* 最初使用ascii（127个）编码表示常用字符，单字节就可以表示（0x00-0x7f）;
* Iso8859-1为单字节编码，总共表示256个字符，当找不到对应字符映射时，会将字符映射到0x3f对应的字符，即为？。
* 出现GBK编码来表示包含汉字的更多字符，全角状态下中英文均为双字节编码，半角时实际使用ascii 编码，所以为单字节。GBK中字符是一个或者两个字节，单字节00–7F这个区间和ASCII是一样的，双字节字符的第一个字节在81-FE之间，通过这个可以判断是单字节还是双字节；
* 后出现更为强大的unicode编码、utf16、utf-8等国际通用编码字符集，其中utf16为双字节编码（即所有字符均为双字节），utf8则为变长编码（1-6字节不等）。

Utf-8编码格式：

单字节编码

0xxxxxxx

双字节编码

110xxxxx 10xxxxxx

三字节编码

1110xxxx 10xxxxxx 10xxxxxx

因此当gbk格式编码转到utf-8后，文件变大了。

参见《深入分析java web内幕》70页 编码，解码

字符编码集：如ascii、GBK、unicode

编码字符集：gbk编码集包含了 gbk2312、gbk编码；Unicode包含utf8、utf-16

几种编码的比较：

对于中文gbk与utf编码集均可处理，gbk与gbk2312类似但gbk表示范围更大；utf8与utf16编码规则有区别，utf16编码效率更高，适合在本地磁盘和内存中使用，不适合在网络传输中使用。utf-8则更适合网络传输，单个字符损害不会影响其他字符，效率则介于gbk与utf-16 之间。