

Кодирование текстовой информации

Двоичное кодирование текстовой информации в компьютере

Текстовая информация – это информация, выраженная с помощью естественных и формальных языков в письменной форме (прописные и строчные буквы, цифры, знаки, математические символы).

Для представления текстовой информации достаточно 256 символов. Можно вычислить количество информации, чтобы закодировать каждый знак:

$$N=2^I \quad 256=2^8 \quad 2^8=2^I \quad I=8 \text{ битов}$$

Тексты вводятся в память компьютера с помощью клавиатуры. На клавишах написаны привычные нам буквы, цифры, знаки препинания и другие 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 десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111. Таким образом, человек различает символы по их начертанию, а компьютер - по их коду.

СИМВОЛ	КОД	СИМВОЛ	КОД	СИМВОЛ	КОД
А	00000	Л	01011	Ц	10110
Б	00001	М	01100	Ч	10111
В	00010	Н	01101	Ш	11000
Г	00011	О	01110	Щ	11001
Д	00100	П	01111	Ъ	11010
Е	00101	Р	10000	Ы	11011
Ж	00110	С	10001	Ь	11100
З	00111	Т	10010	Э	11101
И	01000	У	10011	Ю	11110
Й	01001	Ф	10100	Я	11111
К	01010	Х	10101		

Символ	Двоичный код	Десятичный код	Символ	Двоичный код	Десятичный код
A	01000001	65	N	01001110	78
B	01000010	66	O	01001111	79
C	01000011	67	P	01010000	80
D	01000100	68	Q	01010001	81
E	01000101	69	R	01010010	82
F	01000110	70	S	01010011	83
G	01000111	71	T	01010100	84
H	01001000	72	U	01010101	85
I	01001001	73	V	01010110	86
J	01001010	74	W	01010111	87
K	01001011	75	X	01011000	88
L	01001100	76	Y	01011001	89
M	01001101	77	Z	01011010	90

Структура таблицы кодировки ASCII

Порядковый номер	Код	Символ
0-31	00000000 - 00011111	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.
31-127	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. Все остальные отражаются определенными знаками.
128-255	10000000 - 11111111	Альтернативная часть таблицы (русская). 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.

Различные кодировки знаков

Символ	Windows	MS-DOS	КОИ-8	Mac	ISO	Unicode
А	192	128	225	128	176	1040
В	194	130	247	130	178	1042
М	204	140	237	140	188	1052
Э	221	157	252	157	205	1069
я	255	239	241	223	239	1103

Спасибо за внимание