

«ТрэкТрэк Распознавание»

ОПИСАНИЕ ПРОГРАММЫ

Листов 7

ООО «ТрэкТрэк»
15 июня 2026 г.



АННОТАЦИЯ

Настоящий документ входит в состав программной документации «ТрэкТрэк Распознавание»

Описание программы раскрывает основные вопросы применения, структуры и функционирования программного обеспечения «ТрэкТрэк Распознавание».

Оглавление

1. Общие положения	4
2. Основные базы данных системы и их состав	5
3. Описание функций системы	6

1. Общие положения

1.1. Общие сведения

Полное наименование системы: «ТрэкТрэк Распознавание».

Сокращенное наименование системы: «ТрэкТрэк Распознавание».

Полное наименование ПО: «ТрэкТрэк Распознавание».

Сокращенное наименование ПО: «ТрэкТрэк Распознавание».

Исполнитель: Общество с ограниченной ответственностью «ТРЭКТРЭК» (ООО «ТРЭКТРЭК»).

1.2 Назначение системы

Ежедневно сотрудники юридических департаментов, правовых департаментов и бухгалтерий сталкиваются с большим количеством входящей корреспонденции: договора, счета-фактуры, товарные накладные, квитанции о приеме грузов и т.п. При этом зачастую документы приходят в бумажном виде либо в виде отсканированных PDF файлов, что делает невозможным выделение и обработку информации в таких документах.

Существующие системы оптического распознавания документов либо не учитывают форматирование и выдают неструктурированный текст, либо выдают документы в общем формате, после чего сотрудникам необходимо вручную переносить данные из распознанных документов для ввода во внутренние информационные системы.

Учитывая потребность в инструменте, который бы закрывал описанные выше задачи и сэкономил бы время сотрудников, возникла необходимость разработки специализированной программы для ЭВМ.

«ТрэкТрэк Распознавание» предназначен для выполнения следующих функций:

- распознавание PDF-файлов и конвертирование их в DOCX с сохранением структуры документа и форматирования
- распознавание PDF-файлов и конвертирование их в XLSX и XML с извлечением атрибутов для счета-фактуры.

1.3 Основные компоненты системы

1.3.1 Программно - аппаратная архитектура системы

Рабочие места пользователей разработаны по технологии тонкого клиента.

Доступ к системе организован в виде web-приложения (фронтэнда). Обмен данными между компонентами системы осуществляется следующими средствами: TCP/IP, RFC, JSON, HTTP/HTTPS.

Фронтэнд обеспечивает взаимодействие пользователя с управляющим сервисом сервера приложений (бэкэндом).

Управляющий сервис сервера приложений (бэкэнд) обеспечивает взаимодействие веб-интерфейса с компонентами системы между собой. Он интегрирует в себя:

– механизмы авторизации, аутентификации и аудита действий пользователей:

- сервисы подсистемы мониторинга деятельности приложения;
- сервисы взаимодействия интерфейса пользователя и БД»;
- сервисы взаимодействия с прикладными микросервисами.

Аутентификация пользователей производится на базе сформированных на сервере учетных записей.

Доступ пользователей и администраторов к системе реализован с рабочих мест пользователей и администраторов по технологии «тонкий клиент», по протоколу HTTP посредством Web-браузера.

Обновление приложений производится путем обновления файлов сервера приложения, SQL-скриптов обновления БД, обновления образов для ПО Docker с удаленного внешнего сервера-репозитории версий приложений.

1.3.2. Функциональная структура системы

Программа для ЭВМ разработана в составе следующих функций:

- функция авторизации и администрирования
- функция распознавания PDF-файла и конвертирования его в формат DOCX
- функция распознавания PDF-файла и конвертирования его в формат XLSX с извлечением атрибутов для счета-фактуры
- функция сохранения распознанного документа в XML-файл для счета-фактуры

2. Основные базы данных системы и их состав

В систему входят следующие базы данных:

- БД Postgres «ТрэкТрэк Распознавание»;
- БД Redis «ТрэкТрэк Распознавание»;

Базы данных используются для внутренних нужд бэкенда, в частности для хранения учетных записей и задач на распознавание и конвертирование.

3. Описание функций системы

«ТрэкТрэк Распознавание» работает в составе следующих функций, подсистем и баз данных:

- функция авторизации и администрирования
- функция распознавания PDF-файла и конвертирования его в формат DOCX
- функция распознавания PDF-файла и конвертирования его в формат XLSX с извлечением атрибутов

3.1 Функция авторизации и администрирования

Функция предназначена для организации доступа пользователей в систему с учетом ролевой модели.

Функция администрирования позволяет администраторам добавлять и удалять учетные записи пользователей.

3.2. Функция распознавания PDF-файла и конвертирования его в формат DOCX

Функция предназначена для преобразования PDF-файла в документ Word с сохранением исходного форматирования. Пользователь выбирает PDF-файл у себя на компьютере. Файл передается в бэкенд, распознается и возвращается DOCX-файл с распознанным текстом с сохранением форматирования, формул и таблиц.

3.3. Функция распознавания PDF-файла и конвертирования его в формат XLSX с извлечением атрибутов

Функция предназначена для преобразования PDF-файла в документ Excel с извлечением атрибутов для файла счета-фактуры.

Пользователь выбирает PDF-файл у себя на компьютере. Файл передается в бэкенд, распознается и возвращается XLSX-файл с извлеченными атрибутами. При этом подсвечиваются поля с обнаруженными ошибками распознавания либо опечатками.

При этом извлекаются следующие атрибуты:

Таблица 1 Атрибуты, извлекаемые из счета-фактуры

№ на образце	Наименование атрибута	Формат данных
2	№	число
3	Дата	дата
4	Продавец	текст
5	ИНН продавца	текст
6	Покупатель	текст
7	ИНН покупателя	текст
8	Наименование товара	текст
9	Условное обозначение единицы измерения	текст
10	Количество	число
11	Цена	число
12	Стоимость без налога	число
13	Сумма налога	число
14	Стоимость с налогом	число

Приложение N 1
к постановлению Правительства Российской Федерации
от 26 декабря 2011 г. N 1137
(в ред. Постановления Правительства РФ от 25.05.2017 № 625 и от 19.08.2017 № 981)

СЧЕТ-ФАКТУРА № 1 от 29 октября 2017 г.
Исправление № --- от ---

Продавец ЗАО "Торговый дом Комплексный"
Адрес: г.Москва Подбельского 17-А, 45
ИНН/КПП продавца: 0056123412 / 0234567891
Грузоотправитель и его адрес: ЗАО "Торговый дом Комплексный", Адрес: г.Москва Подбельского 17-А, 45
Грузополучатель и его адрес: НТЦ "Нанотроника", Адрес: Москва, ул. Зеленоградская, д. 23
К платежно-расчетному документу № _____ от _____
Покупатель: НТЦ "Нанотроника"
Адрес: Москва, ул. Зеленоградская, д. 23
ИНН/КПП покупателя: 0234567801 / 0234567891
Валюта: наименование, код, Российский рубль, 643
Идентификатор государственного контракта, договора (соглашения) (при наличии) _____

8	Наименование товара (описание выполненных работ, оказанных услуг), имущественного права	Код вида товара	Код	9 Единица измерения (количество)	10 Количество (объем)	11 Цена (тариф) за единицу измерения	12 Стоимость товаров (работ, услуг), имущественных прав без налога - всего	13 В том числе сумма акцизов	14 Налоговая ставка	Сумма налога, предъявляемая покупателю	Стоимость товаров (работ, услуг), имущественных прав с налогом - всего	Страна происхождения товара		Регистрационный номер таможенной декларации
												Цифровой код	Краткое наименование	
	Вентилятор BNA/TONE ALPINE 160вт, напольный, оконный	-	796	шт	10	771.92	7719.15	без акциза	18%	1389.45	9108.60	789	Болгария	01234/11020/7654 321
	Вентилятор ZPONS (Тайв.), напольный	-	796	шт	1	1213.01	1213.01	без акциза	18%	218.34	1431.36	156	Китай	11234/11020/7654 321
	Всего к оплате						8932.16	X		1607.79	10539.95			

Руководитель организации или иное уполномоченное лицо (подпись) _____ /Бушуев Э. Р./ (ф.и.о.)
Индивидуальный предприниматель или иное уполномоченное лицо (подпись) _____ (ф.и.о.)
Главный бухгалтер или иное уполномоченное лицо (подпись) _____ /Ковалев С. А./ (ф.и.о.)
(реквизиты свидетельства о государственной регистрации индивидуального предпринимателя)

Примечание. 1. Первый экземпляр счета-фактуры, составленного на бумажном носителе - покупателю, второй экземпляр - продавцу.
2. При составлении организацией счета-фактуры в электронном виде пометка "Главный бухгалтер (подпись) (Ф.И.О.)" не формируется.

Печать формы с датой напечатания: 01.10.2017 14:00:00. Сайт: <http://nfdi.ru>
Размещение данных на 10-значный код страны: 0000000000 (или) другие реквизиты (по желанию клиента)

Рисунок 1 Поля для извлечения из Счет-фактуры

При этом после извлечения атрибутов производятся базовые проверки для контроля правильности распознавания:

Таблица 2 Правила автопроверок для счета-фактуры

Проверяемый атрибут	Правило проверки
Стоимость без налога (12)	= Количество (10) * Цена (11)
Сумма налога (13)	= Стоимость без налога (12) * Ставка налога
Стоимость с налогом (14)	= Стоимость без налога (12) + Сумма налога (13)
Всего к оплате по столбцу Стоимость без налога (12)	= Сумма строк по столбцу Стоимость без налога (12)
Всего к оплате по столбцу Сумма налога (13)	= Сумма строк по столбцу Сумма налога (13)
Всего к оплате по столбцу Стоимость с налогом (14)	= Сумма строк по столбцу Стоимость с налогом (14)

Если счет-фактура состоит из 2 и более листов, то комплектность документа определяется следующим образом:

Таблица 3 Правила автопроверок целостности многостраничного счета-фактуры

Проверяемый атрибут	Правило проверки
Всего к оплате по столбцу Стоимость без налога (12)	= Сумма строк по столбцу Стоимость без налога (12)
Всего к оплате по столбцу Сумма налога (13)	= Сумма строк по столбцу Сумма налога (13)
Всего к оплате по столбцу Стоимость с налогом (14)	= Сумма строк по столбцу Стоимость с налогом (14)