由于人类的语言有太多，因而表示这些语言的符号太多，无法用计算机中一个基本的存储单元byte 来表示，因而必须要经过拆分或一些翻译工作，才能让计算机能理解。

编码的原因可以总结为：   
计算机中存储信息的最小单元是一个字节即 8 个 bit，所以能表示的字符范围是 0~255 个  
人类要表示的符号太多，无法用一个字节来完全表示要解决这个矛盾必须需要一个新的数据结构 char，从 char 到 byte 必须编码

目前的编码格式很多，例如GB2312、GBK、UTF-8、UTF-16 这几种格式都可以表示一个汉字

https://www.cnblogs.com/yaya-yaya/p/5768616.html