**ASCII, Unicode和UTF-8**

**ASCII**

单字节码(1byte=8bit)，使用7位二进制数来表示所有的大小写字母，数字0-9，标点符号，以及其他控制符号，共定义了128个字符。主要用于表示英语，无法表示中文

**Unicode**

双字节码，将所有语言都统一到一套编码里，但比ASCII编码多一倍的存储空间，一般采用四位16进制数表示，即\uhhhh

**UTF-8**

是针对Unicode的一种可变长编码，可以用来表示Unicode中的任何字符，根据不同的使用频率将字符编码为不同长度，常用的英文字母被编码成1个字节，汉字通常是3个字节，只有很生僻的字符才会被编码成4-6个字节，是电子邮件，网页以及其他存储或传送文字的应用中优先采用的编码。ASCII编码实际上可以被看成是UTF-8编码的一部分，所以，大量只支持ASCII编码的历史遗留软件可以在UTF-8编码下继续工作。

**计算机中通用的字符编码工作方式**

1 在计算机内存中，统一使用Unicode编码，当需要保存到硬盘的时候，就转换为UTF-8编码

2 当用记事本编辑的时候，从文件读取的UTF-8字符被转换为Unicode字符到内存里，编辑完成后，保存的时候再转换为UTF-8保存到文件  
3 浏览网页时，服务器会把动态生成的Unicode内容转换为UTF-8再传输到浏览器