

**КОДИРОВАНИЕ**

**ТЕКСТОВОЙ ИНФОРМАЦИИ**



# Текст на внешних носителях сохраняется в виде файла.

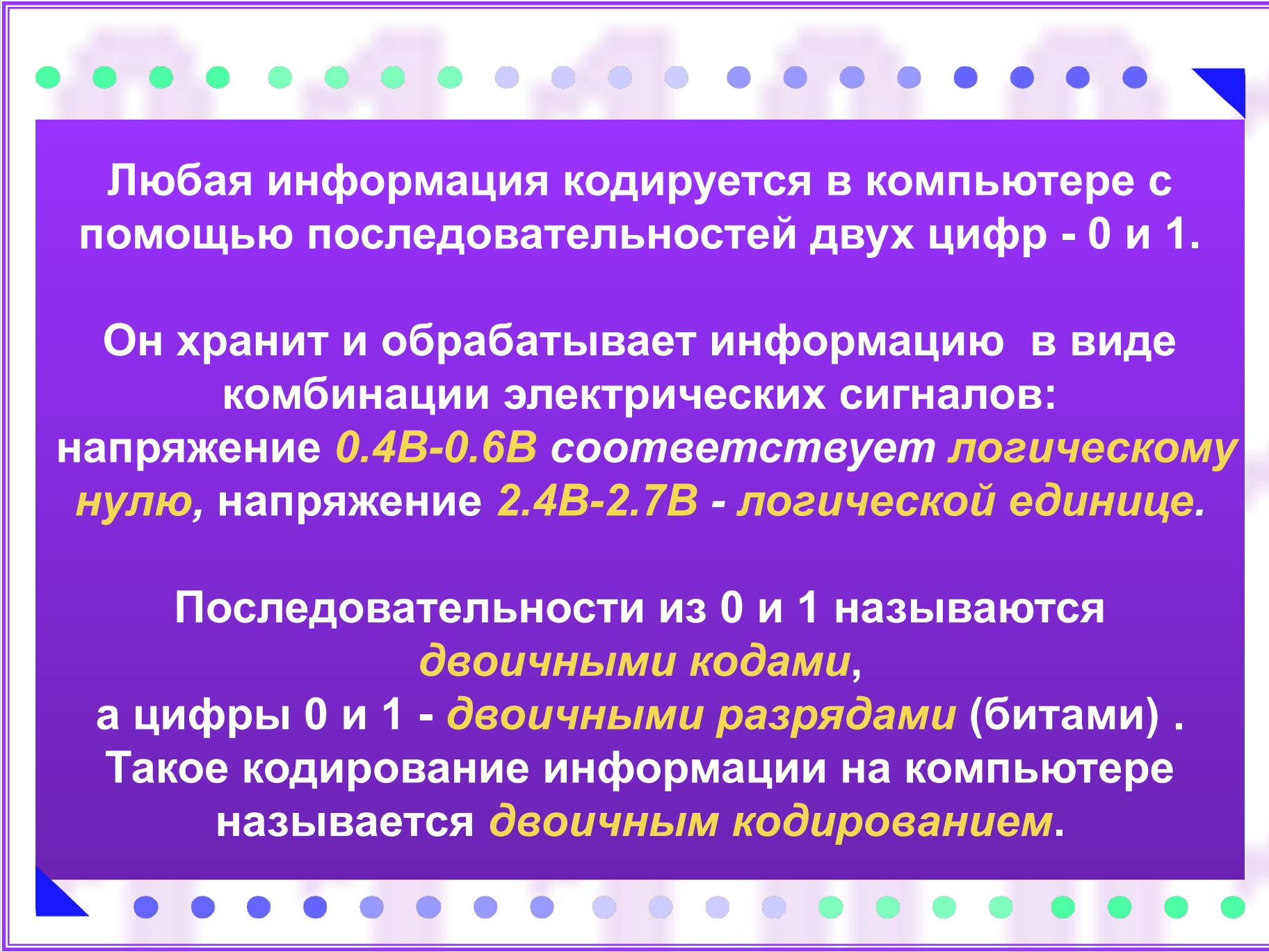
## Преимущества компьютерного документа:

- Компактное размещение
- Легко удалить
- Легко размножить
- Можно быстро переслать на большое расстояние

## Недостаток:

- Прочитать файл можно только с помощью компьютера





Любая информация кодируется в компьютере с помощью последовательностей двух цифр - 0 и 1.

Он хранит и обрабатывает информацию в виде комбинации электрических сигналов:  
напряжение *0.4В-0.6В* соответствует *логическому нулю*, напряжение *2.4В-2.7В* - *логической единице*.

Последовательности из 0 и 1 называются *двоичными кодами*,  
а цифры 0 и 1 - *двоичными разрядами* (битами) .  
Такое кодирование информации на компьютере называется *двоичным кодированием*.

# Обработкой текстовой информации на компьютере начали заниматься с 60 – х годов 20 века.



Первоначально в ЭВМ использовался лишь латинский алфавит (26 букв). Для обозначения каждого символа хватило бы пяти разрядов (битов). Но в тексте есть знаки препинания, десятичные цифры и др. Поэтому в первых англоязычных компьютерах байт - машинный слог - включал шесть битов. Затем семь - не только для того, чтобы отличать большие буквы от малых, но и для увеличения числа кодов управления принтерами, сигнальными лампочками и прочим оборудованием.

В 1964 году появились мощные ЭВМ IBM-360, в которых окончательно *байт* стал равен *восемь битам*, так как для представления текстовой информации в компьютере начал использоваться алфавит мощностью 256 СИМВОЛОВ.

$$2^8 = 256$$

# Международная кодировка ASCII

|          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| sp<br>32 | !<br>33  | "<br>34  | #<br>35  | \$<br>36 | %<br>37  | &<br>38  | '<br>39  | (<br>40  | )<br>41  | *<br>42  | +<br>43  | ,<br>44  | -<br>45  | .<br>46  | /<br>47  |
| 0<br>48  | 1<br>49  | 2<br>50  | 3<br>51  | 4<br>52  | 5<br>53  | 6<br>54  | 7<br>55  | 8<br>56  | 9<br>57  | :<br>58  | ;<br>59  | <<br>60  | =<br>61  | ><br>62  | ?<br>63  |
| @<br>64  | A<br>65  | B<br>66  | C<br>67  | D<br>68  | E<br>69  | F<br>70  | G<br>71  | H<br>72  | I<br>73  | J<br>74  | K<br>75  | L<br>76  | M<br>77  | N<br>78  | O<br>79  |
| P<br>80  | Q<br>81  | R<br>82  | S<br>83  | T<br>84  | U<br>85  | V<br>86  | W<br>87  | X<br>88  | Y<br>89  | Z<br>90  | [<br>91  | \<br>92  | ]<br>93  | ^<br>94  | _<br>95  |
| `<br>96  | a<br>97  | b<br>98  | c<br>99  | d<br>100 | e<br>101 | f<br>102 | g<br>103 | h<br>104 | i<br>105 | j<br>106 | k<br>107 | l<br>108 | m<br>109 | n<br>110 | o<br>111 |
| p<br>112 | q<br>113 | r<br>114 | s<br>115 | t<br>116 | u<br>117 | v<br>118 | w<br>119 | x<br>120 | y<br>121 | z<br>122 | {<br>123 | <br>124  | }<br>125 | ~<br>126 |          |

Коды со 128 по 255 являются национальными, т. е. в национальных кодировках одному и тому же коду соответствуют различные символы.

В настоящее время существует пять различных кодовых таблиц для кодирования русских букв:

***KOI8-R, CP1251, Mac, ISO, Unicode***

Примеры кодовых таблиц для кодирования русских букв:

***KOI8-R***

***CP1251***

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 |
| 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 |
| 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 |
| 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 |
| 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 |
| 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 |
| 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 |
| 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 |
| 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 |
| 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 |
| 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 |
| 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 |
| 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 |
| 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 |
| 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 |



|     |     |     |     |     |     |     |     |     |     |      |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|
| —   |     | Г   | Г   | └   | └   | ┌   | ┌   | ┐   | ┐   | ┘    | ■   | ■   | ■   | ■   | ■   |
| 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138  | 139 | 140 | 141 | 142 | 143 |
| ▯   | ▯   | ▯   | ┌   | ■   | ●   | √   | ≈   | ≤   | ≥   | nbsp | ┘   | ◦   | 2   | •   | ÷   |
| 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154  | 155 | 156 | 157 | 158 | 159 |
| =   |     | F   | ё   | П   | Г   | Г   | П   | Г   | Г   | Г    | Г   | Г   | Г   | Г   | Г   |
| 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170  | 171 | 172 | 173 | 174 | 175 |
| Г   | Г   | Г   | Ё   | Г   | Г   | Г   | Г   | Г   | Г   | Г    | Г   | Г   | Г   | Г   | ©   |
| 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186  | 187 | 188 | 189 | 190 | 191 |
| ю   | а   | б   | ц   | д   | е   | ф   | г   | х   | и   | й    | к   | л   | м   | н   | о   |
| 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202  | 203 | 204 | 205 | 206 | 207 |
| п   | я   | р   | с   | т   | у   | ж   | в   | ь   | ы   | з    | ш   | э   | щ   | ч   | ъ   |
| 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218  | 219 | 220 | 221 | 222 | 223 |
| Ю   | А   | Б   | Ц   | Д   | Е   | Ф   | Г   | Х   | И   | Й    | К   | Л   | М   | Н   | О   |
| 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234  | 235 | 236 | 237 | 238 | 239 |
| П   | Я   | Р   | С   | Т   | У   | Ж   | В   | Ь   | Ы   | З    | Ш   | Э   | Щ   | Ч   | Ъ   |
| 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250  | 251 | 252 | 253 | 254 | 255 |



|      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Á    | à   | ,   | è   | „   | …   | †   | ‡   | €   | ‰   | É   | <   | Й   | Ї   | Ó   | Ú   |
| 128  | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 |
| á    | ‘   | ’   | “   | ”   | •   | –   | —   | è   | ™   | é   | >   | ò   | í   | ó   | ú   |
| 144  | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 |
| nbsp | ỳ   | Ы   | Э   | И   | Ы   | !   | §   | Ё   | ©   | Ю   | «   | ¬   | shy | ®   | Я   |
| 160  | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 |
| °    | ±   | Ы   | Э   | ’   | μ   | ¶   | •   | ё   | №   | ю   | »   | э   | ю   | я   | 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 |
| Р    | С   | Т   | У   | Ф   | Х   | Ц   | Ч   | Ш   | Щ   | Ъ   | Ы   | Ь   | Э   | Ю   | Я   |
| 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 |
| р    | с   | т   | у   | ф   | х   | ц   | ч   | ш   | щ   | ъ   | ы   | ь   | э   | ю   | я   |
| 240  | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 |

# ЭТО ИНТЕРЕСНО!

Слово «ЭВМ» записано в форме десятичных кодов во всех пяти кодировках на основе кодировочных таблиц:

KOI8-R: 252 247 237

CP1251: 221 194 204

Mac: 157 130 140

ISO: 205 178 188

Если перевести коды в двоичную систему счисления, то слово «ЭВМ» будет в различных кодировках иметь вид:

KOI8-R: 11111100 11110111 11101101

CP1251: 11011101 11000010 11001100

Mac: 10011101 10000010 10001100

ISO: 11001101 10110010 10111100

