

Представление текста в памяти компьютера

7 класс

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

Почему это происходит?

Ответить на этот вопрос вам поможет информация, собранная в этой презентации.

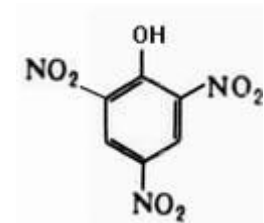
После просмотра презентации отправьте ответ на главный вопрос в виде текстового ответа.

Не забудь ответить и на вопросы к презентации.

С точки зрения компьютера **текст** - это любая последовательность символов из компьютерного алфавита.



Совсем не обязательно, чтобы это был текст на одном из естественных языков (русском, английском и др.). Это могут быть математические или химические формулы, номера телефонов, числовые таблицы. Самое главное, чтобы все используемые символы входили бы в компьютерный алфавит



Запиши в тетрадь, что такое компьютерный текст.



Представление текстовой информации

Для представления информации в компьютере используется алфавит мощностью 256 символов.

Чему равен информационный вес одного символа такого алфавита?

Вспомним из алфавитного подхода к измерению информации формулу, связывающую информационный вес символа алфавита и мощность алфавита:

$$2^i = N, \text{ где}$$

i – информационный вес одного символа

N – мощность алфавита

Дано:
 $N = 256$

$$2^i = N$$

$$2^i = 256$$

т.е. мы должны ответить на вопрос: два в какой степени даёт 256. Это 8 ($2^8 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 256$)

$$2^i = 2^8$$

$$i = 8 \text{ (бит)}$$

$i = ?$

Ответ: информационный вес одного символа алфавита, состоящего из 256 символов, равен 8 бит

[Щелкни здесь и ответь на вопросы в тетради](#)

Представление текстовой информации

Всякая информация представляется в памяти ЭВМ в двоичном виде, а это значит, что каждый символ представляется 8-разрядным двоичным кодом.

Например, буква «А» – имеет код 11000000 (8 цифр 0 и 1, т.е. 8 разрядов)

буква «Р» - имеет код 11010000 (8 цифр 0 и 1, т.е. 8 разрядов)

8 бит = 1 байту, поэтому **двоичный код каждого символа в компьютерном тексте занимает 1 байт** памяти.

Запиши ответы на вопросы в тетради:

- Как представлен символ в памяти компьютера?
- Сколько «весит» один символ компьютерного текста в битах? Почему?

Представление текстовой информации

Теперь возникает вопрос, какой именно восьмиразрядный двоичный код поставить в соответствие каждому символу. Понятно, что это дело условное, можно придумать множество способов кодировки.

Все символы компьютерного алфавита пронумерованы от 0 до 255. Каждому номеру соответствует восьмиразрядный двоичный код от 00000000 до 11111111. Этот код просто порядковый номер символа в двоичной системе счисления.

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

Запиши ответ на вопрос в тетради:

- Что такое таблица кодировки?

Представление текстовой информации

Международным стандартом для ПК стала таблица **ASCII** (читается аски) (Американский стандартный код для информационного обмена). На практике можно встретиться и с другой таблицей – **КОИ-8** (Код обмена информацией), которая используется в глобальных компьютерных сетях.

Откройте учебник и рассмотрите таблицу кодов ASCII

Таблица кодов ASCII

Первая половина

Таблица кодов ASCII делится на две части.

Международным стандартом является лишь первая половина таблицы, т.е. символы с номерами от 0 (00000000), до 127 (01111111). Сюда входят строчные и прописные буквы латинского алфавита, десятичные цифры, знаки препинания, всевозможные скобки, коммерческие и другие символы.

Символы с номерами от 0 до 31 принято называть управляющими. Их функция – управление процессом вывода текста на экран или печать, подача звукового сигнала, разметка текста и т.п.

Символ 32 - пробел, т.е. пустая позиция в тексте. Все остальные отражаются определенными знаками.

Запиши ответы на вопросы в тетради:

- Что описано в первой половине таблицы кодов ASCII
- Какой номер имеет – пробел?

Таблица кодов ASCII

Первая половина таблицы

Обратите внимание:

- в таблице кодировки буквы (прописные и строчные) располагаются в алфавитном порядке
- цифры упорядочены по возрастанию значений.

Такое соблюдение лексикографического порядка в расположении символов называется **принципом последовательного кодирования алфавита**.

Запиши ответ на вопрос в тетради:

- в чем заключается принцип последовательного кодирования алфавита

Таблица кодов ASCII

Вторая половина таблицы

Вторая половина кодовой таблицы ASCII, называемая кодовой страницей (128 кодов, начиная с 10000000 и кончая 11111111), может иметь различные варианты, каждый вариант имеет свой номер.

Кодовая страница в первую очередь используется для размещения национальных алфавитов, отличных от латинского. В русских национальных кодировках в этой части таблицы размещаются символы русского алфавита. Для букв русского алфавита также соблюдается принцип последовательного кодирования.

Запиши ответы на вопросы в тетради:

- Что описано во второй половине таблицы кодов ASCII. Как она называется?

Кодирование текста

В настоящее время существуют пять различных кодировок русского алфавита :

- [КОИ8-Р](#);
- [Windows](#);
- [MS-DOS](#);
- [Macintosh](#);
- [ISO](#).

Из-за этого часто возникают проблемы с переносом русского текста с одного компьютера на другой, из одной программной системы в другую.