



Система интеллектуального анализа документов

Описание ПО

Всего листов: 6

Аннотация

В данном документе представлено описание программного обеспечения (ПО) - Системы интеллектуального анализа документов (СИАД), включающее в себя обзор функций системы, ее цель, назначение и классификацию по ОКПД2 и классификатору программного обеспечения. Система предназначена для автоматизации проверки документов с использованием OCR, ИИ и интеллектуального анализа, с целью повышения эффективности и безопасности обработки документов.

Содержание

Аннотация	2
1. Общие сведения	4
1.1. Глоссарий	4
1.2. Наименование системы	5
1.3. Область применения системы	5
1.4. Краткое описание возможностей системы	5
1.5. Функции системы	5
1.6. Цель системы	6
1.7. Назначение системы	6
1.8. Класс критичности системы	6
1.9. Классификация системы по ОКПД2 и классификатору программного обеспечения	6

1. Общие сведения

1.1. Глоссарий

В тексте настоящего документа представлены следующие сокращения (см. **Таблица 1-1**).

Таблица 1-1 — Перечень сокращений

Сокращение/аббревиатура	Значение
ИИ	Искусственный интеллект
ИОИ	Интеллектуальная обработка информации
НПА	Нормативно-правовой акт
ПО	Программное обеспечение
СИАД	Система интеллектуального анализа документов
Система	Система интеллектуального анализа документов
ЭП	Электронная подпись
OCR	Optical Character Recognition - технология оптического распознавания символов

В тексте настоящего документа представлены следующие термины и определения (см. **Таблица 1-2**).

Таблица 1-2 — Перечень терминов и определений

Термин	Определение
Граф	Объект, состоящий из вершин и соединяющих их рёбер
Документ, электронный документ	В целях настоящего документа термин используется в следующих значениях: — вне Системы интеллектуального анализа документов – текстовый документ в формате *.pdf, *.doc, *.docx, *.rtf, *.txt, *.csv, *.odt, *.xls, *.xlsx, содержащий в том числе нормативные ссылки на другие документы. Носителем документа является файл (файл с документом); — в Системе интеллектуального анализа документов – сущность, объединяющая все элементы и структуры хранения, которые содержат информацию (в том числе обработанную с использованием искусственного интеллекта) из файла с документом
Проверка	Процесс анализа и оценки документа на соответствие ожиданиям и потребностям Пользователей с помощью правил обращения к Большой языковой модели
Сигнатура файла	Набор байтов, обеспечивающий идентификацию типа файла
Сущность	Однозначно идентифицируемый конкретный или абстрактный объект (элемент)
Файл документа, файл	Цифровой носитель информации, содержащий текстовую и графическую информацию

1.2. Наименование системы

Полное наименование системы – Система интеллектуального анализа документов.

Сокращенное наименование системы – СИАД.

1.3. Область применения системы

Областью применения Системы является деятельность по подготовке и актуализации текстов документов.

1.4. Краткое описание возможностей системы

Система предоставляет следующие возможности:

- поиск по тексту документа ссылок на нормативно-правовые акты (НПА), а также последующее определение актуальности данных ссылок;
- определение возможности подписания документа электронной подписью (ЭП);
- проверка названий документов (НПА), на которые есть ссылки по тексту документа;
- проверка наличия приложений к основной части документа, которые упоминаются по тексту;
- анализ структуры документа на предмет наличия обязательных разделов;
- при наличии в документе раздела «Список ссылок», определяется полнота такого списка в соответствии с составом ссылок, упоминаемых по тексту.

1.5. Функции системы

Функции, которые должны выполняться Системой:

- загрузка файлов документов в Систему;
- распознавание текста файлов документов, загруженных в Систему с помощью ИОИ;
- фрагментация и векторизация текста документа;
- обработка текста документов с помощью ИИ;
- обеспечение хранения документов в Системе;
- интеллектуальная верификация документов;
- поддержка принятия решений по исправлению текста документов в части нормативных ссылок – формирование и предоставление рекомендаций Пользователю Системы;
- построение, хранение и анализ графа связей документов;
- определение, хранение и изменение информации о версиях документов;
- корректировка правил выполнения процессов обработки текста документа с помощью ИИ;

- визуализация результатов работы функций Системы в графическом интерфейсе.

1.6. Цель системы

Цель Системы - автоматизировать проверку подлинности и соответствия документов, используя OCR, извлечение ключевых данных и интеллектуальный анализ. Система должна обеспечивать быструю, точную и надежную верификацию различных типов документов, снижая риски мошенничества и оптимизируя документооборот. Ключевые требования включают загрузку документов, распознавание текста, извлечение данных, верификацию и удобный интерфейс. Система предназначена для организаций, работающих с большим количеством документов, и направлена на повышение безопасности и эффективности процессов проверки.

1.7. Назначение системы

Система предназначена для распознавания содержания электронных документов, их обработки (в том числе с применением технологий ИИ), обеспечения систематизированного хранения документов и связей между ними, а также для поддержки принятия решений по редактированию текстов документов.

Система предназначена для решения следующих взаимосвязанных задач:

- распознавание содержания документов следующих форматов: *.pdf, *.doc, *.docx, *.rtf, *.txt, *.csv, *.odt, *.xls, *.xlsx;
- обработка текста документа, в том числе с применением технологий искусственного интеллекта (ИИ);
- обеспечение систематизированного хранения документов и связей между ними.

Обобщенным назначением Системы является информационная поддержка пользователя в принятии решений о необходимости редактирования текста документа.

1.8. Класс критичности системы

Класс критичности Системы – Business Operational.

1.9. Классификация системы по ОКПД2 и классификатору программного обеспечения

Основной класс:

- 11.04 Средства интеллектуального анализа данных (Data Mining).

Коды продукции в соответствии с Общероссийским классификатором продукции по видам экономической деятельности:

- 58.29.2 Обеспечение программное прикладное на электронном носителе;
- 58.29.3 Обеспечение программное для загрузки;
- 62 Продукты программные и услуги по разработке программного обеспечения; консультационные и аналогичные услуги в области информационных технологий;
- 62.01.29 Оригиналы программного обеспечения прочие.