



Кодирование текстовой информации

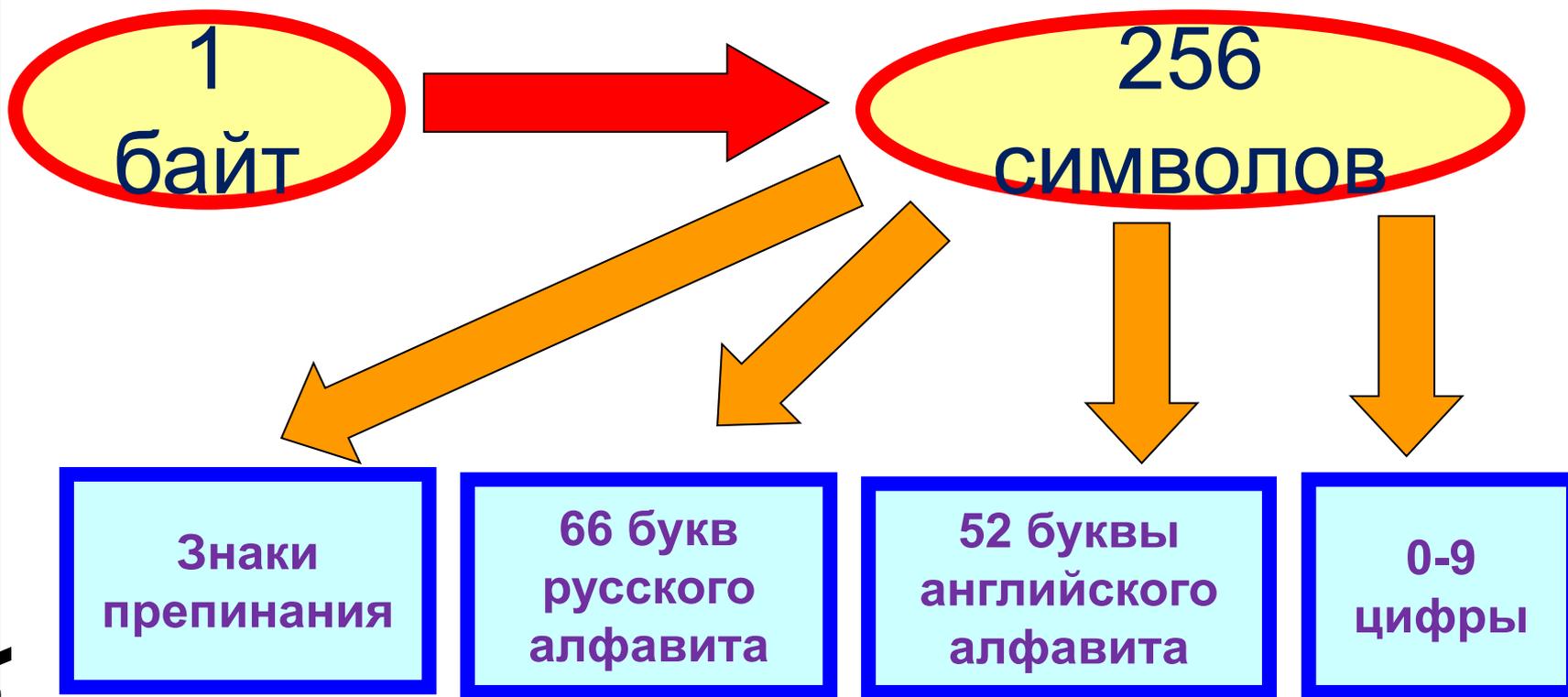
Автор: Сергеева Светлана Александровна,
учитель информатики МБОУ СОШ № 27,
г. Дзержинска, Нижегородской обл.

Текстовая информация – это информация, выраженная с помощью естественных или формальных языков в письменной форме.



Двоичное кодирование текстовой информации

Для кодирования **1 символа** используется
1 байт информации.



**При двоичном кодировании
текстовой информации**
каждому символу ставится в
соответствие своя уникальная
последовательность из восьми
нулей и единиц, свой
уникальный код
от 00000000 до 11111111
(десятичный код от 0 до 255).



Присвоение символу конкретного двоичного кода – это вопрос соглашения, которое фиксируется в кодовой таблице.

Кодовая таблица – таблица, в которой устанавливается соответствие между числовыми кодами и символами.



0 – 32 функциональные клавиши и операции: перевод строки, ввод пробела и т.д.

33 – 127 интернациональные: соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания.



ГДПНОВАРБСИРЕ.

ОПНЬС ХГБХВЕМХЪ, МН, Й ЯНФЮКЕМХЧ, МЕ ГИЮЧ БЮБЕЦН Х.Н., РНКЭЙН
ТМЛХКХЧ - ЧДХЮЮ (ЯСДЪ ОН e-mail). ЕЯКХ НЪХАЮЧЯЭ, РИ ОНОЛЮВЗРЕ. ЦРН
ОХЪЕР РНР ЛНКНДНИ ВЕКНБЕЙ, Я ЙНРНШЛ БШ ПЮГЦНБЮПХБЮКХ МЮ "БАНКЕ-
2000" В ПЕДЮЙКХХ "1 ЯЕМРЪАПЪ". Я БЮБЕИ ЯРНПНШ ОПНГБСВЮКН
ОПЕДКНФЕИХЕ ОНГМЮЙНЛХРЭ ЛЕМЪ Я ВЕКНБЕЙНЛ, ГЮМХЛЮЧЫХЛЯЪ БНОПНЮЛХ
ЙНЛОЗЧРЕПНЦН НАСВЕМХЪ В МЮВЮКЭМШУ ЙКЮЯЧЮУ.

ЙНПНРЙН Н РНЛ, БЕЛ ЛШ ГЮМХЛЮЕЛЯЪ. МЮБЕ ХЯЯКЕДНБЮМХЕ МНЯХР
МЕЯЙНКЭЙН ОЯХУНКНЦХВЕЯЙХИ НРРЕИМЪЙ. ЙНМЮПЕРИИ - ЦРН ДХТТЕПЕМЖХЮКХЪ
ДЕРЕИ ОН ЯРЕОЕИХ БНГЮЯРЮМХЪ ВГЮХЛНДЕИЯРБХЪ Я БШВХЯКХРЕКЭИИ
РЕУМХЙИИ Х ДОКЭМЕИЬЮЪ ХУ ОЯХУНКНЦХВЕЯЙЮЪ ДХЮЦМНЯРХЮЮ:

1. Minimum - МЕР ЙНИРЮЙРЮ Я ЙНЛОЗЧРЕПНЛ;
2. йнлозчреп днло (хкх бне банкш);
3. йнлозчреп в банке;
4. Maximum - йнлозчреп в банке х днло (бне банкш).



128 – 255 национальные (кириллица),
т.е. в национальных кодировках
одному и тому же коду соответствуют
различные символы.



Кодировки знаков

Двоичный код	Десятичный код	КОИ8	CP1251	CP866	Mac	ISO
0000 0000	0					
.....						
0000 1000	8	Удаление последнего символа (клавиша Backspace)				
.....						
0000 1101	13	Перевод строки (клавиша Enter)				
.....						
0010 0000	32	Пробел				
0010 0001	33	!				
.....						
0101 1010	90	z				
.....						
0111 1111	127	□				
.....	128	-	Ъ	А	А	К
.....						
1100 0010	194	Б	В	-	-	Т
.....						
1100 1100	204	Л	М	:	:	Ь
.....						
1101 1101	221	Щ	Э	-	Ё	Н
.....						
1111 1111	255	ь	я	Нераз. пробел	Нераз. пробел	п

Хронологически одним из первых стандартов кодирования русских букв на компьютерах был код КОИ – 8 («Код обмена информационный – 8 битный»). Эта кодировка применяется в компьютерах с операционной системой UNIX.





Наиболее распространенная кодировка – это стандартная кириллическая кодировка **Microsoft Windows**, обозначаемая сокращением **CP1251** («CP» означает «Code Page»). Все Windows – приложения, работающие с русским языком, поддерживают эту кодировку.





Для работы в среде операционной системы **MS-DOS** используется «альтернативная» кодировка, в терминологии фирмы Microsoft – кодировка **CP 866**.



Фирма Apple разработала для компьютеров **Macintosh** свою собственную кодировку русских букв (**Mac**).





Международная организация по стандартизации (International Standards Organization, ISO) утвердила в качестве стандарта для русского языка еще одну кодировку под названием **ISO 8859 – 5**.



Стандарты кодировок:

1. КОИ-8 - UNIX
2. CP1251 («CP» означает «Code Page») - Microsoft Windows
3. CP 866 - MS-DOS
4. Mac - Macintosh
5. ISO 8859 – 5



В последнее время появился новый международный стандарт **Unicode**, который отводит на каждый символ не один байт, а два, и поэтому с его помощью можно закодировать 65 536 различных символов. Эту кодировку поддерживает платформа Microsoft Windows.



Решение задач

N – количество символов в алфавите
(мощность алфавита)

i – информационный «вес» одного
символа

$$N = 2^i$$

k – количество символов в тексте

$$I_{m.} = k \cdot i$$



1. Сколько килобайт составляет сообщение, содержащее 12288 бит?

2. Каков информационный объем слова ПРОГРАММИРОВАНИЕ:

А) в 16-битной кодировке;

Б) в 8-битной кодировке.

3. Книга, набранная с помощью компьютера, содержит 150 страниц, на каждой странице – 40 строк, в каждой строке – 60 символов. Каков объем информации в книге?



Домашнее задание

п. 2.1, № 2.1, 2.2 (стр. 52) письменно

