链接：https://blog.csdn.net/feather_wch/article/details/79799718

1、ASCII码的由来和作用

1. 20实际60年代，美国制定的 字符编码 ， 规定了英文字符与二进制间的关系
2. ASCII码 一共规定了 128个字符的编码 ， 包括 32个不可见控制符号
3. ASCII码 只占用了一个 字节 的后面 7位

2、Unicode编码的特点

1. Unicode 只是符号集，只规定了 符号 的 二进制代码，没有规定 如何存储
2. 对于 英文字符 来说， 很多字节都是0 是极大的浪费，而对于 部分语言(中文等) 又需要更多的字节。
3. Unicode 不同的 存储方式 导致了 Unicode 难以推广

3、UTF-8的由来和特点

1. UTF-8 是 Unicode 在 互联网 上使用最多的 实现方式 ---UTF-16(2字节或者4字节)，UTF-32(4个字节)使用很少。
2. UTF-8 是一种 变长的编码方式 ---根据不同符号能改变字节长度(1字节~4字节)

4、UTF-8的编码规则

1. 单字节符号 ： 字节第一位 为 0，后面 7位 为该符号的 Unicode码 (因此， 英文字母 的 UTF-8 和 ASCII码 相同)
2. n字节的符号 ， 第一个字节 放置 n个1 +一个 0 ， 后面 每个字节 的开头为 10 ， 其余的 位 用 Unicode码 填充。(这样第一字节开头有几个 1 就表示该 字符 占据几个 字节)

Unicode符号范围 (十六进制)	UTF-8编码方式（二进制）
0000 0000-0000 007F	0xxxxxxx
0000 0080-0000 07FF	110xxxxx 10xxxxxx
0000 0800-0000 FFFF	1110xxxx 10xxxxxx 10xxxxxx
0001 0000-0010 FFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

5、Little endian和Bit endian(大端和小端)

1. 如 Unicode码-4E25
2. Big endian：存储时 4E在前 ， 25在后
3. Little endian：存储时 25在前，4E在后
4. Unicode规定 ： 文件开头 头两个字节是 FE FF-表示大头 ， 乳沟是 FF FE-表示小头

6、GBK编码

> 1. GBK 的文字编码是 双字节 来表示的(中、英文字符均使用双字节)
> 2. 为区分中文，将其最高位都定成1。

大小端的故事

> 这两个古怪的名称来自英国作家斯威夫特的《格列佛游记》。在该书中，小人国里爆发了内战，战争起因是人们争论，吃鸡蛋时究竟是从大头(Big-endian)敲开还是从小头(Little-endian)敲开。为了这件事情，前后爆发了六次战争，一个皇帝送了命，另一个皇帝丢了王位。