**ИТ в ПД, 3 курс. Строго для группы 16!**

**Изучите теоретическое содержание учебного материала.**

**Выполните задания по теме.**

**ВНИМАНИЕ!!! Указания к выполнению.**

Работа – конспект и ответы на задания – выполняются **письменно на двойном листе в клетку или в новой тетради**.

**НА ПОЧТУ ПРИСЫЛАТЬ НЕ НАДО, ПРОВЕРЮ ПО ОКОНЧАНИИ ДИСТАНТА**

**ТЕМА. Базовые системные программные продукты и пакеты прикладных программ в области профессиональной деятельности**

**1. Сканеры**

**Сканер** — устройство ввода, назначение которого оцифровка информации, хранящейся на нецифровом носители, с её последующей передачей в память ПК в виде графики.

Ввод данных в компьютер — это одна из самых утомительных и подверженных ошибкам операций, сканеры облегчают эту работу.

  

Принцип работы.

Лампа освещает сканируемый текст, отражённые лучи попадают на фотоэлемент, состоящий из множества светочувствительных ячеек. Каждая из них под действием света приобретает электрический заряд. Аналого-цифровой преобразователь ставит в соответствие каждой ячейке числовое значение, и эти данные передаются в компьютер.

Основные характеристики сканера

* **Тип сканера**. По типу, сканеры подразделяются на планшетные, протяжные и слайд-сканеры. **Планшетный** тип сканера наиболее распространен. Конструктивно сканер состоит из стеклянной основы (планшета) и считывающего механизма под ней. В процессе сканирования документ или книга неподвижны, а информация снимается за счет перемещения линии считывающих фотоэлементов вдоль оригинала. Сканер имеющий **протяжной** тип, осуществляет сканирование лишь отдельных листов бумаги, которые проходят через считывающую светочувствительную линию. К сожалению, такой подход не позволяет оцифровывать объемные носители информации. **Слайд-сканер**. Данный тип сканера используется для сканирования материалов имеющих прозрачную или полупрозрачную основу. К таким нецифровым носителям, например, можно отнести фотопленку или рентгеновский снимок.
* **Тип датчика**

Contact Image Sensor, CIS — представляет собой линию фотоэлементов, которая перемещается вдоль сканируемого материала, и строка за строкой передает информацию с носителя в виде электроимпульсов. Для подсветки оригинала используются светодиоды, расположенные вблизи фотоэлементов. Недостатком данного типа датчиков является малая глубина резкости.

Charge-Coupled Device, CCD – не что иное, как интегральная микросхема, обладающая линией фоточувствительных элементов. При построении картинки, используется оптическая конструкция -включающая в себя зеркало и объектив. Для подсветки сканируемого материала применяется люминесцентная подсветка. Плюсом CCD-сканеров является хорошая глубина резкости и цветопередача. К минусам, можно отнести большую толщину и вес датчика, а также стоимость в сравнении с CIS.



* **Разрешение** сканера зависит от количества фоточувствительных элементов линейки на один дюйм по оси X и минимальным интервалом хода шагового двигателя, также на дюйм, по оси Y. **Разрешение – основной параметр при выборе сканера**, подавляющее большинство моделей имеет минимальное разрешение 600 х 1200 точек на дюйм (dpi). Почему минимальное? Существует **понятие улучшенного разрешения**, когда готовое изображение формируется по принципу интерполяции. Для того чтобы получить картинку превышающую разрешение оригинала, пространство между соседними точками заполняется по принципу градации яркости и цвета исходя из данных полученных оптическим способом о цвете и яркости оригинальных точек.
* **Оптическая плотность**. Способность сканера отличить сканируемый материал от «полной темноты», своего рода, параметр «слепоты» считывающих светочувствительных датчиков. Чем выше чувствительность, тем лучше результат при сканировании темных малоконтрастных исходников.
* **Глубина цвета.** Влияет на цветопередачу при сканировании исходных материалов. Различают два вида цветопередачи: **внутренняя** – количество цветов, различаемых системой сканера; **внешняя** – количество цветов, которые сканер отправляет на ПК. В основном используется 24-битная цветопередача, что вполне достаточно для задач офиса или дома. Но, если работать с графикой, необходим сканер с большей разрядностью.
* **Скорость сканирования.** Чем выше этот показатель, тем лучше. Но, нужно помнить, что скорость также зависит от выбранного разрешения и площади сканируемого материала.
* **Формат сканирования**. Максимальный формат оригинала, который может отсканировать устройство - поддерживаемый формат нецифрового носителя (А4, А3). Сканеры, поддерживающие формат A4, получили большее распространение.

**2. Системы оптического распознавания символов (OCR)**

Очень часто появляется необходимость перевести в электронный вид текст каких-то документов, или даже книг. Можно затратить определённое время и просто набрать этот текст с помощью клавиатуры. Но, чем больше исходный текст, тем больше времени будет затрачено на его ввод в память компьютера. Поэтому для ввода текстов в память компьютера с бумажных носителей используют **сканеры** и **программы распознавания символов**.

**Системы оптического распознавания символов** используются при создании электронных библиотек и архивов путем перевода книг и документов в цифровой компьютерный формат.

После обработки документа **сканером** получается графическое изображение документа (графический образ). Но графический образ еще не является текстовым документом. Человеку достаточно взглянуть на лист бумаги с текстом, чтобы понять, что на нем написано. С точки зрения компьютера, документ после сканирования превращается в набор разноцветных точек, а вовсе не в текстовый документ.

Проблема распознавания текста в составе точечного графического изображения является весьма сложной. Подобные задачи решают с помощью специальных программных средств, называемых **средствами распознавания образов** - OCR (Optical Character Recognition — оптическое распознавание символов). Современные алгоритмы распознавания текста не ориентируются ни на конкретный шрифт, ни на конкретный алфавит. Системы оптического распознавания символов являются «самообучающимися» (для каждого конкретного документа они создают соответствующий набор шаблонов символов), и поэтому скорость и качество распознавания многостраничного документа постепенно возрастают.

Наиболее широко известна и распространена такая программа отечественных производителей — **ABBY FineReader**.

Эта программа предназначена для распознавания текстов на русском, английском, немецком, украинском, французском и многих других языках (на **179** языках), а также для распознавания смешанных двуязычных текстов.

Возможности программы **ABBY FineReader**:

* Работает с разными моделями сканеров.
* Позволяет из бумажных документов, PDF-файлов и цифровых фото сделать редактируемый текст.
* Позволяет объединять сканирование и распознавание в одну операцию, работать с пакетами документов (многостраничными документами) и с бланками.
* Позволяет редактировать распознанный текст и проверять его орфографию.
* Сохраняет внешний вид документа, а также его структуру, то есть, расположение слов, абзацев, таблиц, изображений, заголовков и нумерация страниц останутся такими же, как и в оригинале.
* Экспортирует тексты в Word, Excel, PowerPoint или Outlook.

Преобразование бумажного документа в электронный вид происходит в пять этапов. Каждый из этих этапов программа ***FineReader*** может выполнять как автоматически, так и под контролем пользователя. Если все этапы проводятся автоматически, то преобразование документа происходит за один прием.

*Пять этапов процесса обработки документа* с помощью программы ***ABBY FineReader***:

* 1. Сканирование документа (кнопка Сканировать).
	2. Сегментация документа (кнопка Сегментировать).
	3. Распознавание документа (кнопка Распознать).
	4. Редактирование и проверка результата (кнопка Проверить).
	5. Сохранение документа (кнопка Сохранить).

Рассмотрим этапы более подробно.

1) На этапе **сканирования** производится получение изображений при помощи сканера и сохранение их в виде, удобном для последующей обработки. Чтобы начать сканирование, надо включить сканер и щелкнуть на кнопке **Сканировать**.

2) Второй этап работы — **сегментация**, разбиение страницы на блоки текста. Если страница содержит колонки, иллюстрации, врезки, подрисуночные подписи или таблицы, то порядок распознавания требует коррекции. Содержимое страницы разбивается на блоки, внутри каждого из которых распознавание осуществляется в естественном порядке. Блоки нумеруются, исходя из порядка включения их в документ. При автоматической сегментации (кнопка**Сегментировать**) определение границ блоков осуществляется автоматически. При этом учитываются поля документа, просветы между колонками, рамки.

3) Процесс **распознавания** текста после сегментации начинается с щелчка на кнопке **Распознать** и полностью автоматизирован.

4) Когда распознавание данной страницы завершается, полученный текстовый документ отображается в окне **Текст**. Заключительные этапы работы позволяют **отредактировать** полученный текст с помощью средств, напоминающих текстовый редактор **WordPad**. **Провести проверку**орфографии с учетом трудностей распознавания позволяет кнопка **Проверить**.

5) По щелчку на кнопке **Сохранить** запускается **Мастер сохранения** результатов. Он позволяет сохранить распознанный текст или передать его в другую программу (например, в **Microsoft Word**) для последующей обработки полученный текст можно сохранить в виде форматированного или неформатированного документа.

Есть и другие программы, позволяющие распознавать документы.

**CuneiForm**— это программа для оптического распознавания текста документов, представленных в виде изображений, в редактируемый вид. Результаты работы программы можно редактировать в офисных программах и текстовых редакторах, сохранять в популярных форматах, проводить по ним полнотекстовый поиск.

**ABBYY FineReader Professional 11.0.513.139-**Профессиональная программа для оптического распознавания текста.

**ABBYY PDF Transformer 3.0 -**Универсальное решение для преобразования документов PDF с возможностью оптического распознавания текста.

1. **Контрольные вопросы (выполняются письменно на двойном листе):**
	1. Дать определение сканера.
	2. Опишите принцип работы сканера.
	3. Самостоятельно составьте и начертите схему - классификацию «Характеристики сканеров».
	4. Как производится распознавание документа?
	5. Опишите этапы процесса обработки документа с помощью программы ABBY FineReader.