

Презентация
на тему:
Кодирование текстовой
Информации

ГРУППА 311



Понятие текстовой информации

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

К текстовой информации относятся:

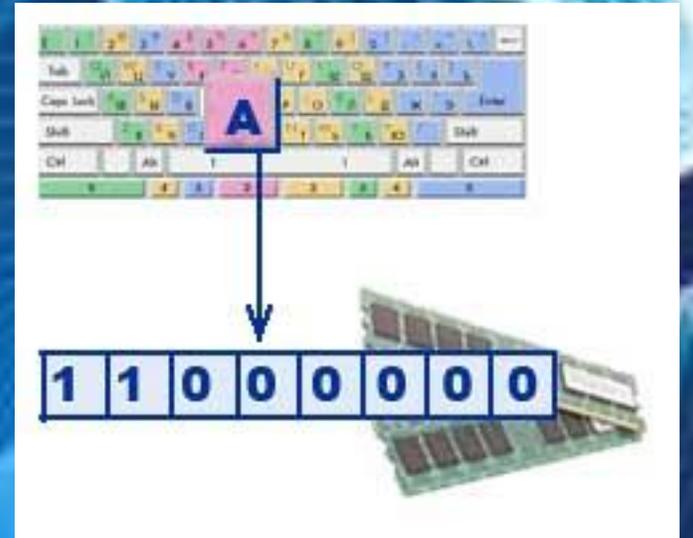
- Буквы русского алфавита
- Буквы латинского алфавита
- Цифры
- Знаки
- Математические символы

Представление текстовой информации

Для кодирования текстовой информации достаточно 256 различных знаков!

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.



Человек различает знаки по их начертанию, а компьютер – по их двоичным кодам. При вводе в компьютер текстовой информации происходит её двоичное кодирование, изображение знака преобразуется в его двоичный код. Пользователь нажимает на клавиатуре клавишу со знаком, и в компьютер поступает определенная последовательность из восьми электрических импульсов (двоичный код знака). Код знака хранится в оперативной памяти компьютера.

В процессе вывода знака на экран компьютера производится обратное кодирование, т. е. преобразование двоичного кода знака в его изображение.



Преобразование текстовой информации

Ввод символа с клавиатуры

A

f

8

@



Оперативная память

11000000

01100110

00001000

01000000

Преобразование текстовой информации

Оперативная память

00111100

01101101

11011001

10111001



Вывод на экран

V

m

Щ

No

Таблица кодировки

Таблица кодировки – это таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды)

Международным стандартом для ПК считается таблица **ASCII – American Standard Code for Information Interchange** (Американский стандартный код для информационного обмена)



Кодовая таблица

Для представления символов и соответствующих им кодов используется *кодовая таблица*.

В качестве стандарта во всем мире принята таблица ASCII (American Standard Code for Information Interchange – Американский стандартный код для обмена информацией). Условно таблица разделена на части:

- от 0 до 32 коды соответствуют операциям;
- с 33 по 127 соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания;
- со 128 по 255 являются национальными.

Современные кодовые таблицы

В настоящее время существует несколько различных кодовых таблиц для русских

букв:

Windows

Mac

ISO

Unicode

S

MS-DOS

KOI-8

Тексты, созданные в одной кодировке, не будут правильно отображаться в другой!



Современные кодовые таблицы

Mac

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

КОИ-8

128	144	160	176	192	208	224	240
129	145	161	177	193	209	225	241
130	146	162	178	194	210	226	242
131	147	163	179	195	211	227	243
132	148	164	180	196	212	228	244
133	149	165	181	197	213	229	245
134	150	166	182	198	214	230	246
135	151	167	183	199	215	231	247
136	152	168	184	200	216	232	248
137	153	169	185	201	217	233	249
138	154	170	186	202	218	234	250
139	155	171	187	203	219	235	251
140	156	172	188	204	220	236	252
				205	221	237	253
				206	222	238	254
				207	223	239	255

MS-DOS

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

128	А	144	Р	160	а	176	Ё	192	Г	208	А	224	р	240	Е
129	Б	145	С	161	б	177	■	193	Г	209	Т	225	с	241	ё
130	В	146	Т	162	в	178	■	194	Г	210	Т	226	т	242	ё
131	Г	147	У	163	г	179	+	195	Г	211	Т	227	у	243	ё
132	Д	148	Ф	164	д	180	+	196	Г	212	Т	228	ф	244	ё
133	Е	149	Х	165	е	181	+	197	Г	213	Т	229	х	245	ё
134	Ж	150	Ц	166	ж	182	+	198	Г	214	Т	230	ц	246	ё
135	З	151	Ч	167	з	183	+	199	Г	215	Т	231	ч	247	ё
136	И	152	Ш	168	и	184	+	200	Г	216	Т	232	ш	248	ё
137	Й	153	Щ	169	й	185	+	201	Г	217	Т	233	щ	249	ё
138	К	154	Ъ	170	к	186	+	202	Г	218	Т	234	ъ	250	ё
139	Л	155	Ы	171	л	187	+	203	Г	219	Т	235	ы	251	ё
140	М	156	Ь	172	м	188	+	204	Г	220	Т	236	ь	252	ё
141	Н	157	Э	173	н	189	+	205	Г	221	Т	237	э	253	ё
142	О	158	Ю	174	о	190	+	206	Г	222	Т	238	ю	254	ё
143	П	159	Я	175	п	191	+	207	Г	223	Т	239	я	255	ё

Windows

ISO

163	Г	179	Г	195	У	211	г	227	у	243	г
164	Є	180	Д	196	Ф	212	д	228	ф	244	є
165	Ѕ	181	Е	197	Х	213	е	229	х	245	ѕ
166	І	182	Ж	198	Ц	214	ж	230	ц	246	і
167	І	183	З	199	Ч	215	з	231	ч	247	і
168	Ј	184	И	200	Ш	216	и	232	ш	248	ј
169	Љ	185	Р	201	Щ	217	р	233	щ	249	љ
170	Њ	186	К	202	Ъ	218	к	234	ъ	250	њ
171	Ћ	187	Л	203	Ы	219	л	235	ы	251	ћ
172	К	188	М	204	Ь	220	м	236	ь	252	ќ
173	-	189	Н	205	Э	221	н	237	э	253	ѓ
174	Ў	190	О	206	Ю	222	о	238	ю	254	у
175	U	191	П	207	Я	223	п	239	я	255	џ

В ISO не определены

Кодовая таблица Unicode

В 1991 году был предложен новый международный стандарт кодирования – **Unicode**, который отводит на каждый символ по 2 байта (16 бит)

$$N = 2^{16} = 65\,536$$

取 撚 柏 嬰 滂 嬰

אָבִינִי שְׁבַשְׁמִים תְּקִידָה

α β γ δ ε ζ η θ

اتن بمكايته ان بعض ملوك جبل القبق

Такое количество кодовых комбинаций позволяет закодировать знаки языков практически всех алфавитов мира!

Спасибо за внимание!

