

# 数据表示与信息编码

## ▼ 机器数

### ▼ 机器数

- 带符号的二进制数
- 最高位是符号位
- 0代表正数
- 1代表负数

### ▼ 计算机中采用定长格式的数据

- ▼ 参加运算的结果只能表示在定长范围内
  - 超出长度的溢出

### ▼ 原码

- 简单直观
- ▼ 比较方便地进行乘法运算
  - 尾数相乘，符号位简单相加就能得到乘积
- ▼ 缺点
  - 进行减法时困难

### ▼ 反码

- 一个正数的反码就是它的原码
- 负数的反码是符号位不变，其余各位取反
- 注意机器数为定长数，其前面的一系列0在取反时也要取反

### ▼ 补码

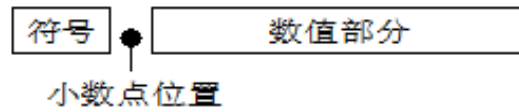
- 正数的补码等于原码
- ▼ 负数的补码等于它的反码加1
  - 最低位加1，进位不改变符号位
- 补码的补码还原为原码
- ▼ 计算机内部存储一般用补码形式
  - 用于简化减法运算

### ▼ 定点数和浮点数

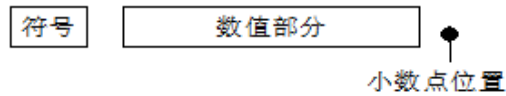
## ▼ 定点数

### ▼ 小数点固定在某一位置

#### ▪ 定点纯小数



#### ▪ 定点纯整数



## ▼ 浮点数

### ▪ 32位二进制浮点格式

数的符号	阶码符号	阶码值	尾数
1 bit	1 bit	7 bit	23bit

## ▼ 数据压缩

### ▼ 无损压缩

#### ▼ 解压后的数据和压缩的数据相同，多用于文本数据的压缩

##### ▼ 霍夫曼压缩

- 频率相关编码数据压缩
- 没有重码
- RLE编码

### ▼ 有损压缩

#### ▼ 解压后的数据不能完全重现压缩前的数据，往往用于多媒体数据的压缩

- JPEG
- MPEG
- MP3

### ▼ 压缩比

- 压缩后码位（大小）比压缩前码位（大小）

### ▼ 图像

#### ▼ 主要格式

- ▼ BMP

- 无损压缩
- GIF
- JPG
- PNG
- 矢量图
- 点位图

## ▼ 进制

### ▼ 常用进制

- 十进制 (D)
- 二进制 (B)
- 八进制 (Q)
- 十六进制 (H)

### ▼ 数制转换

#### ▼ 二进制&八进制

- 以小数点为界，分别将3位2进制数与1位8进制数对应

010	110	101	.	001	010	二进制
↓	↓	↓		↓	↓	
2	6	5	.	1	2	八进制

#### ▼ 二进制&十六进制

- 以小数点为界，分别将4位2进制数与1位8进制数对应

1011	0101	.	0010	1000	二进制
↓	↓		↓	↓	
B	5	.	2	8	十六进制

#### ▼ 十进制&R进制

- 模R取余，再用商除以R取余，直到商为0，各余数倒数排列

#### ▼ 例

- 1234 Mod 16 余数为2 (低位)，商为77
- 77 Mod 16 余数为13 (D)，商为4
- 4 Mod 16 余数为4 (高位)，商为0

### ▼ 小数转换

- 将十进制小数部分乘以2，将进位按序排列
- $0.625 \times 2$  积为1.25 进位位为1（高位）小数部分积0.25  
 $0.25 \times 2$  积为0.5 进位位为0 小数部分积0.5  
 $0.5 \times 2$  积为1.0 进位位为1（低位）小数部分积0

## ▼ 编码

### ▼ ASCII

- 7位二进制编码，被确立为国际标准

### ▼ 拓展ASCII(Latin-1)

- 8位编码

### ▼ Unicode

- UTF

### ▼ 中文编码

- GB2312-1980

### ▼ BIG5

- 港澳台地区繁体汉字

### ▼ GBK

- GB2312-1980的拓展，支持Unicode和BIG5中的所有汉字

### ▼ GB18030,2005版

- 超大型中文编码字符集