

计算机中的数据表示形式

2019年4月2日 星期二 11:18

1. 信息的二进制表示方式

计算机中需要存储视频、音频、图片、文档等内容，在底层都是转换为二进制来进行存储；
字符转二进制，将字符编号，将编号转为二进制存储在计算机底层；

2. 码表

码表就是字符到二进制的映射关系。

世界上有很多码表，通常由国家或国际标准化组织定义，在一定范围内使用。

Latin-1 也叫ASCII 或 ISO8859-1	只有英文、数字、常见符号 历史上出现最早的码表	一个字符占用一个字节 byte	 ASCII码对照表
GB2312	兼容Latin-1 在Latin-1的基础上增加了常用汉字 由中国国家标准定义	一个字符占用两个字节2byte=16bit	略
GBK	兼容GB2312,增加了更多的汉字 由中国国家标准定义	一个字符占用两个字节	略
unicode编码	是一种映射规则，将全世界所有的字符都对应到一个数字上	只是为每个符号指定了一个编码，并没有指定如何映射成二进制	 Unicode编码表（全）
Utf-16/utf-32	兼容latin1 是unicode编码的实现，占用2/4个字节	将unicode定义的符号对应的数字转为定长二进制来代表该符号 java语言的底层使用utf-16来代表字符	略
Utf-8	兼容latin1 是unicode编码的实现，是一种变长的编码形式，用1至4个字节来表示符号，常用符号用1个字节表示，次常用符号用2个	将unicode定义的符号对应的数字转为变长二进制来代表该符号，节省空间，通常在将数据写出磁盘 通过网络发送等操作时优先使用utf-8	 彻底弄懂Unicode ...

字节，非常用符号用3个字节表示，极特殊符号用4个字节表示。 其中汉字在utf-8码表中占3个字符。		
--	--	--