

0011 0010 10 0

Представление текстов в памяти компьютера. Кодировочные таблицы.



Устройство компьютера

0011 0010 10



Имея компьютер, можно создавать
тексты, не тратя на это много времени и
бумагу.

Носителем текста становится память ПК.

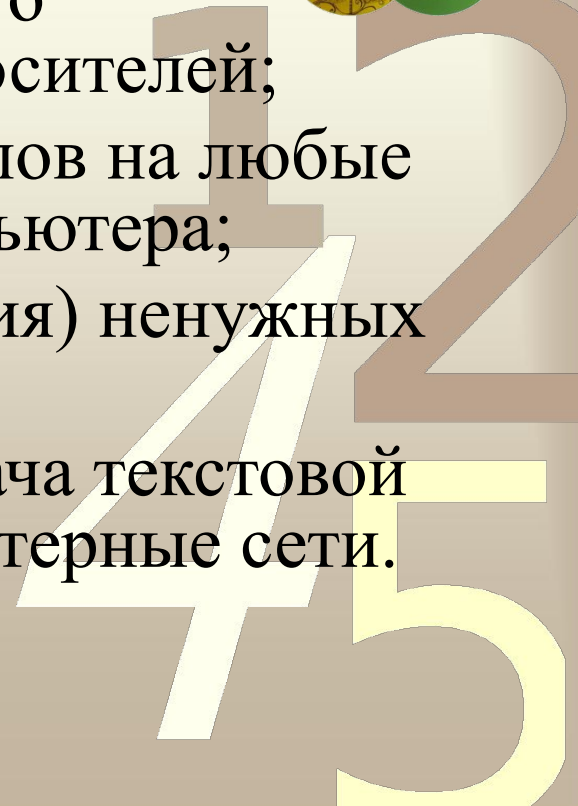
Текст на внешних носителях
сохраняется в виде файла.



Преимущества файлового хранения текстов по сравнению с бумагой:



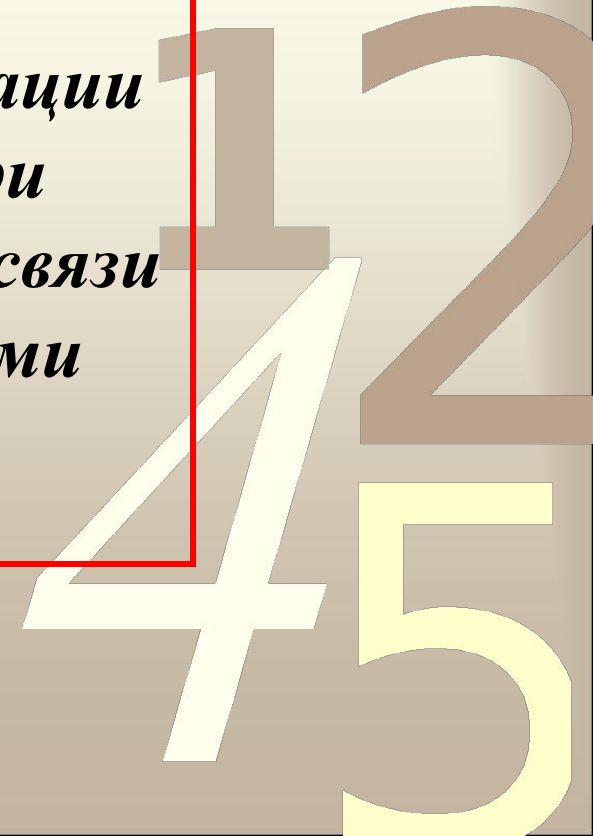
- возможность внесения изменений;
- компактность хранения текстовых документов;
- возможность многократного использования внешних носителей;
- легкость копирования файлов на любые носители с помощью компьютера;
- легкость удаления (стирания) ненужных текстовых файлов;
- оперативная связь – передача текстовой информации через компьютерные сети.



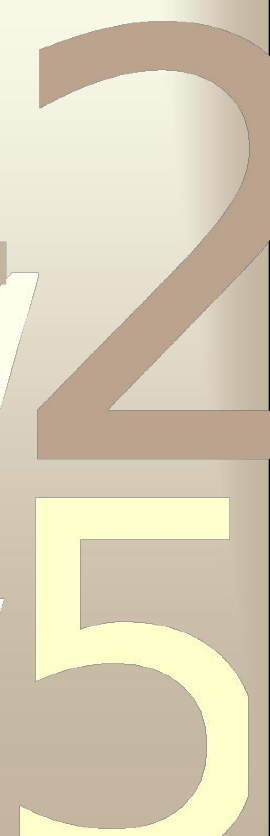
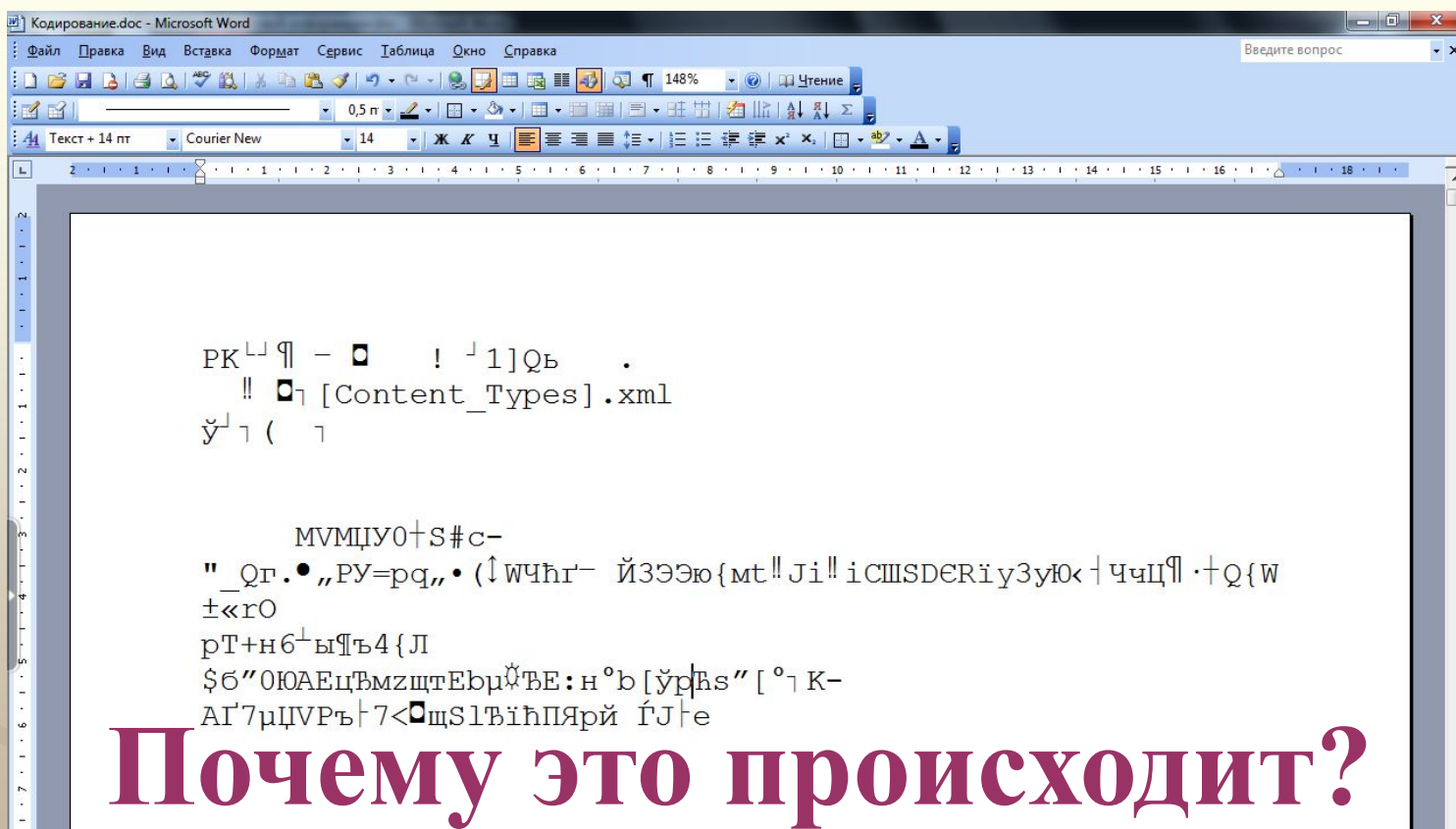
0011 0010 1010 1101 0001 0100 1011

Самое поразительное отличие
компьютерного текста от бумажного – это
создание в нем гипертекста.

*Гипертекст – это способ организации
текстовой информации, внутри
которой установлены смысловые связи
(гиперсвязи) между ее различными
фрагментами.*



Иногда бывает так, что текст, состоящий из букв русского алфавита, полученный с другого компьютера, невозможно прочитать - на экране монитора видна какая-то "абракадабра".



С точки зрения компьютера текст состоит из отдельных символов. К числу символов принадлежат не только буквы (заглавные или строчные, латинские или русские), но и цифры, знаки препинания, спецсимволы типа "=", "(", "&" и даже пробелы между словами.

Множество символов, с помощью которых записывается текст, называется *алфавитом*.

Число символов в алфавите – это его *мощность*.



Определение количества информации:

$$N = 2^b$$

где N – мощность алфавита (количество символов),
 b – количество бит (информационный вес символа).

Т.к. в алфавите 256 символов, тогда

$$256 = 2^8, \text{ т.е. вес 1 символа – 8 бит.}$$

Единице измерения 8 бит присвоили название **1 байт**:

$$1 \text{ байт} = 8 \text{ бит.}$$

Двоичный код каждого символа в компьютерном тексте занимает 1 байт памяти.



Пример.

0011 0010 1010 1101 0001 0100 1011

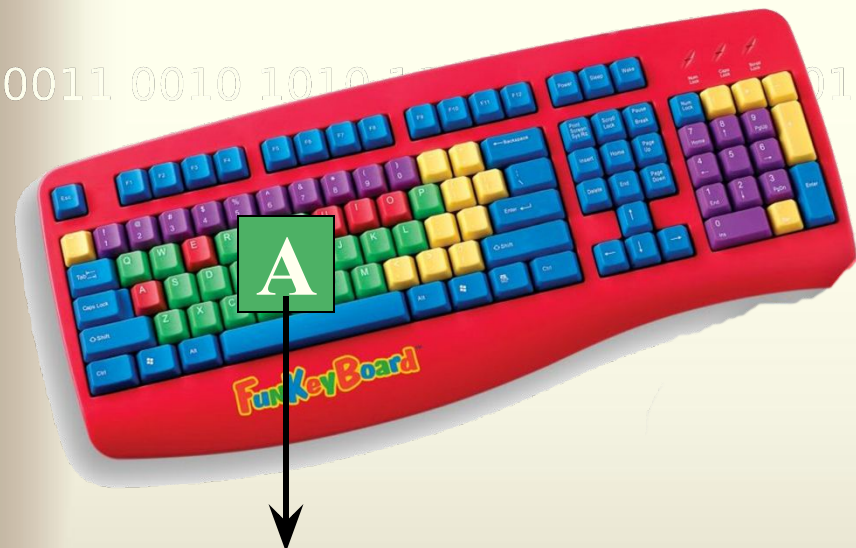
Сколько бит памяти компьютера
занимает слово микропроцессор?

Решение.

Микропроцессор – 14 символов,
значит занимает 14 байт

$14 \text{ байт} \times 8 = 112 \text{ бит}$



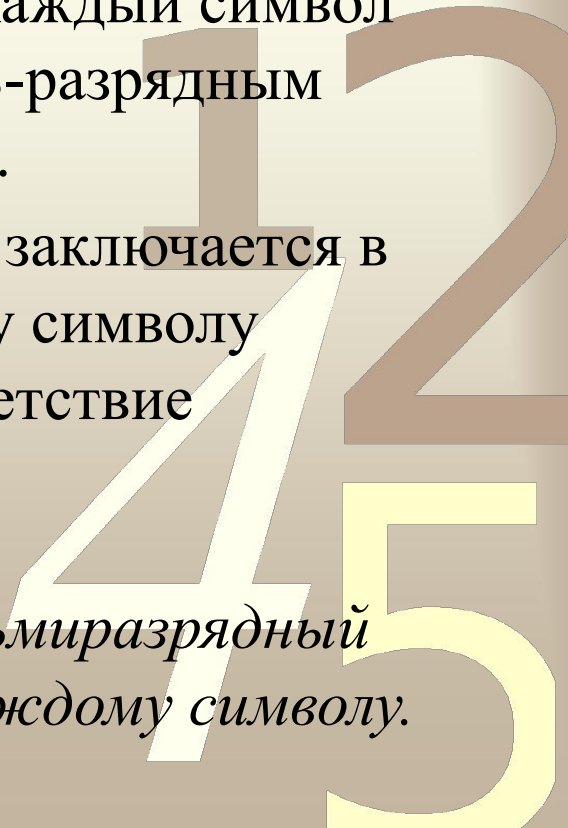


| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|

Тексты вводятся в память компьютера с помощью клавиатуры. На клавишах написаны привычные нам символы. В оперативную память они попадают в двоичном коде. Это значит, что каждый символ представляется 8-разрядным двоичным кодом.

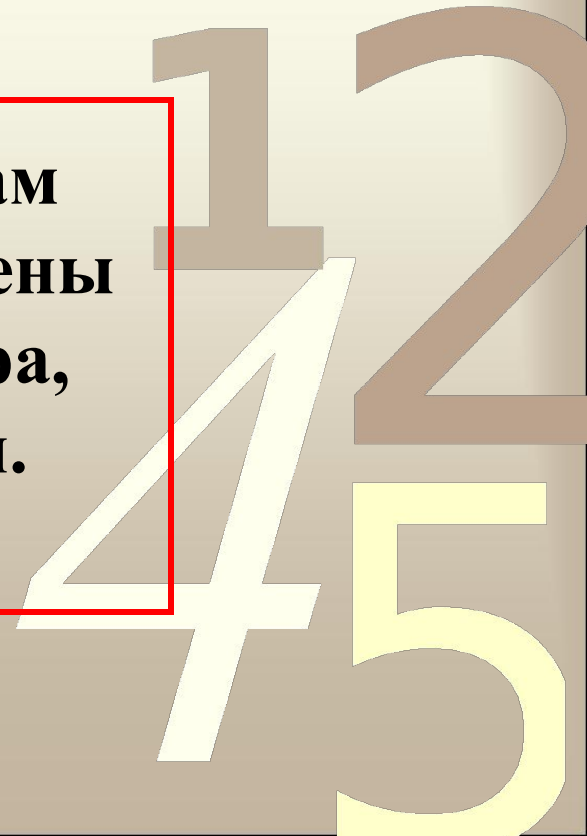
Кодирование заключается в том, что каждому символу ставится в соответствие уникальный код.

Теперь возникает вопрос, какой именно восьмиразрядный двоичный код поставить в соответствие каждому символу.



Все символы компьютерного алфавита пронумерованы от 0 до 255. Каждому номеру соответствует восьмиразрядный двоичный код от 00000000 до 11111111.

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера, называется таблицей кодировки.



0011 0010 1010 1101 0001 0100 1011

Для разных типов ЭВМ используются различные таблицы кодировки.

Международным стандартом для ПК стала таблица *ASCII*

(Американский стандартный код для информационного обмена).



Структура таблицы кодировки ASCII

Таблица кодов ASCII делится на две части.

| Порядковый номер | Код | Символ |
|------------------|------------------------|---|
| 0 - 31 | 00000000 – 00011111 | Управляющие символы. Процесс вывода текста на экран или печать, подача звукового сигнала, разметка текста. |
| 32 - 127 | 00100000 – 01111111 | Стандартная часть таблицы (английский). Сюда входят строчные и прописные буквы латинского алфавита, десятичные цифры, знаки препинания, всевозможные скобки, коммерческие и другие символы. Символ 32 - пробел, т.е. пустая позиция в тексте. |
| 128 - 255 | 10000000 – 11111111 | Вторая половина может иметь различные варианты. Кодовая страница используется для размещения национальных алфавитов. Для нас в этой это символы русского алфавита. |

*В таблице кодировки буквы
(прописные и строчные) располагаются
в алфавитном порядке, а цифры
упорядочены по возрастанию значений.
Такое соблюдение лексикографического
порядка в расположении символов
называется **принципом
последовательного кодирования
алфавита.***



Попробуем с помощью таблицы ASCII
представить, как будут выглядеть слова
в памяти компьютера.

| Слова | Память |
|-------|----------|
| file | 01100110 |
| | 01101001 |
| | 01101100 |
| | 01100101 |
| disk | 01100100 |
| | 01101001 |
| | 01110011 |
| | 01101011 |



0011 0010 1010 1101 0001 0100 1011

Пример.

0011 0010 1010 1101 0001 0100 1011

Сколько бит памяти компьютера
занимает выражение **жесткий диск**?

Решение.

Жесткий диск – 12 символов,

значит занимает 12 байт

$12 \text{ байт} \times 8 = 96 \text{ бит}$



С помощью таблицы ASCII закодировать и декодировать слова

0011 0010 1010 1101 0001 0100 1011

1) **Link**

01001100 01101001 01101110 01101011

2) **Класс**

11001010 11101011 11100000 11110001
11110001

3) 01010111 01101001 01101110 01100100
01101111 01110111 01110011

Windows



Домашнее задание

0011 0010 1010 1101 0001 0100 1011

§ 13; вопросы, №6 письменно

12
45