальный размер одного файла: 10 МБ.
- Имя файла должно быть уникальным, не начинаться с точки, не быть пустым, иметь длину не более 100 символов и не содержать запрещенных символов (`\ / : * ? " < > |`).
- Качество сканирования: от 300 до 1200 dpi.
- Угол наклона проекции изображения: не более 5 градусов.
- Документ должен полностью находиться в кадре, иметь контрастный фон, не иметь геометрических искажений, бликов, теней и посторонних объектов.
- На одном листе PDF не должно быть изображено несколько документов.
- Распознавание рукописного текста не поддерживается.

4.2. Выходные данные

Результатом работы ПО является сводный отчет по результатам распознавания в формате .xls (Excel).

Структура отчета:

- **Лист статистики:** Содержит общие показатели, а также таблицу ошибок с указанием идентификатора файла, его имени и причины сбоя.
- **Листы с данными:** Формируются отдельно для каждого типа распознанного документа. Столбцы соответствуют настроенной конфигурации.

5. Затрачиваемые ресурсы и требования к среде функционирования

Вычислительные и временные ресурсы

- Время обработки зависит от объема и сложности документов. В среднем обработка 5 документов занимает около 2 минут.
- Вычислительные ресурсы серверной части (процессорное время, оперативная память) обеспечиваются на стороне владельца ПО и масштабируются в зависимости от нагрузки.

Требования к аппаратно-программным средствам на стороне клиента

Для работы с клиентскими компонентами ПО необходимо устройство (персональный компьютер или мобильное устройство) с доступом к сети Интернет.

- **Для работы с телеграм-ботом:** ОС Windows 7 и выше, macOS 10.11 и выше, Linux, iOS 13.0 и выше или Android 5.0 и выше. Установленное приложение «Телеграм».
- **Для работы с файлами:** персональный компьютер с установленным десктопным SFTP-клиентом (например, FileZilla, WinSCP), поддерживающим протокол SFTP. Актуальные технические требования к SFTP-клиенту уточняйте в документации его разработчика.