## *Двоичное кодирование текстовой информации. Различные кодировки кириллицы*

Как отмечалось ранее, для представления информации в памяти ЭВМ используется двоичный способ кодирования.

Если каждому символу какого-либо алфавита сопоставить определённое целое число, то с помощью двоичного кода можно кодировать и текстовую информацию. Для хранения двоичного кода одного символа может быть выделен 1 байт = 8 битам. Учитывая, что каждый бит принимает значение 0 или 1, количество их возможных сочетаний в байте равно $2^8 = 256$. значит, с помощью 1 байта можно получить 256 разных двоичных кодовых комбинаций и отобразить с их помощью 256 различных символов. Такое количество символов вполне достаточно для представления текстовой информации, включая прописные и строчные буквы русского и латинского алфавитов, цифры, знаки, псевдографические символы и т. д. 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. Важно, что присвоение символу конкретного кода – это вопрос соглашения, которое фиксируется в кодовой таблице. Кодирование текстовой информации с помощью байтов опирается на несколько различных стандартов, но первоосновной для всех стал ASCII (**A***merican* **S***tandard* **C***ode for* **I***nformation* **I***nterchange*), разработанный в США в национальном институте ANSI (**A***merican* **N***ational* **S***tandards* **I***nstitute*). В системе ASCII закреплены две таблицы кодирования – базовая и расширенная. Базовая таблица закрепляет значения кодов от 1 до 127, а расширенная относится к символам с номерами 128 до 255. первые 33 кода (с 0-го по 32-й) соответствуют не символам, а операциям (перевод строки, ввод пробела и т.д.). Коды с 33-го по 127-й являются интернациональными и соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания. Коды с 128-го по 255-й являются национальными, т.е. в национальных кодировках одному и тому же коду соответствуют различные символы.

В языках, использующих кириллический алфавит, в том числе русском, пришлось полностью менять вторую половину таблицы ASCII, приспосабливая ее под кириллический алфавит. В частности, для представления символов кириллицы используется так называемая «альтернативная кодировка».

В настоящее время существует несколько различных кодовых таблиц для русских букв (КОИ-8, СР-1251, СР-866, Mac, ISO), поэтому тексты, созданные в одной кодировке, могут неправильно отображаться в другой.

После появления ОС Windows от фирмы Microsoft выяснилось, что альтернативная кодировка по некоторым причинам для неё не подходит. Передвинув русские буквы в таблице на место символов псевдографики (появилась возможность – ведь псевдографика в Windows не требуется), получили кодировку Windows 1251 (Win-1251).

Но компьютерные технологии постоянно совершенствуются, и в настоящее время всё большее число программ начинает поддерживать шестнадцатибитовый стандарт Unicode, который позволяет кодировать практически все языки и диалекты жителей Земли в силу того, что кодировка включает в себя 65 536 различных двоичных кодов.

Международная организация по стандартизации (**I***nternational* **O***rganization for* **S***tandardization*) разработала свой код, способный соперничать с Unicode. Здесь для кодировки символов используется комбинация из 32 бит.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | | ☺ 1 | ☻ 2 | ♥ 3 | ♦ 4 | ♣ 5 | ♠ 6 | • 7 | 8 | o 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| **1** | ➢ 16 | ➢ 17 | 18 | 19 | 20 | 21 | - 22 | 23 | ↑ 24 | ↓ 25 | → 26 | ← 27 | ↔ 28 | 29 | 30 | 31 |
| **2** | 32 | ! 33 | " 34 | # 35 | $ 36 | % 37 | & 38 | ' 39 | ( 40 | ) 41 | * 42 | + 43 | , 44 | - 45 | . 46 | / 47 |
| **3** | 0 48 | 1 49 | 2 50 | 3 51 | 4 52 | 5 53 | 6 54 | 7 55 | 8 56 | 9 57 | : 58 | ; 59 | < 60 | = 61 | > 62 | ? 63 |
| **4** | @ 64 | A 65 | B 66 | C 67 | D 68 | E 69 | F 70 | G 71 | H 72 | I 73 | J 74 | K 75 | L 76 | M 77 | N 78 | O 79 |
| **5** | P 80 | Q 81 | R 82 | S 83 | T 84 | U 85 | V 86 | W 87 | X 88 | Y 89 | Z 90 | [ 91 | \ 92 | ] 93 | ^ 94 | _ 95 |
| **6** | ` 96 | a 97 | b 98 | c 99 | d 100 | e 101 | f 102 | g 103 | h 104 | i 105 | j 106 | k 107 | l 108 | m 109 | n 110 | o 111 |
| **7** | p 112 | q 113 | r 114 | s 115 | t 116 | u 117 | v 118 | w 119 | x 120 | y 121 | z 122 | { 123 | \| 124 | }] 125 | ~ 126 | □ 127 |
| **8** | А 128 | Б 129 | В 130 | Г 131 | Д 132 | Е 133 | Ж 134 | З 135 | И 136 | Й 137 | К 138 | Л 139 | М 140 | Н 141 | О 142 | П 143 |
| **9** | Р 144 | С 145 | Т 146 | У 147 | Ф 148 | Х 149 | Ц 150 | Ч 151 | Ш 152 | Щ 153 | Ъ 154 | Ы 155 | Ь 156 | Э 157 | Ю 158 | Я 159 |
| **A** | а 160 | б 161 | в 162 | г 163 | д 164 | е 165 | ж 166 | з 167 | и 168 | й 169 | к 170 | л 171 | м 172 | н 173 | о 174 | п 175 |
| **B** | р 176 | с 177 | т 178 | у 179 | ф 180 | х 181 | ц 182 | ч 183 | ш 184 | щ 185 | ъ 186 | ы 187 | ь 188 | э 189 | ю 190 | я 191 |
| **C** | └ 192 | ┴ 193 | ┬ 194 | ├ 195 | ─ 196 | ┼ 197 | ╞ 198 | ╟ 199 | ╚ 200 | ╔ 201 | ╩ 202 | ╦ 203 | ╠ 204 | ═ 205 | ╬ 206 | ╧ 207 |
| **D** | ╨ 208 | ╤ 209 | ╥ 210 | ╙ 211 | ╘ 212 | ╒ 213 | ╓ 214 | ╫ 215 | 216 | ┘ 217 | ┌ 218 | ▉ 219 | ▄ 220 | ▌ 221 | ▐ 222 | ▀ 223 |
| **E** | р 224 | с 225 | т 226 | у 227 | ф 228 | х 229 | ц 230 | ч 231 | ш 232 | щ 233 | ъ 234 | ы 235 | ь 236 | э 237 | ю 238 | я 239 |
| **F** | Ё 240 | ё 241 | >= 242 | <= 243 | ⌠ 244 | ⌡ 245 | 246 | ≈ 247 | ° 248 | ∙ 249 | • 250 | √ 251 | ∞ 252 | ` 253 | ■ 254 | □ 255 |

Рис. 1. Таблица альтернативной кодировки.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | | 1 | 2 | © 3 | Ё 4 | § 5 | є 6 | · 7 | 8 | o 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| **1** | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| **2** | 32 | ! 33 | " 34 | # 35 | $ 36 | % 37 | & 38 | 39 | ( 40 | ) 41 | * 42 | + 43 | , 44 | . 45 | 46 | / 47 |
| **3** | 0 48 | 1 49 | 2 50 | 3 51 | 4 52 | 5 53 | 6 54 | 7 55 | 8 56 | 9 57 | : 58 | ; 59 | < 60 | = 61 | > 62 | ? 63 |
| **4** | @ 64 | A 65 | B 66 | C 67 | D 68 | E 69 | F 70 | G 71 | H 72 | I 73 | J 74 | K 75 | L 76 | M 77 | N 78 | O 79 |
| **5** | P 80 | Q 81 | R 82 | S 83 | T 84 | U 85 | V 86 | W 87 | X 88 | Y 89 | Z 90 | [ 91 | \ 92 | ] 93 | ^ 94 | _ 95 |
| **6** | ` 96 | a 97 | b 98 | c 99 | d 100 | e 101 | f 102 | g 103 | h 104 | i 105 | j 106 | k 107 | l 108 | m 109 | n 110 | o 111 |
| **7** | p 112 | q 113 | r 114 | s 115 | t 116 | u 117 | v 118 | w 119 | x 120 | y 121 | z 122 | { 123 | \| 124 | } ] 125 | ^ 126 | □ 127 |
| **8** | Ъ 128 | ┌ 129 | ‚ 130 | 131 | „ 132 | … 133 | † 134 | ‡ 135 | 136 | ‰ 137 | Љ 138 | ‹ 139 | Њ 140 | Ќ 141 | Ћ 142 | Џ 143 |
| **9** | ђ 144 | ‘ 145 | ’ 146 | “ 147 | ” 148 | 149 | – 150 | — 151 | 152 | ™ 153 | љ 154 | › 155 | њ 156 | ќ 157 | ћ 158 | џ 159 |
| **A** | 160 | Ў 161 | ў 162 | Ј 163 | ¤ 164 | Ґ 165 | ¦ 166 | § 167 | Ё 168 | © 169 | ª 170 | « 171 | 172 |  173 | ® 174 | Ї 175 |
| **B** | ° 176 | ± 177 | І 178 | і 179 | ґ 180 | µ 181 | 182 | 183 | ё 184 | № 185 | є 186 | 187 | Ѕ 188 | ѕ 189 | ї 190 | • 191 |
| **C** | А 192 | Б 193 | В 194 | Г 195 | Д 196 | Е 197 | Ж 198 | З 199 | И 200 | Й 201 | К 202 | Л 203 | М 204 | Н 205 | О 206 | П 207 |
| **D** | Р 208 | С 209 | Т 210 | У 211 | Ф 212 | Х 213 | Ц 214 | Ч 215 | Ш 216 | Щ 217 | Ъ 218 | Ы 219 | Ь 220 | Э 221 | Ю 222 | Я 223 |
| **E** | а 224 | б 225 | в 226 | г 227 | д 228 | е 229 | ж 230 | з 231 | и 232 | й 233 | к 234 | л 235 | м 236 | н 237 | о 238 | п 239 |
| **F** | р 240 | с 241 | т 242 | у 243 | ф 244 | х 245 | ц 246 | ч 247 | ш 248 | щ 249 | ъ 250 | ы 251 | ь 252 | э 253 | ю 254 | я 255 |

Таблица кодировки кириллицы в Windows (CP-1251).