常用编码方式中英文对照表

|  |  |
| --- | --- |
| **编码方式** | **说明** |
| ASCII | 7位编码，编码范围是0x00-0x7F。ASCII字符集包括英文字母、阿拉伯数字和标点符号等字符。其中0x00-0x1F和0x7F共33个控制字符 |
| GB2312 | 基于区位码设计的，区位码把编码表分为94个区，每个区对应94个位，每个字符的区号和位号组合起来就是该汉字的区位码。区位码一般 用10进制数来表示，如1601就表示16区1位，对应的字符是“啊”。在区位码的区号和位号上分别加上0xA0就得到了GB2312编码 |
| GBK | GB2312编码的超集，向下完全兼容GB2312，同时GBK收录了Unicode基本多文种平面中的所有CJK汉字。同 GB2312一样，GBK也支持希腊字母、日文假名字母、俄语字母等字符，但不支持韩语中的表音字符（非汉字字符）。GBK还收录了GB2312不包含的 汉字部首符号、竖排标点符号等字符 |
| GB18030 | 向下兼容GBK和GB2312，兼容的含义是不仅字符兼容，而且相同字符的编码也相同。GB18030收录了所有Unicode3.1中的字符，包括中国少数民族字符，GBK不支持的韩文字符等等 |
| ISO-8859-1 | 单字节编码，向下兼容ASCII，其编码范围是0x00-0xFF，0x00-0x7F之间完全和ASCII一致，0x80-0x9F之间是控制字符，0xA0-0xFF之间是文字符号 |
| UTF-16 | 变长编码，用两个字节对BMP内的字符编码，用4个字节对超出BMP（基本多语言面，Basic Multilingual Plane, BMP）范围的辅助平面内的字符作编码 |
| UTF-8 | UCS字符集（通用字符集，Universal Character Set）的另一种编码方式，UTF-16的每个单元是两个字节（16位），而UTF-8的每个单元是一个字节（8位）。UTF-16中用一个或两个双字节表示一个字符，UTF-8中用一个或几个单字节表示一个字符 |