

## Обучающиеся группы № 36.

Приветствую вас на дистанционном обучении

Работы прошу высылать на адрес [msn@apt29.ru](mailto:msn@apt29.ru)

(Мамонов Сергей Николаевич)

Задание:

1. Внимательно ознакомиться с теоретическими сведениями лекционного занятия.
2. Сделать краткий конспект теоретического материала в тетради.
3. Письменно ответить на контрольные вопросы.
3. Результат работы прислать для проверки на электронный адрес [msn@apt29.ru](mailto:msn@apt29.ru)

### Лекционное занятие 4

#### Технологии обработки текстов

Одним из самых массовых видов информационных технологий являются технологии обработки текстов.

С распространением персональных компьютеров как основных инструментов создания и обработки текстов появилось новое значение термина «документ». Понятие **«текстовый документ»** мы будем трактовать следующим образом: это текст, созданный на компьютере и сохраненный в файле. Документ, созданный на компьютере, помимо текста, может включать в себя формулы, диаграммы, рисунки, таблицы, колонтитулы и пр. В тексте могут использоваться разнообразные шрифты, меняться размеры полей. Принято говорить, что такой документ отформатирован определенным образом. Примерами документов, с которыми нам приходится иметь дело в повседневной жизни, могут служить письма, заявления, объявления, приказы, инструкции, статьи, рассказы и пр. Применяемые для целей создания и обработки документов прикладные программы называются **текстовыми редакторами и текстовыми процессорами**.

Возможности текстовых процессоров выходят за рамки базовых функций. Текстовый процессор предоставляет пользователю возможность автоматизированного создания алфавитных и предметных указателей, оглавлений, сносок, управление печатью. В текст можно включать рисунки, таблицы, формулы и пр. Можно форматировать текст.

**Форматирование** определяет внешний вид текста, а не его содержание. Отметим, что одно с другим часто бывает связано. Например, при составлении официального заявления на чье-либо имя реквизиты адресата записываются в правом верхнем углу. Многие документы представляются в форме таблиц. Заголовки разделов документа принято выделять жирным шрифтом или подчеркиванием, располагать посередине строки.

Файл с отформатированным текстом включает в себя как содержание текста, так и данные по его форматированию при выводе на экран или на печать.

Популярным текстовым процессором является Microsoft Word, входящий в состав пакета Microsoft Office. В настоящее время все большую популярность получает текстовый процессор OpenOffice.org Writer, входящий в состав универсального офисного пакета OpenOffice.org, работающего на разных операционных платформах. Этот пакет относится к свободно распространяемому программному обеспечению.

Развитие текстовых процессоров идет по пути повышения их интеллектуальных возможностей, к числу которых относятся автоматическая проверка правописания и перевод текста с одного языка на другой.

### **Проверка правописания**

Под проверкой правописания следует понимать проверку орфографии, грамматики и стилистики. Проверка правописания может быть произведена только для текста, введенного на одном из языков, которые текстовый процессор поддерживает. Чаще всего текстовый процессор сам определяет язык текста.

**Проверка орфографии** возможна при наличии встроенных словарей. Следует обратить внимание, что текстовый процессор пропускает слова, которые написаны правильно с точки зрения орфографии, однако неверно используются в контексте. Например: «березовая почка», «березовая печка», «березовая точка».

Текстовые процессоры предоставляют пользователю возможность создавать собственные словари, редактировать существующие (добавлять и удалять слова). Пользовательские словари используются для хранения правильно написанных слов, отсутствующих в основном словаре. Файлы словарей имеют расширение dic, однако являются обычными текстовыми файлами и с ними можно работать, например в текстовом редакторе Блокнот.

**Проверка грамматики** выявляет такие ошибки, как неправильное использование предлогов, нарушение согласования слов в предложении и т.д. Грамматическая проверка производится на основе фиксированного набора правил. Набором используемых правил можно управлять, по мере надобности подключая или отключая определенные правила.

**Проверка стилистики** позволяет выявлять и в документе малоупотребимые, просторечные слова и выражения.

Параметрами проверки правописания (орфографии, грамматики, стиля) можно управлять (настраивать). Например, в текстовом процессоре Microsoft

Word эту возможность можно реализовать, выполнив команду **Сервис** → **Параметры** и переключившись на вкладку **Правописание**. По умолчанию установлены оптимальные для большинства случаев параметры проверки правописания.

### **Оптическое распознавание текстов**

Сканер позволяет вводить графическую информацию с листа бумаги. В настоящее время существует потребность переноса текстовой информации с бумажных носителей (книги, журналы, газеты) в память компьютера и сохранения этой информации в текстовых файлах. Это во многом связано с созданием электронных библиотек, куда кроме современных изданий помещаются и издания, вышедшие в свет много лет назад и поэтому не имеющие электронной версии. После **сканирования** вся информация, в том числе и текстовая, представлена как изображение. Такой текст можно просмотреть, распечатать, а редактировать только с помощью графического редактора, который не обладает удобными средствами для работы с текстовой информацией.

Обычно при сканировании получают файл форматов **TIFF**, **TGA**, **PCX** или **BMP**. Для перевода в более общепринятый сегодня формат **JPEG** используют специальные программы преобразования графических форматов — **конверторы**.

Программа сканирования обычно имеет *следующие возможности*:

1. перед началом работы можно задать режим сканирования: выбрать оптическое разрешение — обычно можно задать значения 300, 600, 1200 dpi (если сканер поддерживает разрешения до 1200 dpi) — и палитру — черно-белую, цветную (256 цветов), цветную с максимально возможным цветовым разрешением;
2. очень удобно использовать *режим предварительного сканирования*. В этом режиме происходит быстрое сканирование изображения в режиме низкого разрешения, после чего пользователю предоставляется возможность выбрать более точно область сканирования или поправить расположение исходного изображения на планшете;
3. при сканировании можно сразу задать масштаб и применить несколько простейших фильтров.

Для преобразования изображения в текстовый документ предназначены специальные программы **оптического распознавания текстов**. Наиболее известной и часто используемой является программа Fine Reader (разработчик — фирма Abby). Можно сказать, что с помощью программ распознавания компьютер учится «читать» печатные и рукописные документы/

### **Основные этапы ввода в память компьютера печатных текстовых документов:**

*1-й этап — сканирование.* Положите лист сканируемого текста на планшет и выберите режим сканирования, необходимую область

сканирования либо всю страницу целиком. Укажите ориентацию сканируемого текста — книжную или альбомную. Процесс сканирования страницы может занять несколько минут. После того как процесс завершен, в левом рабочем поле появится иконка, обозначающая отсканированную страницу.

Аналогично выполните сканирование остальных страниц.

*2-й этап — распознавание.* Выберите на левом поле страницу, с которой вы будете работать, при этом справа высветится ее отсканированное изображение. Выберите язык распознаваемого текста: русский, русско-английский, английский. Затем необходимо выполнить разбивку текста на блоки. Разбивка может осуществляться автоматически, с помощью встроенной процедуры, либо вручную. Ручная разбивка может потребоваться, если структура текста на странице достаточно сложна — в текст врезаны таблицы, диаграммы, рисунки. После разбивки текста можно приступать собственно к распознаванию. Распознавание производится по блокам и в зависимости от мощности компьютера может занимать от нескольких секунд до нескольких минут. На Рабочем поле уже распознанные страницы отличаются наличием в иконке второго листа.

После того как для всех страниц выполнено распознавание, необходимо выполнить проверку орфографии. Это позволит исправить некоторые возможные ошибки распознавания. Окончательную коррекцию следует выполнить вручную.

*3-й этап — заключительный.* Полученный документ можно распечатать, сохранить в виде текстового файла или же экспортировать непосредственно в текстовый редактор Word.

### **Понятие шаблона документа**

**Шаблон** – это файл в формате dot, предназначенный для создания типичных файлов в формате doc и содержащий некоторую информацию, которая будет использоваться без изменений во многих документах пользователя. Все документы по умолчанию создаются на основе шаблона обычный. Существует большое количество шаблонов писем, факсов, отчетов, разных бланков и т.п.

В шаблоне хранятся разнообразные элементы, составляющие основу документа Word. Для создания шаблона нужно войти в редактор Word и создать там новый документ с помощью команды **Файл/Создать**. В появившемся диалоговом окне «Создание документа» выбрать вкладку «Шаблоны на моем компьютере», а затем вкладку «Другие документы». В открывшейся вкладке выбрать необходимый ярлык.

Далее, с помощью диалога Мастера, создают необходимый шаблон документа.

### **Контрольные вопросы:**

1. Назовите базовые функции текстового редактора.
2. Что такое текстовый документ?

3. Что такое форматирование текста?
  4. Перечислите интеллектуальные функции текстового процессора.
  5. Как производится проверка орфографии?
  6. Проведите проверку правописания какого-либо документа.
- Получите статистический отчет об удобочитаемости этого документа.