

字符串不同编码的来源

ASCII

- 由于计算机是美国人发明的，最早只有127个字符被编码到计算机里，也就是大小写字母、数字和一些符号。

Unicode

- 由于全球有上百种语言，各有各国的标准，为了解决冲突，Unicode把所有语言统一到一套编码中，就不会乱码问题了；最常用的是2byte表示一个字符，偏僻的4个字节。

UTF-8

- 如果文体是英文的话，用Unicode编码比用ASCII编码多出一倍的存储空间，在存储和传输上不划算，本着节约精神，将Unicode编码转化为"可变长编码"UTF-8编码。
- 在计算内存中，统一使用Unicode编码，当需要保存到硬盘或传输时转化为UTF-8编码。###python第一行
- `# -*- coding:utf-8 -*-`