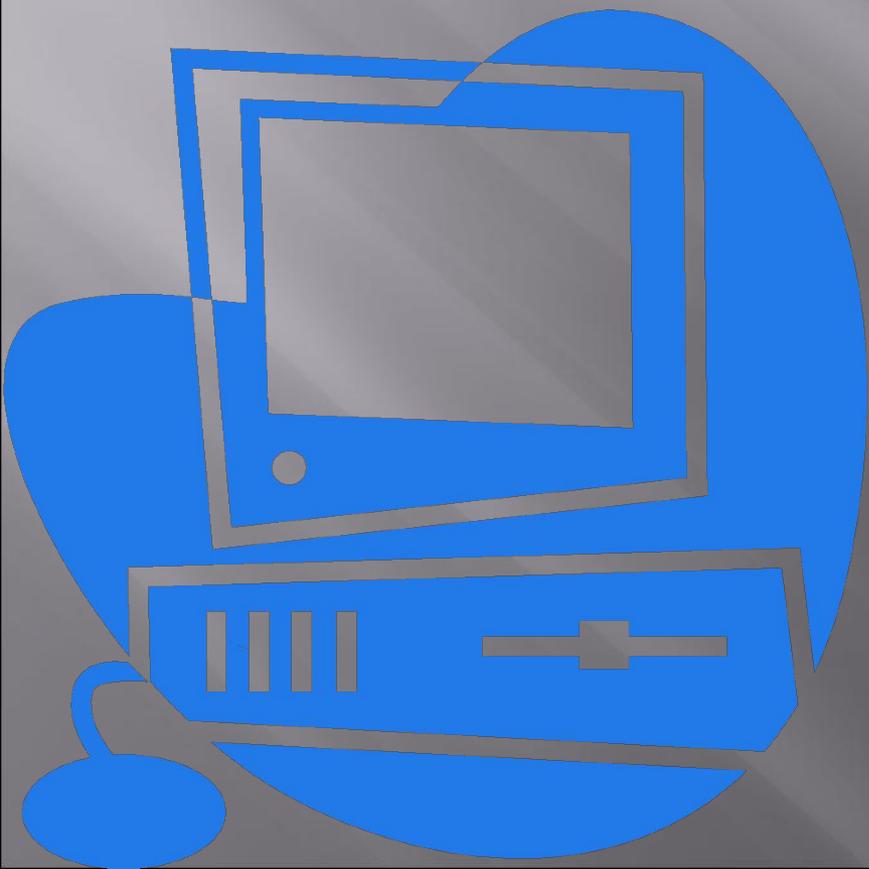


ОБРАБОТКИ ТЕКСТОВОЙ И ИЗОБРАЗИТЕЛЬНОЙ ИНФОРМАЦИИ



Выполнил:
Студент третьего курса
Группа ДТмоБ-3-1
Муродов М.Х.

Проверил:
Ткачук Ю.Н

Технология обработки текстовой информации

1. Создание и редактирование документов
2. Форматы текстовых файлов
3. Форматирование документа
4. Гипертекст
5. Компьютерные словари и системы машинного перевода текстов
6. Системы оптического распознаения документов



Создание и редактирование документов

Для обработки текстовой информации на компьютере используют приложения общего назначения –

текстовые редакторы,

которые позволяют создавать, редактировать, форматировать, сохранять и распечатывать текст



Текстовые редакторы

```
graph TD; A[Текстовые редакторы] --- B[Простые текстовые редакторы: простейшее форматирование шрифта (Windows Блокнот)]; A --- C[Текстовые процессоры: вставка списков и таблиц, проверка орфографии и т.д. (Microsoft Word и StarOffice Writer)]; A --- D[Настольные издательские системы: подготовка к изданию журналов и газет (Adobe PageMaker)];
```

Простые текстовые редакторы: простейшее форматирование шрифта (Windows Блокнот)

Текстовые процессоры: вставка списков и таблиц, проверка орфографии и т.д. (Microsoft Word и StarOffice Writer)

Настольные издательские системы: подготовка к изданию журналов и газет (Adobe PageMaker)

Создание документа

- ✓ Создание документа начинается с выбора *шаблона*.
- ✓ Для создания документов со сложной структурой используют *Мастера*
- ✓ В процессе создания документа в текстовом редакторе пользователь вводит символы с клавиатуры

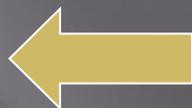


Редактирование документа

Редактирование документа производится путём копирования, перемещения или удаления выделенных символов или фрагментов текста.

Чтобы заменить одно многократно встречающееся слово на другое:

[Правка – Заменить]



Технология обработки изобразительной информации
Технология обработки изобразительно информации базируется на 3 этапах:

1. изобразительная информация 2. система обработки 3. технология обработки или последовательность операций. Когда говорим об изображении мы имеем совокупность свойств на входе системы и совокупность свойств которую должны получить на выходе системы. На входе системы изображение называется оригиналом.

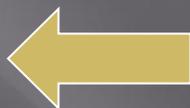
Вставка объектов в документ

Механизм встраивания
и внедрения объектов

(OLE - Object Linking Embedding)

позволяет копировать и вставлять
объекты из одного приложения в другое

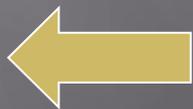
[Вставка – Объект]



Проверка орфографии и синтаксиса

Для проверки орфографии и синтаксиса используются специальные программные модули, которые включаются в состав текстовых процессоров и редакционно-издательских систем.

Наиболее часто встречающиеся опечатки исправляет функция *Автозамена*



Форматы текстовых файлов

Формат файла определяет способ хранения текста в файле

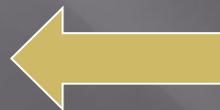
Для преобразования текстового файла из одного формата в другой используют

программы - конверторы

Форматы текстовых файлов

Универсальные
(могут быть прочитаны большинством редакторов)

Оригинальные
(используются отдельными текстовыми редакторами)



Наиболее распространенные форматы текстовых файлов

Только текст (Text Only) (**TXT**)

Текст в формате **RTF**
(Rich Text Format)

Документ Word (**DOC**)

Works 4.0 для Windows (**WPS**)

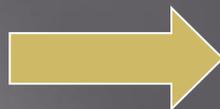
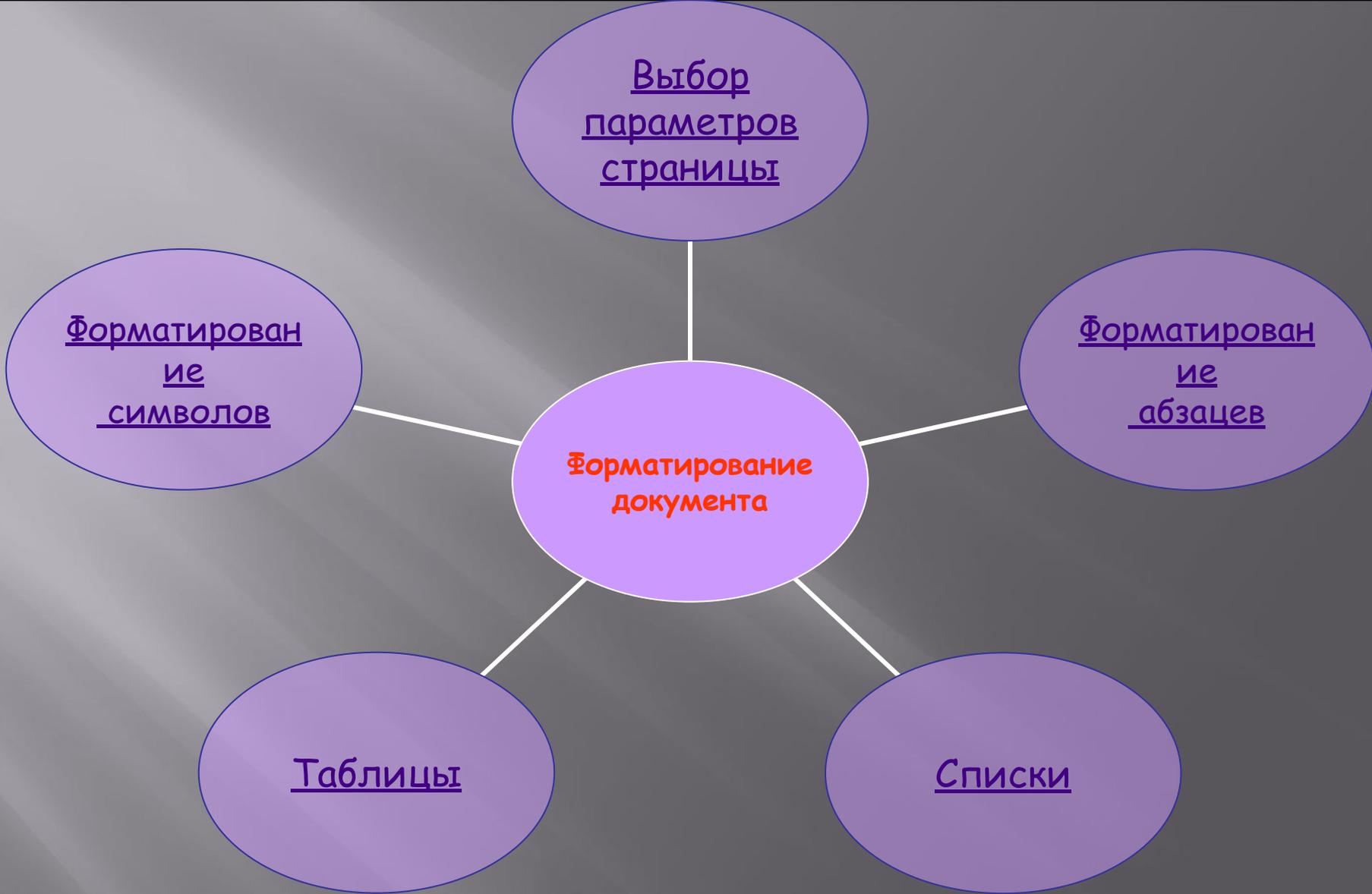
HTML-документ (**HTM, HTML**)



Сохранение и открытие документа в определенном формате

- ▣ Ввести команду **[Файл – Сохранить как]**
- ▣ Присвоить документу имя
- ▣ Выбрать в перечне требуемый формат





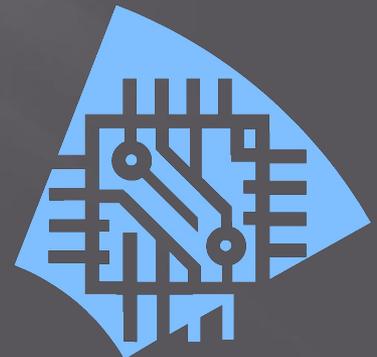
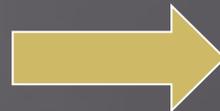
Выбор параметров страницы

- Существуют две возможные ориентации страницы: **книжная** и **альбомная**
- На странице можно устанавливать требуемые размеры полей, использовать колонтитулы [**Файл – Параметры страницы**]
- Страницы документа требуется нумеровать [**Вставка - Номера страниц**]



Форматирование абзацев

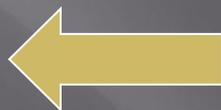
В компьютерных документах **абзац** – любой текст, заканчивающийся управляющим символом (маркером) конца абзаца. Ввод конца абзаца обеспечивается нажатием **{Enter}**.



Выравнивание абзацев

Выравнивание отражает расположение текста относительно границ полей страницы. Чаще всего используют четыре способа выравнивания абзацев:

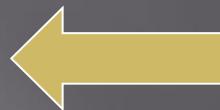
- **По левому краю** – левый край ровный, а правый рваный.
- **По центру** – оба края имеют неровные очертания, однако каждая строка абзаца симметрична относительно середины.
- **По правому краю** – правый край ровный, а левый рваный.
- **По ширине** – оба края ровные, то есть располагаются точно по границам страницы. В этом случае последняя строка абзаца ведет себя как при левостороннем выравнивании.



Отступы и интервалы

Чаще всего абзац начинается отступом первой строки. Весь абзац целиком может иметь отступы слева и справа, которые отмеряются от границ полей страницы.

Для выравнивания абзаца ввести команду
[Формат – Абзац]



Списки

Списки применяются для размещения в документе различных перечней. Существуют списки различных типов:

- ▣ **Нумерованные списки**, когда элементы списка сопровождаются арабскими или римскими числами и буквами;
- ▣ **Маркированные списки**, когда элементы списка отмечаются с помощью специальных символов-маркеров.

Возможно создание и вложенных списков, причем вкладываемый список может по своему типу отличаться от основного. Ввести команду **[Формат-Список]** и на диалоговой панели *Списки* на вкладке *Многоуровневый* выбрать требуемый тип многоуровневого списка.



Таблицы

Таблица – объект, состоящий из строк и столбцов, на пересечении которых образуются ячейки. С помощью таблиц можно форматировать документы. При размещении в таблице чисел можно производить над ними вычисления.

Преобразовать имеющийся текст в таблицу
можно с помощью команды
[Таблица - Преобразовать в таблицу]



Вставить документ в таблицу:
[Таблица – Вставить в таблицу]

Автоматически отформатировать внешний вид таблицы:
[Таблица – Автоформат]

Форматирование таблицы вручную:
[Формат – Границы и заливка]

Задать точную ширину столбца:
[Таблица – Высота и ширина ячейки]

Форматирование символов

Символы являются теми основными объектами, из которых состоит документ.

Символы – это буквы, цифры, пробелы, знаки пунктуации, специальные символы.

Символы можно форматировать.

Основные свойства символов:
шрифт, размер, начертание и цвет.





Гипертекст

Гипертекст позволяет структурировать документ путем выделения в нем слов-ссылок (**гиперссылок**).

Гиперссылка состоит из двух частей: указателя ссылки и адресной части ссылки.

Указатель ссылки – объект, который визуально выделяется в документе.

Адресная часть гиперссылки представляет собой название закладки в документе, на который указывает ссылка.

Для создания гиперссылки:
ввести команду **[Вставка-Гиперссылка]**.

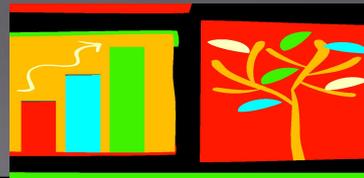
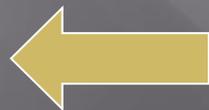


Компьютерные словари и системы машинного перевода текстов

Компьютерные словари могут содержать переводы сотен тысяч слов и словосочетаний, а также предоставляют пользователю дополнительные возможности:

- словари могут являться многоязычными
- словари могут кроме основного словаря общеупотребительных слов содержать десятки специализированных словарей по областям знаний
- словари обеспечивают быстрый поиск словарных статей
- словари могут являться мультимедийными.

Среди российских словарей следует выделить словарь **Lingvo**, который содержит более 1,2 миллиона слов и словосочетаний, систему электронных словарей «**Контекст**» и словарь «**Мультилекс**».

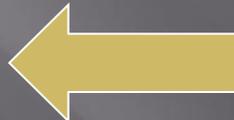




Системы машинного перевода

Системы машинного перевода осуществляют перевод текстов, основываясь на формальном «знании» языка и использовании словарей.

Лучшими среди российских систем машинного перевода считаются **PROMT** и «Сократ».

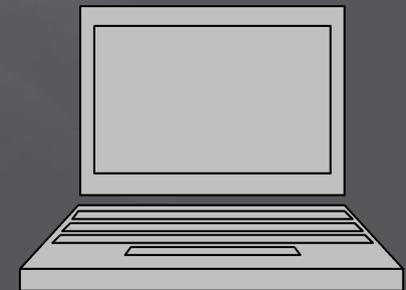


Системы оптического распознавания документов

Системы оптического
распознавания
символов (**FineReader**, **CuneiForm**)

Системы оптического
распознавания форм

Системы распознавания
рукописного текста
(продукты фирмы **Apple**)



Обработка текстовой информации

Основные приемы редактирования

1. Откройте файл Фраза.

Теперь поднимитесь

2. Выделите эту фразу щелчком на левом поле напротив строки и скопируйте в буфер обмена, нажав на панели инструментов кнопку Копировать.

3. Вставьте буквосочетание из буфера обмена 7 раз с помощью кнопки Вставить.

4. Разделите буквосочетания пробелами и запятыми так, чтобы получившиеся фразы имели смысл.

Единицами текстовой информации являются

- Символ – наименьшая единица символьной информации
- Слово – набор символов, ограниченный пробелами
- Абзац – группа смежных строк файла, первая из которых может начинаться отступом влево по отношению к другим строкам абзаца, нулевым отступом или отступом вправо
- Строка – произвольная последовательность символов между левой и правой границами абзаца

Обработка текстовой информации

Основные приемы редактирования

1. Откройте файл Лишнее.doc.

Приставка, предлог, суффикс, окончание,
корень;

Треугольник, отрезок, длина, квадрат, круг ...

2. В каждой строке найдите лишнее слово.
Выделите его двойным щелчком и удалите,
нажав клавишу Delete на клавиатуре.

3. Сохраните файл с изменениями в папке Мои
документы под именем Нет_лишнего.

Текстовый редактор

- Текстовый редактор – это программа для работы с текстами.
- Текст состоит из букв, цифр, знаков препинания и специальных символов, которые можно ввести с помощью клавиатуры компьютера.
- Фотографии, чертежи, схемы и таблицы текстовый редактор не обрабатывает.

Двоичное кодирование текстовой информации

Начиная с 60-х годов, компьютеры все больше стали использовать для обработки текстовой информации и в настоящее время большая часть ПК в мире занято обработкой именно текстовой информации.

Традиционно для кодирования одного символа используется количество информации = 1 байту (1 байт = 8 битов).

Единицами текстовой информации являются

- Символ – наименьшая единица символьной информации
- Слово – набор символов, ограниченный пробелами
- Абзац – группа смежных строк файла, первая из которых может начинаться отступом влево по отношению к другим строкам абзаца, нулевым отступом или отступом вправо
- Строка – произвольная последовательность символов между левой и правой границами абзаца