**ASCII-American Standard Code for Information Interchange**

* **Character repertoire:**  
  ASCII 严格来讲就是7个bit大小的字符集，也就是code point介于0-127之间的字符集合。
* **Character code:**   
  32-126之间的字元是可打印字元，其他是控制字元。
* **Character encoding:**   
  ASCII可以没有任何编码就可以在计算机中用一个字节表示，也就是每个code point被表示成等价的单字节二进制形式。

**LATIN1-8bit sing byte coded graphic character sets.**

* **Character repertoire:**Aka ISO-8859-1。是7bit ASCII 字元集的扩充的一种，是8个bit大小的字符集，也就是code point 介于0-255之间的字符集和。
* **Character code:**包含有191个可打印字符，其余是控制字符或者扩展的欧洲特殊字符。
* **Character encoding:**与ASCII相似， 每个code point被表示成等价的单字节二进制形式。

**UTF8-8bit Unicode Transfer Format**

* **Character repertoire:**包含世界上大部分书写系统的使用的字符，大概1百万个code point (1,114,112 = 220 + 216 )。
* **Character code:**

2,684个保留字符。 98893个图形字符。 435个控制，格式化等特殊用途的字符 。

* **Character encoding:**是针对Unicode的一种变长的字符编码。能表示Unicode标准中任意一个字符。 UTF8可以将一个字符编码乘1到4个字节大小来表示。

1. 128US-ASCII 字符需要一个字节（U+0000-U+007F)

2. 对于带有区分符号的Latin字母和来自西欧的字符需要两个字节(U+0080-U+7FF**)**

3. BMP其他的字符需要3个字节

4. Non-BMP的其他字符需要4个字节。

**ASCII, LATIN1, UTF8 关系：**

    为了向后兼容， Unicode分配128ASCII和256 LATIN1字符的code point没有改变，与它们在ASCII和Latin1的code point 相同。 因此一个只包含ASCII字符的UTF8文件等同于ASCII文件。 同理，每个正确编码的ASCII文件也是有效地UTF8文件。对于12-256的LATIN1字符，因为其编码的特殊性-UTF8需要2个字节来表示一个 LATIN1字符，也就是二者编码的文件当然不能等价。

     如果只考虑各个编码单字节的外围：ASCII(0-127), LATIN1(0-255), UTF8(0-253), 一个UTF8编码的字符串也能被保存为LATIN1文件，但显然是乱码的。反过来，LATIN1编码的文件是无法存为一个UTF8的文件，因为一个大于 253的字符在UTF8中是不存在的。