

## 1.1 1.2 数据、信息与进位制

### 1. 数据的定义

- 数据是对客观事物的符号表示。
- 在计算机科学中，数据表现形式可以是文字、图形、图像、音频、视频等。

### 2. 信息及信息的特征

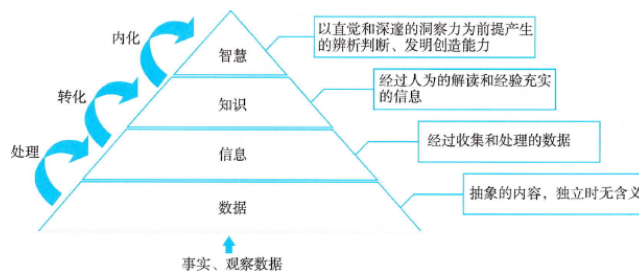
- 信息是数据、信号中所包含的含义
- 信息具有**载体依附性**、时效性、共享性、可加工处理性等特征。

### 3. 知识

- 知识是人类在社会实践中所获得的认识和经验的总和，也是人类在实践中认识客观世界的成果，它包括对事实、信息的描述及在教育 and 实践中获得的技能。
- 知识是可以积累与传承的。

### 4. 数据、信息与知识的关系

- 数据经过解释后产生的意义就是信息，数据是信息的载体，单纯的数字是没有意义。
- 通过归纳、演绎、比较等手段对信息进行挖掘，将万千信息中有价值的部分与已存在的人类知识体系相结合，形成知识。
- 数据、信息与知识的关系可以通过图表示



### 5. 数字化

- 将模拟信号转换为数字信号的过程称为数字化
- 将模拟信号转换成数字信号一般需要经过采样、量化与编码

### 6. 进制转换

- 将十进制  $\rightarrow$   $k$  进制：除  $k$  取余法，将  $k$  进制  $\rightarrow$  十进制 数可用按权相加法（即  $\times k^{i-1}$ ），e.g.  
★  $25D = 11001B, \therefore$

$$25 \div 2 = 12 \dots 1$$

$$12 \div 2 = 6 \dots 0$$

$$6 \div 2 = 3 \dots 0$$

$$3 \div 2 = 1 \dots 1$$

$$1 \div 2 = 0 \dots 1$$

$$\star D2H = 208D, \therefore 13 \times 16^1 + 2 \times 16^0 = 208$$

- 二进制数转换成十六进制数，从二进制数的低位开始，每四位二进制数转换成一位十六进制数；反之，每一位十六进制数可转换成四位二进制数。e.g.

$$\star D2H = 11010010B, \therefore D_{(16)} \rightarrow 13_{(10)} \rightarrow 1101_{(2)}, 2_{(16)} \rightarrow 2_{(10)} \rightarrow 11_{(2)}$$

$$\star 110111B = 37H, \therefore 0111_{(2)} \rightarrow 7_{(10)} \rightarrow 7_{(16)}, 11_{(2)} \rightarrow 3_{(10)} \rightarrow 3_{(16)}$$

## 一、单选题 (每题仅有一个正确选项)

1. (作业本 1.1) 下列有关数据的说法, 正确的是 ( )  
A. 数据必须由数字组成 B. 虚假的数据不能承载任何信息  
C. 数据的价值往往取决于其所承载的信息 D. 所有的数据都是人为创造的
2. (作业本 1.1) 下列关于数据和信息的说法, 正确的是 ( )  
A. 数据的表现形式只能是文字和图像 B. 同一信息对所有人而言其价值是相同的  
C. 有数字才能被输入到计算机中进行处理 D. 信息是数据经分析、解释后得到的
3. (作业本 1.1) 世界第一高峰——珠穆朗玛峰位于中国和尼泊尔两国边界上, 海拔 8848.86 米, 是喜马拉雅山脉的主峰。结合上述事例, 下列关于数据、信息、知识的描述正确的是 ( )  
A. 若在纸上单独书写“8848.86”这个数, 它就已经被赋予了一定的意义  
B. 当人们看到海拔 8000 多米的高度时, 会联想到缺氧、寒冷等词汇, 这是知识的体现  
C. “珠穆朗玛峰峰顶海拔过高, 不宜人类居住。”这体现了人类的智慧  
D. 不同国籍的人引用珠穆朗玛峰高度采用不同的数据, 说明存在虚假数据
4. (作业本 1.1) 下列关于数据和信息的说法, 正确的是 ( )  
A. 数据自古就有, 但信息只在计算机被发明后才出现  
B. 数据量的大小决定了其所承载信息价值的高低  
C. 同一数据出现在不同的应用情境, 表示的信息可能不同  
D. 计算机技术的发展已经使人类可以处理世界上所有数据
5. (作业本 1.1) 下列有关信息的说法, 不正确的是 ( )  
A. 信息超出有效期后不再具有任何价值  
B. 信息无处不在, 且呈现形式多样  
C. 信息的传播、存储必须依附于某种载体  
D. 信息经过加工、处理可以具有更高的使用价值
6. (作业本 1.1) 下列有关计算机处理数据的说法, 不正确的是 ( )  
A. 用计算机处理数据可以获得较高的效率  
B. 数据量较大时, 只有计算机才能进行处理  
C. 计算机已成为处理数据最主要的工具  
D. 计算机可以处理文字、图形、图像、音频、视频等数据
7. (作业本 1.1) 下列有关数据、信息和知识的说法, 正确的是 ( )  
A. 一般而言, 数据都是真实可信的, 而信息可能存在虚假  
B. 人类能采集到的信息中都蕴含着很高的价值  
C. 只要有足够多的信息, 就能建构任意知识  
D. 随着人类社会的发展, 已有的知识也有可能继续完善
8. (作业本 1.1) 古代的数学家发现正多边形的边数越多, 其形状就越接近于圆。魏晋时期数学家刘徽利用割圆术计算出圆周率的近似值为 3.14, 并以此来计算圆的面积。下列有关该事例中数据、信息、知识的说法, 不正确的是 ( )  
A. 刘徽从正六边形开始, 每次倍增边数, 最后割成 192 边, 该过程中采集到了数据  
B. 从此之后, 3.14 这个数据在任何地方都表示圆周率  
C. 刘徽采用割圆术这一方法, 是智慧的体现  
D. 圆面积的近似值可以用 3.14 乘以半径的平方获得, 这是知识

9. (作业本 1.2) 要表示 10 种不同的信息, 需要的二进制位数至少为 ( )  
 A. 1                                      B. 2                                      C. 4                                      D. 10
10. (作业本 1.2) 有十进制数 115, 将其用一个字节的二进制表示, 结果是 ( )  
 A. 0111 0011                      B. 0011 0011                      C. 0110 0111                      D. 1100 1110
11. (作业本 1.2) 十进制数 111 转换为十六进制是 ( )  
 A. 6F                                      B. D7                                      C. 37                                      D. D3
12. (作业本 1.2) 算式:  $01101101B + 16D$  的结果是 ( )  
 A. 11101101B                      B. 01111101B                      C. 80D                                      D. 234D
13. (作业本 1.2) 某二进制数共 5 位, 其中最低位和最高位均为 1, 其他位未知, 该二进制数转换为十进制后的最小值和最大值分别是 ( )  
 A. 9, 15                                      B. 17, 31                                      C. 33, 63                                      D. 32, 64
14. (作业本 1.2) 某数学题如下: 设  $\{a_n\}$  是集合  $\{2^t + 2^s | 0 \leq s < t, s, t \in \mathbb{Z}\}$  中所有的数从小到大排列成的数列, 即  $a_1 = 3, a_2 = 5, a_3 = 6, a_4 = 9, a_5 = 10, a_6 = 12, \dots$  将数列  $\{a_n\}$  各项按照上小下大、左小右大的原则写成如下的三角形数表:
- |   |   |    |    |   |
|---|---|----|----|---|
|   |   | 3  |    |   |
|   | 5 | 6  |    |   |
|   | 9 | 10 | 12 |   |
| * | * | *  | *  | * |
| * | * | *  | *  | * |
- 那么,  $a_{100}$  是 (提示, 将前面几个数转成二进制找规律) ( )  
 A. 1                                      B. 2                                      C. 3                                      D. 4
15. (作业本 1.2) 在计算机中, 原码和补码是两种常见的数字表示方法。原码表示法是在数值前面增加一位符号位 (即最高位为符号位): 正数为 0, 负数为 1, 其余位表示数值的大小。在补码表示法中, 正整数的补码与原码相同, 负整数的补码是将其原码除符号位外的所有位取反 (0 变 1, 1 变 0, 符号位为 1 不变) 后加 1。若用一个字节来表示, 则十进制数 -123 的补码是 ( )  
 A. 11111011                      B. 01111011                      C. 10000100                      D. 10000101

## 1.3 1.5 ASCII 编码、大数据

### 1. 数字化

- 将模拟信号转换为数字信号的过程称为数字化
- 将模拟信号转换成数字信号一般需要经过采样、量化与编码

### 2. 容量单位

- 基本单位与最小单位换算关系： $1B = 8bit$
- $1KB = 2^{10}B$
- $1MB = 2^{10}KB$
- $1GB = 2^{10}MB$
- $1TB = 2^{10}GB$
- $1PB = 2^{10}TB$

### 3. 字符编码

#### ASCII（美国信息交换标准码）

- 用途：将键盘上的所有字符编码成二进制
- 特点：占 1Byte，有效位 7bit，因此有 128 个字符，十进制范围是 0 ~ 127，十六进制是 0 ~ 7F
- 常见 ASCII 码："0" 的 ASCII 码是 48，"A" 是 65，"a" 是 97

#### 汉字编码

- 外码（输入码）：常用的输入码有拼音码、五笔字形码等
- 交换码：为了方便各数字系统（不仅是计算机系统）之间的汉字信息的通信交换，如 GB2312、GBK
- 机内码（处理码）：在计算机内部汉字代码都用机内码，在磁盘上记录汉字代码也使用机内码
- 字形码：用于汉字的输出字模，用于汉字显示，分为点阵法和矢量法

### 4. 大数据特征（4V）

- 数量（Volume）巨大
- 速度（Velocity）快
- 多样（Variety）类型
- 价值（Value）密度低

### 5. 大数据思维

- 大数据要分析的是全体数据，而不是抽样数据。
- 对于数据不再追求精确性，而是能够接受数据的混杂性。
- 不一定强调对事物因果关系的探求，而是更加注重它们的相关性。

### 6. 大数据对社会影响

- 提高工作效率与生活便捷；让决策更精确；增加新的就业机会；也带来了信息、个人隐私泄露等社会问题

#### 一、单选题 (每题仅有一个正确选项)

1. (作业本 1.2) 在电脑上，利用麦克风、声卡等设备录制语音，以文件形式存储在电脑中，并通过音箱进行播放。下列说法正确的是 ( )

- A. 模拟信号通过麦克风直接转换成数字信号  
B. 声卡将语音信号转换成模拟数据并存放在电脑中  
C. 音箱中发出的声音属于数字信号  
D. 数字信号存储为文件需要经过编码
2. (作业本 1.2) 下列有关模拟信号与数字信号的说法, 正确的是 ( )  
A. 用水银温度计测得液体温度属于数字信号  
B. 传感器可以采集模拟信号  
C. 模拟信号未经数字化就可以被计算机处理  
D. 模拟信号转换为数字信号后无法再转换为模拟信号
3. (作业本 1.3) 下列有关 ASCII 编码的说法, 正确的是 ( )  
A. 共有 127 个字符  
B. 每个字符用 1 个字节中的高 7 位编码  
C. 每个字符占用 1 个字节的存储空间  
D. 可以用于汉字字符的编码
4. (作业本 1.3) 下列不属于字符编码的是 ( )  
A. ASCII                      B. Unicode                      C. GB2312                      D. GIF
5. (作业本 1.3) 下列单位换算正确的有 ①  $1\text{Byte} = 8\text{bit}$  ②  $1\text{KB} = 8\text{Byte}$  ③  $1\text{MB} = 1024\text{KB}$  ④  $1\text{GB} = 1024\text{Byte}$  ( )  
A. ①②                      B. ①③                      C. ②③                      D. ②④
6. (作业本 1.3) 在计算机中表示 1 个 GB2312 汉字字符需要 ( )  
A. 1 个字节                      B. 2 个字节                      C. 3 个字节                      D. 4 个字节
7. (作业本 1.3) 在计算机中, 用于存储汉字的编码是 ( )  
A. ASCII 码                      B. 机内码                      C. 拼音码                      D. 五笔字形码
8. (作业本 1.3) 条形码的最后 1 位 (最右边 1 位) 数字为校验码, 其计算方式为: (1) 将条形码编码数字包括校验码在内, 按由右至左的顺序进行编号, 校验码的代码位置序号为 1。(2) 校验码的计算步骤如下: ① 从编码位置序号 2 开始, 对所有偶数位的编码数字求和, 将得到的和乘以 3; ② 从编码位置序号 3 开始, 对所有奇数位的编码数字求和; ③ 将步骤 ① 与步骤 ② 的结果相加, 仅保留其个位数字; ④ 用 10 减去步骤 ③ 的结果, 其差的个位数字即为所求校验码的值。现有条形码, 其编码数字为  $978 - 7 - 04 - 049606 - X$ , 其中  $X$  为校验码。则该值是 ( )  
A. 0                      B. 2                      C. 4                      D. 8
9. (作业本 1.5) 下列关于数据与大数据的说法正确的是 ( )  
A. 数据是现代科学研究的重要资源  
B. 大数据的数据量庞大, 价值密度高  
C. 计算机中的数据都以 ASCII 码存储  
D. 大数据的应用降低了用户隐私信息泄露的风险
10. (作业本 1.5) 我们每天都在跟各类软件打交道: 聊天、购物、看新闻和短视频每一次我们的点击和滑动都会成为数据的一部分, 有关组织通过数据的搜集、存储、分析和可视化技术, 解决大数据海量、高速、多变、价值密度低的问题, 使数据从散乱的信息变成知识和智慧帮助组织解决发展中遇到的实际问题, 对于上述描述。下列说法不正确的是 ( )

- A. 你的每次上网行为及相关数据可能会被采集  
B. 你在上网时的每次操作，蕴含着巨大的价值  
C. 你在聊天、购物、看新闻和短视频等过程中，有可能泄露个人隐私  
D. 网站提供给你的“个性化推荐”，依赖于你的上网行为
11. (作业本 1.5) 关于大数据的特征和大数据对社会影响，说法正确的是 ( )  
A. 大数据的特征主要包括数据量大、速度快、数据类型多、价值密度低  
B. 大数据消除数字鸿沟和信息不对称  
C. 大数据价值密度低，降低了信息泄露的风险  
D. 大数据数据量大，给人们的生活带来了不便
12. (作业本 1.5) 下列数据中属于大数据的是：① 社交平台上产生的数据，② 交通摄像头记录的数据，③ 某校历次考试成绩数据，④ 电商平台交易数据 ( )  
A. ①②③                      B. ①②④                      C. ①③④                      D. ②③④
13. (作业本 1.5) 下列有关大数据的说法，正确的是 ( )  
A. 大数据可以对全体数据进行分析  
B. 在大数据时代，样本数据分析法已经不再使用  
C. 大数据采集的数据以结构化数据为主  
D. 用大数据进行数据处理时，必须保证每个数据都准确无误
14. (作业本 1.5) 某用户在网上购买了一件商品，电商根据交易平台的大数据，给该用户进行个性化推荐时，最不需要考虑的是 ( )  
A. 该商品一般与什么商品组合销售  
B. 购买过该商品的其他用户还会购买什么商品  
C. 该用户曾经购买过什么产品  
D. 该用户为什么购买这件商品

## 1.3 1.4 图像、音频编码、数据管理与安全

### 1. 声音编码

未压缩 Wave 格式文件的存储容量计算，本质是：采样点数  $\times$  每个点容量，具体地：采样频率 (Hz)  $\times$  时间 (s)  $\times$  声道数  $\times$  量化位数 (单位：bit)。

### 2. 图像编码

- 数字图像包括矢量图形和位图图像。
- 矢量图形是用点、线或多边形等基于数学方程的几何图元表示的图像。如  $x^2 + y^2 \leq r^2$  表示半径是  $r$  的实心圆。特点是：容量小、色彩少 (8 位色 = 256 色)、放大不会失真
- 位图图像又称栅格图或点阵图，是由像素点 (pixel) 组成的。容量大、色彩丰富 (32 位真色彩可以到  $2^{32}$  色)、放大有锯齿状失真

### 3. 图像容量计算

未压缩 BMP 格式文件的存储容量计算，本质上也是：像素点数  $\times$  每个点的容量，具体地：水平像素  $\times$  垂直像素  $\times$  颜色位深度 (单位：bit)。已知颜色数量时，需要转成位深度，如 256 色就是 8bit。

### 4. 数据管理

- 计算机数据管理经历了人工管理、文件管理和数据库管理三个阶段
- 数据库是数据管理的主要方式
- 现代数据库可以结构化数据、半结构化数据和非结构化数据

### 5. 文件格式

- 文档：.txt, .docx, .wps, .html, .pdf
- 图像：.bmp, .jpg, .png, .gif
- 音频：.wav, .mp3, .wma
- 视频：.avi, .mpg, .mp4, .mov
- 应用程序 (可执行)：.exe

### 6. 数据安全

- 威胁数据安全的因素：硬盘驱动器损坏、操作失误、黑客入侵、感染计算机病毒、遭受自然灾害等。
- 保护存储数据的介质 (如硬盘) 的安全：磁盘阵列、数据备份、异地容灾等手段。
- 提高数据本身的安全
  - 数据加密：提高数据的**保密性**。通过加密算法和加密密钥将明文转变为密文，而解密则逆过程。
  - 数据校验：是为保证数据的**完整性**进行的一种验证操作。通常用一种指定的算法对原始数据计算出一个校验值，接收方按同样的算法计算出一个校验值，如果两次计算得到的校验值相同，则说明数据是完整的。常见的数据校验方法有 MD5、CRC、SHA-1 等。

### 一、单选题 (每题仅有一个正确选项)

1. (作业本 1.3) 某 Wave 格式的音频文件，其采样频率为 44.1kHz，量化位数为 16bit，2 个声道，其 1 秒的数据量为 ( )  
A. 176.4B                      B. 1.38KB                      C. 172.3KB                      D. 1.35MB
2. (作业本 1.3) 小明要参加学校多媒体作品比赛，其中对于图像类作品的要求：尺寸为  $1280 \times 720$  像素，颜色位深度为 24 位，其文件大小不得超过 450KB。根据这一要求，小明上交的图像作品，其压缩比至少是 ( )  
A. 2 : 1                      B. 4 : 1                      C. 6 : 1                      D. 8 : 1

3. (作业本 1.3) 一部 4K 影片 (3840×2160), 其频为 30fps (帧/秒), 颜色位深度为 24 位, 若不进行压缩, 则其 1 分钟视频的数据量是 ( )
- A. 695MB                      B. 5.56GB                      C. 41.7GB                      D. 333.7GB
4. (作业本 1.4) 为了保障个人信息安全, 下列措施有效的是 ( )
- A. 关闭防火墙软件                      B. 提升自身的信息安全意识
- C. 个人敏感信息保存在 U 盘中                      D. 个人账户的密码不要定期更改
5. (作业本 1.4) 2014 年, 中国铁路售票网站 12306 发生个人数据泄露事件, 犯罪嫌疑人通过收集某游戏网站以及其他多个网站泄露的用户名与密码信息, 尝试登录 12306 网站进行“撞库”非法获取用户信息, 谋取非法利益。结合该事例, 为保护个人信息安全, 下列行为最合理的是 ( )
- A. 用户要提高密码强度                      B. 不同网站采用不同的密码
- C. 每天改变一次密码                      D. 计算机安装杀毒软件
6. (作业本 1.4) 下列文件中属于网页文件的有 ①hello.wav ②index.htm ③nothing.wma ④admin.html ⑤cookie.txt ⑥bg.jpg ( )
- A. ①③                      B. ②④                      C. ②⑤                      D. ④⑥
7. (作业本 1.4) 对下列文件类型归类正确的是 ( )
- A. movie.avi 视频, picnic.jpg 音乐                      B. sky.mp3 音乐, psd.txt 文档
- C. movie.avi 图片, brief.docx 文档                      D. brief.docx 视频, game.rar 压缩
8. (作业本 1.4) 下列有关计算机数据管理的说法, 不正确的是 ( )
- A. 不同的编码规则决定了不同的文件格式
- B. 计算机一般采用树形目录结构来管理文件
- C. 数据库系统的使用提高了数据的可维护性
- D. 传统数据库技术一般用于非结构化数据管理
9. (作业本 1.4) 下列有关数据安全的说法。正确的是 ( )
- A. 经过加密后的数据就不会被破坏
- B. 通过数据校验可以检测数据完整性
- C. 提升数据安全, 只需要做好存储介质的保护
- D. 数据安全问题都由人为因素引发
10. (作业本 1.4) 为了预防自然灾害引起的数据损坏, 一般可采用的措施有 ( )
- A. 安装磁盘阵列系统    B. 对数据进行加密    C. 安装防火墙    D. 建设异地容灾系统