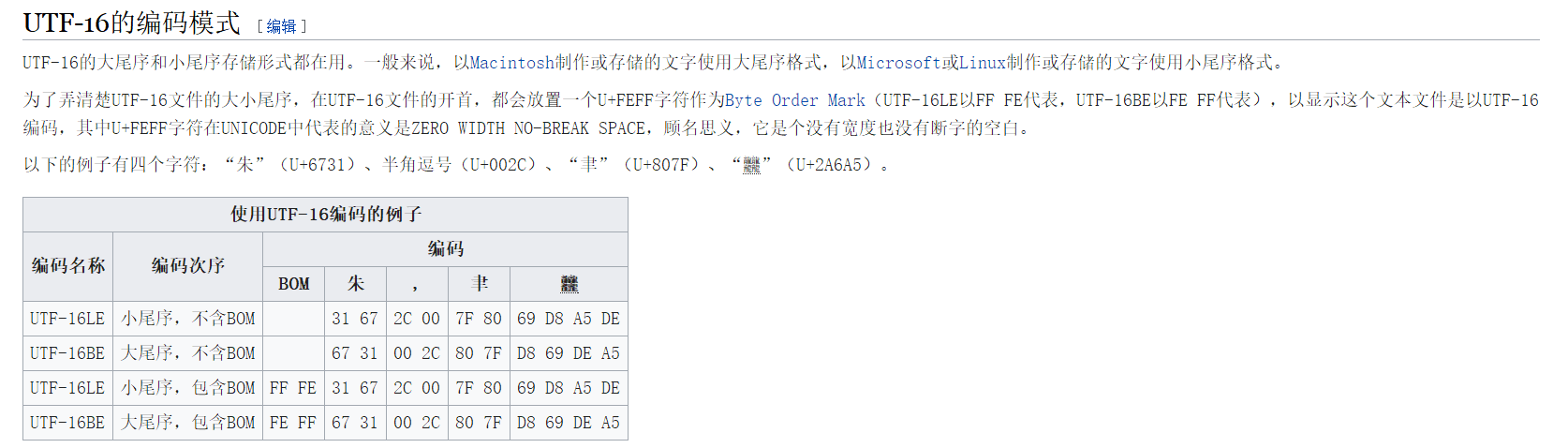
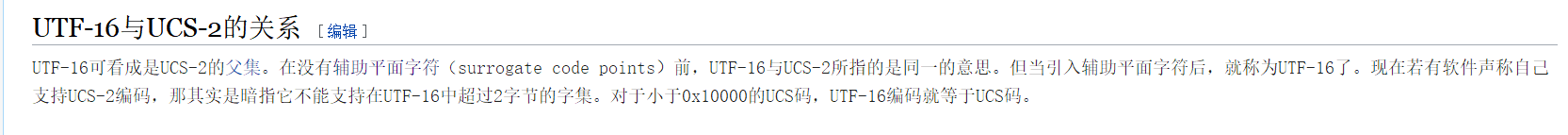
Unicode系列的编码表示可以使用&#nnn 或者 &#xmmm ; （n表示10进制数字，m表示16进制字符）

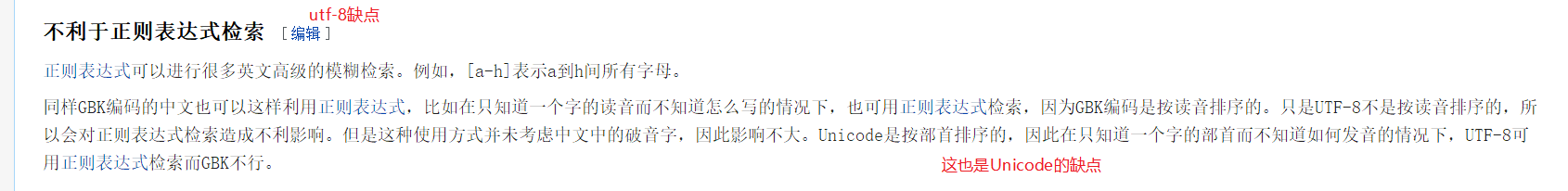
## Unicode

* ***Unicode***一般使用2个字节存储字节码，比如：U+4AE0； 但是这只表示0号平面的字符（英文：Basic Multilingual Plane，简写BMP。又称为“零号平面”、plane 0）
* 在0号平面以外的字符，需要使用五个或六个数字（16进制）
* 在Unicode 3.0里使用“U-”然后紧接着八个数字，而“U+”则必须随后紧接着四个数字。
* 多平面（17个平面：0-16）需要21bit，但实际上多平面编码方式仍然占用了4个字节和UTF-32（UCS4）一样

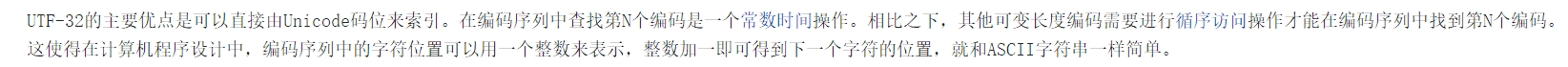
## **Utf-16（UCS2）**

* Utf-16是变长的：Unicode字符的码位，需要1个或者2个16位长的码元来表示，因此这是一个变长表示。
* [大端序](https://zh.wikipedia.org/wiki/%E5%AD%97%E8%8A%82%E5%BA%8F" \o "字节序)（Big-Endian，简写为UTF-16 BE）、[小端序](https://zh.wikipedia.org/wiki/%E5%AD%97%E8%8A%82%E5%BA%8F" \o "字节序)（Little-Endian，简写为UTF-16 LE），目前在PC机上的Windows系统和Linux系统对于UTF-16编码默认使用UTF-16 LE
* 
* 

## **Utf-8**

* UTF-8是变长，使用1-3个字节表示字符码
* UTF-8是为了节省存储空间（比如：一个全是ASCII字符的文件，如果使用Unicode编码会浪费大量的空间）
* 

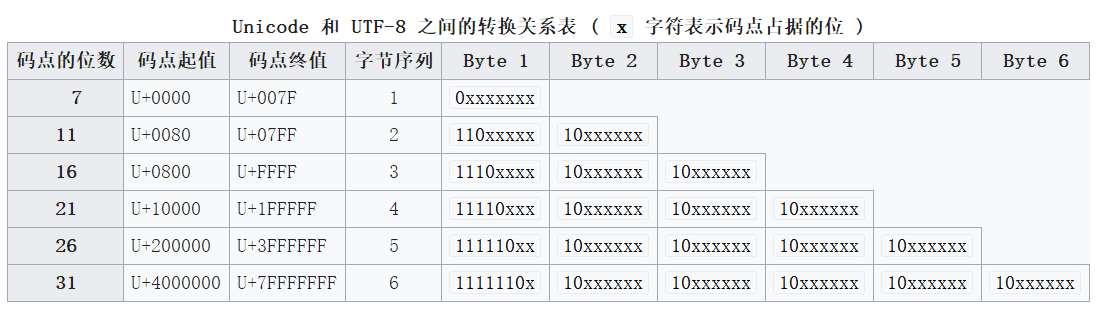
**UTF-32**

* UTF-32长度是固定的；
* 我想utf-32的速度应该会快一些

## **ANSI**

* 非Unicode编码的[ANSI](https://zh.wikipedia.org/wiki/ANSI" \o "ANSI) —— 对于英文系统即[ASCII](https://zh.wikipedia.org/wiki/ASCII" \o "ASCII)编码，中文系统则为[GB2312](https://zh.wikipedia.org/wiki/GB2312" \o "GB2312)或[Big5](https://zh.wikipedia.org/wiki/Big5" \o "Big5)编码
* Windows中的保存格式：除ansi，其余三种为“Unicode”（对应UTF-16 LE）、“Unicode big endian”（对应UTF-16 BE）和“UTF-8”。

## **转码**

* Unicode转 utf-8：
* Unicode转utf-16：