**Бизнес процесс.**

**Потокового сканирования текста для организации электронного архива документов**

**Назначение:**

предназначено для автоматизации перевода бумажных документов в электронный вид, создание электронного документооборота и электронного архива документов.

**Функции потокового сканирования включают:**

–        Предварительное формирование штрих-кодов

–        Формирование сканированного образа документа

–        Формирование электронной учётной карточки, содержащей электронный образ отсканированного документа

–        Передачу корректно отсканированного электронного образа документа в БД

**Возможно осуществление потокового сканирования в пакетном режиме.**

При этом  каждый пакет документов для сканирования может идентифицироваться по:

–        Количеству указанных страниц

–        Пустым страницам-разделителям

–        Штрих-кодам

 Возможный вариант идентификации пакета сканирования определяется настройками сканирования.

**Преимущества использования потокового сканирования для организации электронного архива документов:**

–        Снижение трудозатрат на ввод информации в базы данных

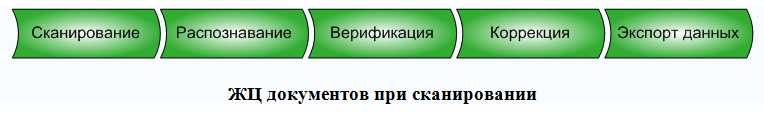
–        Возможность замены бумажного документооборота и бумажных архивов электронной системой документооборота

–        Повышение эффективности управления документами

–        Снижение временных затрат на поиск документов

**Электронный архив документов. Жизненный цикл сканированных документов**

Документы в процессе сканирования проходят определенные этапы, совокупность которых мы обозначим как жизненный цикл сканированных документов (ЖЦ сканирования).



**Этапы ЖЦ сканирования включают:**

1. Сканирование документов
2. Распознавание информации
3. Верификация данных
4. Коррекция данных
5. Экспорт данных

**Этап 1. Сканирование документов**

 Поточные сканеры в большинстве своем оснащены системой автоматической подачи документов, что позволяет достигнуть скорости оцифровки до 200 изображений в минуту на один сканер.

 Для выделения отдельных документов из потока в начале сканируемого документа должна быть нанесена ***штамп-метка/маркер/штрих-код***

 Возможно определение типа и вида документа. Метка должна сопровождаться кодом сканирования, на основании которого может проводиться сопоставление бумажного  и электронного документов.

 Если документ является многостраничным, метка и код сканирования наносятся только на первый лист документа, и в дальнейшем при вводе массива документов они служат признаком конца текущего документа и началом нового.

 Место нанесения метки и кода сканирования может находиться в любой свободной от текста части документа. При отсутствии на лицевой стороне документа свободного места их можно нанести с обратной стороны листа.

**Возможно выделение следующие подэтапы для первого этапа:**

–        **Формирование штамп-метки/маркера/штрих-кода** специализированной программой

* Одномерные штрих-коды
* Двухмерные штрих -коды

–      **Распечатка штамп-метки/маркера/штрих-кода**

–        **Подготовка комплекта документов** для сканирования, нанесение штамп-метки/маркера/штрих-кода

–        **Сканирование пакетов документов**:

* Исправление перекосов, искажений, удаление частей изображений и  др.
* Сборка документов по типам, видам
* Корректировка качества сканирования

Сканирование документов можно осуществлять в рамках локальной сети, удаленно по средствам http, а также с использованием интерфейсов TWAIN, WIA и ISIS.

**Результат выполнения этапа:** отсканированный документ в формате PDF, TIFF, JPEG, JPEG2000, BMP, PNG, PCX, DCX, DjVu, JBIG2.

**Этап 2. Распознавание информации**

 На этапе распознавания используются программные OCR компоненты.

 Системы, поддерживающие распознавание текста, позволяют:

* **Осуществлять распознавание печатного текста на разных языках**, таких как: латинский, греческий, тайский, армянский, японский, корейский, китайский и др. Перечень поддерживаемых языков зависит от возможностей конкретной программы.
* **Осуществлять распознавание документов, содержащих текст сразу на разных языках**. При этом, система сама распознает какой текст к какому языку относится.
* **Распознавать текст различного типа**: типографская печать, печатная машинка, матричный принтер, MICR (E13B), OCR-A, OCR-B.
* **Распознавать штрих коды.**

 В качестве настроек обычно можно указывать максимальное количество неуверенно распознанных символов на страницы для дальнейшей корректировки содержания на этапе Верификации.

**Результат  выполнения этапа:** отсканированный документ в формате Microsoft PDF

**Высокое качество распознавания и восстановления оформления документа:**

* Поддержка 191 языка распознавания печатного текста, включая языки на основе латинского, греческого, тайского, армянского и кириллического алфавитов, а также японский, корейский, китайский и тайский языки.
* Распознавание многоязычных документов.
* Режим быстрого распознавания увеличивает скорость обработки изображений хорошего качества в 2-2,5 раза.
* Поддержано распознавание различных типов текста: типографская печать, печатная машинка, матричный принтер, MICR (E13B), OCR-A, OCR-B.

**Распознавание штрих-кодов:**

* Поддержано распознавание наиболее популярных одномерных и двумерных штрих-кодов.
* Автоматически идентифицируются и распознаются штрих-коды, расположенные на странице под любым углом к горизонтали.

**Контроль качества распознавания:**

* Установка порога качества распознавания изображений, основанная на допустимом количестве неуверенно распознанных символов на страницу.

**Преобразование PDF-файлов**

* Определение текстового слоя и его целостность для обеспечения более быстрого и качественного преобразования PDF-файлов.
* В создаваемых PDF-файлах сохраняются метаданные документа (название, автор, тема, ключевые слова) и восстанавливаются гиперссылки внутри документа.
* Поддержаны алгоритмы шифрования и другие средства разграничения доступа: сохранение результатов распознавания в PDF-файл, защищённый паролем.

**Этап 3. Верификация данных**

Этап верификации включает:

* Проверку полученной от сканирования информации на присутствие значений
* Редактирование распознанного текста
* Формирование учетной карточки документа
* Заполнение атрибутов карточки отсканированного документа
  + Автоматически
  + Вручную

 На этом этапе, в случае обнаружения ошибок при сканировании и всех неточностей, информация помечается специальными маркировками для дальнейшего исправления на этапе Коррекции. При необходимости, этап Верификации можно проводить повторно.

После верификации и устранения ошибок, страница будет отправлена на дальнейшую обработку или проэкспортирована. При неудовлетворительных результатах распознавания, оператор станции верификации может отправить страницу на повторную обработку.

**Результат  выполнения этапа:**  заполненная учетная карточка документа и отсканированный документ в формате PDF. Учетная карточка содержит уникальный ID номер (сформированный по распознанному ШК) и уникальную ссылку на сканобраз документа.

**Этап 4. Коррекция данных**

 Оператор обрабатывает информацию, полученную на стадии верификации данных, вносит изменения в результирующую запись вручную (ввод с клавиатуры, выбор из списка заранее заданных ответов). В некоторых случаях может объединяться с предыдущим этапом верификации данных.

**Результат  выполнения этапа:**  заполненная отредактированная учетная карточка документа и отсканированный документ в формате PDF

**Этап 5. Экспорт данных**

 Оцифрованные структурированные данные выгружаются в необходимом формате для дальнейшего анализа результатов сканирования.

 При этом экспорт данных может осуществляться:

* В локальную или сетевую папку
* В корпоративный портал
* В систему электронного документооборота/электронный архив

 Также может осуществляться рассылка данных по электронной почте.

**Результат выполнения этапа:** размещение/отправка итогового документа на требуемом ресурсе