

Аналитическая обработка документов

Презентация решения

Сентябрь 2019

Аналитическая обработка документов

Область применения, выгоды

Решение по аналитической обработке документов может быть применено в любых областях, связанных с обработкой большого числа документов в электронном формате и документов на бумажных носителях, а также областях, связанных с обработкой текстов со сложно-структурированными данными:

- Банковская сфера
- Бухгалтерский и финансовый учет
- Область законодательства и права
- Работа с нормативными справочниками и конструкторской документацией

Ключевые выгоды от внедрения решения:

- Значительное сокращение трудозатрат на обработку документации
- Исключение влияния человеческого фактора в решении рутинных задач
- Автоматическая проверка сложно зависимых данных на консистентность



Аналитическая обработка документов

Возможные кейсы

Анализ закупочной документации

Выделение номенклатурных позиций, сравнение с архивом прошлых закупок, выделение закупок требующих дополнительного внимания.

Сортировка и разбор входящих документов

Распознавание скан-образов входящих документов, сортировка по соответствующим типам, дополнительный анализ сутевой части

Распознавание чеков

Распознавание скан образов чеков и их последующий анализ для внутреннего бухгалтерского учета

Анализ конструкторской документации

Выделение списка комплектующих из конструкторской документации, формирование номенклатурного перечня, нормализация

Правовая проверка документации

Анализ документов на соответствие нормативно правовой базе

Обработка специализированных документов

Анализ и обработка документов специфичных для конкретной области, например, распознавание паспортов, прав, и т.д.



Аналитическая обработка документов

Архитектура решения

Подсистема распознавания скан-образов

- Определение качества скан-образа
- Распознавание скан-образов

Подсистема лингвистического разбора

- Выделение сущностей
- Формирование модели данных

Подсистема анализа и обработки

- Анализ сутевой составляющей документа
- Преобразование и обработка в соответствии с заложенной логикой

Подсистема распознавания скан-образов

Функциональные особенности

- Определение качества скан-образа (необходимо для последующего анализа сутевой части документа)
- Поддерживаемые входные форматы файлов:
 - Графические: *.PDF, *.JPG, *.TIFF, *.PNG, *.BMP
 - Текстовые: *.PDF, *.DOC, *.DOCX, *.XLS, *.XLSX, *.RTF
 - Архивы графических и текстовых файлов: *.RAR, *.ZIP, *.7z
- Двухэтапный автоматический процесс проверки и исправления распознанных слов при неудовлетворительном качестве скан-образов:
 - Первый уровень: использование общеязыковых словарей
 - Второй уровень: использование настраиваемых словарей для учета специфики конкретной области применения
- Выделение данных из таблиц



Подсистема лингвистического разбора

Функциональные особенности

- Проведение лингвистического анализа и формирование модели данных документа
- Выделение именованных сущностей из неструктурированного текста
- Атрибутирование сущностей (выделение: организаций, имен собственных, географических объектов, дат, чисел, номенклатурных позиций, специфичных терминов...)
- Проведение латентно-семантического индексирования – анализ с точки зрения присутствия в контенте не только ключевых слов, но также их синонимов и слов, близко связанных с тематикой конкретной области применения
- Формирование тегированного структурированного образа документа

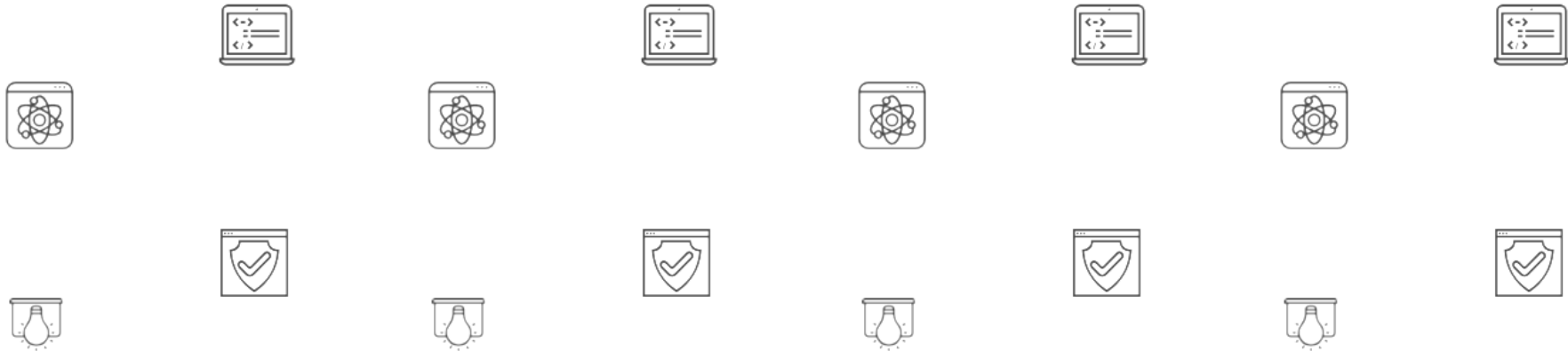


Подсистема анализа и обработки

Функциональные особенности

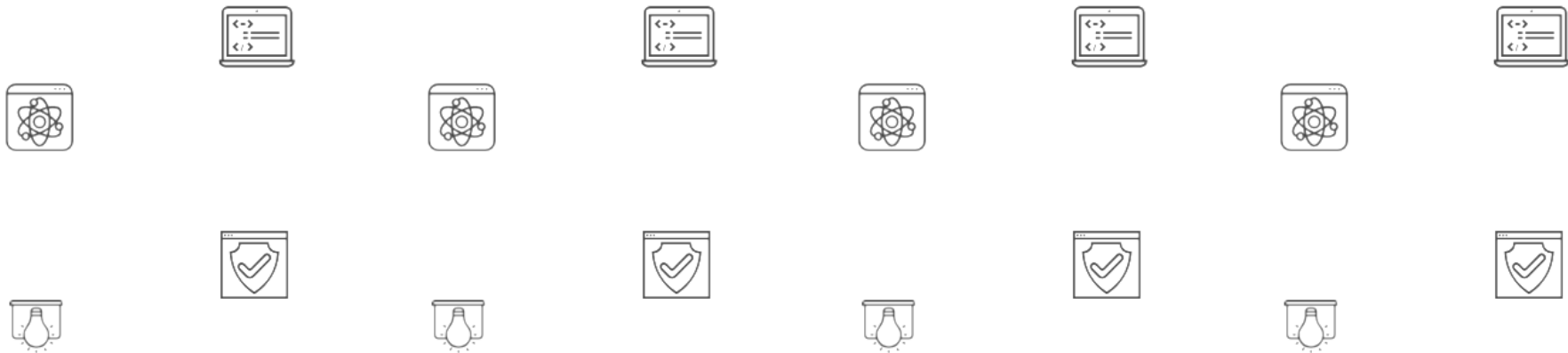
- Определение конкретного типа документа, согласно заданным паттернам тематики конкретной области применения
- Проведение анализа сутевой части документа конкретного типа документа
- Применение нейронных сетей для глубокого анализа документа
- Возможность преобразования входных документов в любые целевые форматы, согласно заданным паттернам тематики конкретной области применения





Интеллектуальный помощник

Дополнительное расширение системы



Интеллектуальный помощник

Описание

Подсистема интеллектуального помощника представляет собой интерфейс получения информации и управления любой системой на естественном языке:

- Производит анализ сутевой части фраз и сообщений, выделение запроса пользователя на конкретные действия:
 - Операции запроса информации
 - Проведение сложной аналитики
 - Команды управления
- Позволяет представить результаты операции в различном формате:
 - Текстовой
 - Табличный
 - Графический
- Возможность использования на входе голосового потока от пользователя

Ключевые выгоды от внедрения решения:

- Предоставление пользователям простого интерфейса к функционально сложным и объемным информационным системам
- Исключение влияния человеческого фактора в решении рутинных задач

