# 字符串编码

## 前言：

我们知道，所有的信息最终都表示为一个二进制的字符串，每一个二进制位（bit）有0和1两种状态。当我们需要把字符'A'存入计算机时，应该对应哪种状态呢，存储时，我们可以将字符'A'用01000010（这个随便编的）二进制字符串表示，存入计算机；读取时，再将01000010还原成字符'A'。那么问题来了，存储时，字符'A'应该对应哪一串二进制数呢，是01000010？或者是10000000 11110101？说白了，就是需要一个规则。这个规则可以将字符映射到唯一一种状态(二进制字符串)，这就是编码。而最早出现的编码规则就是ASCII编码，在ASCII编码规则中，字符'A'既不对应01000010，也不对应1000 0000 11110101，而是对应01000001

01000001，十进制是65,日常做题就都知道这个,A-Z是[65,90]a-z是[97,122]

通常计算机存贮这些字符就是以上面那些数字的二进制。

ASCII是比较古老的编码，只能表示一些英文符号和字母。无法表示其他语言的文字，比如汉子，现在无论数据包还是存储文件的二进制或者16进制数，并不是ASCII嘛

常见编码方法各大变成语言都有封装的函数，比如java有String.getBytes()

很多人不知道具体实现，

下面说说几种常见的编码方法

## UTF-8：