字符编码

一进制

山东肥城

字节

字符

a123212 哈

ASCII

Oct	Char		Dec	Нх	Oct	Html	Chr	Dec	Нх	Oct	Html	Chr	Dec	Нх	Oct	Html C
000	NUL	(null)	32	20	040	@#32;	Space	64	40	100	a#64;	. 0	96	60	140	`
001	SOH	(start of heading)	33	21	041	!	1	65	41	101	A	A	97	61	141	a
002	STX	(start of text)	34	22	042	 4 ;	rr	66	42	102	B	В	98	62	142	498 ;
003	ETX	(end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c
004	EOT	(end of transmission)	36	24	044	\$	ş	68	44	104	¢#68;	D	100	64	144	d
005	ENQ	(enquiry)	37	25	045	%	*	69	45	105	E	E	101	65	145	e,
006	ACK	(acknowledge)	38	26	046	&	6	70	46	106	F	F	102	66	146	f
007	BEL	(bell)	39	27	047	'	1	71	47	107	G	G	103	67	147	g
010	BS	(backspace)	40	28	050	((72	48	110	H	H	104	68	150	a#104;
011	TAB	(horizontal tab)	41	29	051))	73	49	111	a#73;	I	105	69	151	i,
012	LF	(NL line feed, new line)	42	2A	052	&# 4 2;	*	74	4A	112	a#74;	J	106	6A	152	j
013	VT	(vertical tab)	43	2B	053	&#43;</td><td>+</td><td>75</td><td>4B</td><td>113</td><td>K</td><td>K</td><td>107</td><td>6B</td><td>153</td><td>a#107;</td></tr><tr><td>014</td><td>FF</td><td>(NP form feed, new page)</td><td>44</td><td>2C</td><td>054</td><td>a#44;</td><td></td><td>76</td><td>40</td><td>114</td><td>a#76;</td><td>L</td><td>108</td><td>6C</td><td>154</td><td>a#108;</td></tr><tr><td>015</td><td>CR</td><td>(carriage return)</td><td>45</td><td>2D</td><td>055</td><td>a#45;</td><td>F 1.1</td><td>77</td><td>4D</td><td>115</td><td>M</td><td>M</td><td>109</td><td>6D</td><td>155</td><td>a#109;</td></tr><tr><td>016</td><td>SO</td><td>(shift out)</td><td>46</td><td>2E</td><td>056</td><td>a#46;</td><td></td><td>78</td><td>4E</td><td>116</td><td>a#78;</td><td>N</td><td>110</td><td>6E</td><td>156</td><td>@#110;</td></tr><tr><td>017</td><td>SI</td><td>(shift in)</td><td>47</td><td>2F</td><td>057</td><td>/</td><td>/</td><td>79</td><td>4F</td><td>117</td><td>O</td><td>0</td><td>111</td><td>6F</td><td>157</td><td>o</td></tr><tr><td>020</td><td>DLE</td><td>(data link escape)</td><td>48</td><td>30</td><td>060</td><td>a#48;</td><td>0</td><td>80</td><td>50</td><td>120</td><td>O;</td><td>P</td><td>112</td><td>70</td><td>160</td><td>@#112;</td></tr><tr><td>021</td><td>DC1</td><td>(device control 1)</td><td>49</td><td>31</td><td>061</td><td>a#49;</td><td>1</td><td>81</td><td>51</td><td>121</td><td>Q</td><td>Q</td><td>113</td><td>71</td><td>161</td><td>@#113;</td></tr><tr><td>022</td><td>DC2</td><td>(device control 2)</td><td>50</td><td>32</td><td>062</td><td>2</td><td>2</td><td>82</td><td>52</td><td>122</td><td>R</td><td>R</td><td>114</td><td>72</td><td>162</td><td>@#114;</td></tr><tr><td>023</td><td>DC3</td><td>(device control 3)</td><td>51</td><td>33</td><td>063</td><td>3</td><td>3</td><td>83</td><td>53</td><td>123</td><td>4#83;</td><td>S</td><td>115</td><td>73</td><td>163</td><td>@#115;</td></tr><tr><td>024</td><td>DC4</td><td>(device control 4)</td><td>52</td><td>34</td><td>064</td><td>4</td><td>4</td><td>84</td><td>54</td><td>124</td><td>4;</td><td>T</td><td>116</td><td>74</td><td>164</td><td>t</td></tr><tr><td>025</td><td>NAK</td><td>(negative acknowledge)</td><td>53</td><td>35</td><td>065</td><td>5</td><td>5</td><td>85</td><td>55</td><td>125</td><td>U</td><td>U</td><td>117</td><td>75</td><td>165</td><td>u</td></tr><tr><td>026</td><td>SYN</td><td>(synchronous idle)</td><td>54</td><td>36</td><td>066</td><td>4;</td><td>6</td><td>86</td><td>56</td><td>126</td><td>V</td><td>V</td><td>118</td><td>76</td><td>166</td><td>v</td></tr><tr><td>027</td><td>ETB</td><td>(end of trans. block)</td><td>55</td><td>37</td><td>067</td><td>7;</td><td>7</td><td>87</td><td>57</td><td>127</td><td>W</td><td>W</td><td>119</td><td>77</td><td>167</td><td>w</td></tr><tr><td>030</td><td>CAN</td><td>(cancel)</td><td>56</td><td>38</td><td>070</td><td>8</td><td>8</td><td>88</td><td>58</td><td>130</td><td>X</td><td>X</td><td>120</td><td>78</td><td>170</td><td>x</td></tr><tr><td>031</td><td>EM</td><td>(end of medium)</td><td>57</td><td>39</td><td>071</td><td>9</td><td>9</td><td>89</td><td>59</td><td>131</td><td>Y</td><td>Y</td><td>121</td><td>79</td><td>171</td><td>y</td></tr><tr><td>032</td><td>SUB</td><td>(substitute)</td><td>58</td><td>ЗA</td><td>072</td><td>:</td><td>:</td><td>90</td><td>5A</td><td>132</td><td>Z</td><td>Z</td><td>122</td><td>7A</td><td>172</td><td>@#122;</td></tr><tr><td>033</td><td>ESC</td><td>(escape)</td><td>59</td><td>ЗВ</td><td>073</td><td>;</td><td>3</td><td>91</td><td>5B</td><td>133</td><td>[</td><td>. [</td><td>123</td><td>7B</td><td>173</td><td>@#123;</td></tr><tr><td>034</td><td>FS</td><td>(file separator)</td><td>60</td><td>3С</td><td>074</td><td>4#60;</td><td><</td><td>92</td><td>5C</td><td>134</td><td>&#92;</td><td>A.</td><td>124</td><td>7C</td><td>174</td><td>@#124;</td></tr><tr><td>035</td><td>GS</td><td>(group separator)</td><td>61</td><td>ЗD</td><td>075</td><td>@#61;</td><td>=</td><td>93</td><td>5D</td><td>135</td><td>&#93;</td><td>]</td><td>125</td><td>7D</td><td>175</td><td>@#125;</td></tr><tr><td>036</td><td>RS</td><td>(record separator)</td><td>62</td><td>3E</td><td>076</td><td>۵#62;</td><td>></td><td>94</td><td>5E</td><td>136</td><td>	4;</td><td>^</td><td>126</td><td>7E</td><td>176</td><td>a#126;</td></tr><tr><td>037</td><td>US</td><td>(unit separator)</td><td>63</td><td>3F</td><td>077</td><td>?</td><td>2</td><td>95</td><td>5F</td><td>137</td><td>&#95;</td><td></td><td>127</td><td>7F</td><td>177</td><td></td></tr><tr><td></td><td></td><td>-</td><td>'</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>_</td><td></td><td></td><td></td><td></td><td>T</td></tr></tbody></table>										

Ú	4.00										
E.	160	á	176		192	L	208	Ш	224	α	240
æ	161	í	177	*****	193	1	209	₹	225	ß	241
Æ	162	ó	178		194	т	210	π	226	Γ	242
ô	163	ú	179		195	H	211	L	227	π	243
ö	164	ñ	180	4	196	- (212	Ŀ	228	Σ	244
ò	165	Ñ	181	4	197	+	213	F	229	σ	245
û	166	2	182	1	198	F	214	П	230	μ	246
ù	167	۰	183	П	199	⊩	215	#	231	τ	247
ÿ	168	ė.	184	7	200	L	216		232	Φ	248
Ö	169	-	185	4	201	F	217	J	233	Θ	249
Ü	170	4	186		202	<u>JL</u>	218	г	234	Ω	250
¢	171	1/2	187	7	203	īF	219		235	δ	251
£	172	1/4	188	IJ	204	ŀ	220		236	00	252
¥	173	i	189	Ш	205	=	221	1	237	ф	253
R	174	«	190	4	206	#	222		238	ε	254
f	175	»	191	٦	207	≟	223		239	\wedge	255
							S	ource:	www.	Looku	pTabl
	æÆôöòûúÿÖÜ¢£	Æ 162 ô 163 ö 164 ò 165 û 166 ù 167 ÿ 168 Ö 169 Ü 170 ❖ 171 Æ 172 ¥ 173 № 174	æ 161 í Æ 162 ó ô 163 ú ö 164 ñ ò 165 Ñ û 166 ° ù 167 ° ÿ 168 ¿ Ö 169 □ Ü 170 □ ♀ 171 ⅓ £ 172 ⅓ ¥ 173 ; № 174 «	# 161 i 177 # 162 ó 178 ô 163 ú 179 ö 164 ñ 180 ò 165 Ñ 181 û 166 182 ù 167 ° 183 ÿ 168 ¿ 184 Ö 169 □ 185 Ü 170 □ 186 ҫ 171 ½ 187 £ 172 ¼ 188 ¥ 173 j 189 № 174 ≪ 190	## 161 f 177 ## 162 6 178 ## 163 ú 179 6 163 ú 179 6 164 ñ 180 6 165 Ñ 181 166	### 161 f 177 193 ### 162 6 178 194 6 163 ú 179 195 6 164 ñ 180 196 6 165 Ñ 181 197 û 166	## 161 f 177 ## 193	## 161 f 177 ## 193 \(\text{1} \) 209 ## 162 6 178 ## 194 \(\text{7} \) 210 6 163 \(\text{4} \) 179 \ 195 \ 195 \ 211 6 164 \(\text{6} \) 180 \ 196 \ - 212 6 165 \(\text{N} \) 181 \ \dagger 197 \ + 213 6 166 \(\text{4} \) 182 \ \dagger 198 \ \dagger 214 6 167 \(\text{6} \) 183 \ \dagger 199 \ \dagger 215 7 168 \(\text{7} \) 183 \ \dagger 200 \ \dagger 216 6 169 \ - 185 \ \dagger 201 \ \dagger 217 7 \(\text{U} \) 170 \ - 186 \ \dagger 202 \ \dagger 218 6 171 \ \dagger 2 187 \ \dagger 203 \ \dagger 219 6 172 \ \dagger 4 188 \ \dagger 204 \ \dagger 220 7 173 \ 189 \ \dagger 205 \ = 221 7 175 \(\text{8} \) 191 \ \dagger 207 \ \dagger 222 7 175 \(\text{9} \) 191 \ \dagger 207 \ \dagger 223	## 161 i 177	## 161 f 177	æ 161 f 177 193 ⊥ 209 ∓ 225 B Æ 162 ó 178 194 ⊤ 210 π 226 Γ ô 163 ú 179 195 211 ⊥ 227 π ö 164 ñ 180 196 — 212 ½ 228 Σ ò 165 Ñ 181 197 + 213 ឝ 229 σ û 166 182 198 ‡ 214 π 230 μ ù 167 183 π 199 215 231 τ ÿ 168 ¿ 184 ¬ 200 ½ 216 ‡ 232 ф Ö 169 ¬ 185 201 217 」 233 ⊕ Ü 170 ¬ 186 202 ½ 218

混乱时代

你会如何设计?

- 字符表 (Character repertoire)
- 给字符表里的抽象字符编上一个数字,也就是字符集合到一个整数集合的映射。这种映射称为编码字符集(CCS:Coded Character Set)
- 将CCS里字符对应的整数转换成有限长度的比特值,便于以后计算机使用一定长度的二进制形式表示该整数。这个对应关系被称为字符编码表(CEF:Character Encoding Form)

Unicode

UTF-8 UTF-16 GBK

Unicode Transformation Format

"严" unicode 4E25(100111000100101)

从"严"的最后一个二进制位开始,依次从后向前填入格式中的x,多出的位补0。这样就得到了,"严"的UTF-8编码>是"11100100 10111000 10100101",转换成十六进制就是OxE4B8A5。

Byte order mark

0010 0000 1010 1100

字体

- 1. 字体是用来显示字符的,将不可见的内存中的字符显示为可见的图形。
- 2. 日常使用的字符是由特定字符集(比如 Unicode) 定义的。
 - 3. 并非所有字体都遵循你用的字符集。

iconv -f UTF8 -t GB18030 a.csv >b.csv

QA