

# 必修一 数据与计算

## 第一章 数据与信息

### 1.1 数据及其特征

信息技术涵盖了获取、表示、传输、存储和加工信息在内的各种技术。

数据：数据是现实世界客观事物的符号记录，是信息的载体，是计算机加工的对象。

在日常生活中，人们常用符号表达客观事物。

在计算机科学中数据具有如下特征：二进制、语义性、分散性、多样性与感知性。

### 1.2 数据编码

模拟信号与数字信号：模拟信号是指用连续变化的物理量所表达的信息，是传到能量的一种形式，在时间和大小上是连续的；数字信号是离散时间信号的数字化表示，在时间和大小上是离散的。

数字信号的大小常用有限位的二进制数来。

编码的基本方式：数据基本上是通过编码将模拟信号转换为数字信号进行存储和传输，文字、图像、声音等类型的数据都可以经过编码进行存储和传输。

文字编码：单字节编码（ASCII 码和莫尔斯码）和双字节编码（国标码和统一码）

ASCII 码使用指定的 7 位或 8 位二进制数组合来表示 128 或 256 种可能的字符。

英文字母 A 的 ASCII 码为 01000001，转化为十进制（序号）后是多少？ $2^6+2^0$

7 位二进制表示（0000000-1111111）128 种编码。

图像编码：位图又叫点阵图，采用位映射存储格式，每一个像素映射一个数据，存放在以字节位单位的矩阵中。

在计算机二进制系统中，每个 0 和 1 就是一个位（bit），数据存储的最小单位；8 个二进制位称为一个字节（Byte）。

黑白图像，每个像素有 2 种可选颜色（黑、白），称为 1 位图像。16 色图像有 16 种可选颜色，称为 4 位图像（ $2^4=16$ ）；256 色图像称为 8 位图像（ $2^8=256$ ）

有一幅 24 位的位图图像，像素为 1024\*800，试确定其数据文件的大小。

1024\*800\*24，单位是位；1024\*800\*3，注意单位是字节。

声音编码：模拟声音信号（模数转换 A/D）采样\量化\编码成数字信号。

计算机存储容量（文件大小）单位。bit---B---KB---MB---GB---TB

**1B ( 字 节 ) = 8 位**

1KB=1024B      1MB=1024KB      1GB=1024MB      1TB=1024GB

位图图像存储空间计算公式是：

水平像素×垂直像素×每个像素色彩所需位数 / 8（单位：字节 B）

下面表格以分辨率 800×600 的位图图像为例，计算机各种类别位图图像的存储空间。

位图图像类别	总像素	一个像素所占的位数	总的位数(bit 位)	容量(B)
黑白	800×600	1 位 ( $2^1=2$ )	800×600×1	800×600×1/8
256 级灰度	800×600	8 位 ( $2^8=256$ )	800×600×8	800×600×8/8
16 色彩色	800×600	4 位 ( $2^4=16$ )	800×600×4	800×600×4/8
<b>256 色彩色</b>	800×600	<b>8 位</b> ( $2^8=256$ )	800×600× <b>8</b>	800×600×8/8
<b>24 位真彩色</b>	800×600	24 位	800×600× <b>24</b>	800×600×24/8

### 1.3 声音的数字化:

1、自然界的物理量可以分为模拟量和数字量，模拟量的特征是连续变化，数字量的特征是非连续变化。可以通过模数转换器把模拟量转化为数字量。常见的声音格式有：MP3、WAV、MID、WMA等。

通过采样和量化可以把模拟信号转化为数字信号。

**采样**：就是每隔一时间段，从连续的波形上取一点。

**量化**：对采样过来的点，进行数字化。

2、声音的质量取决于采样频率和量化位数。采样频率越高，量化的分辨率越高，所得声音的保真程度也越好，但数据量会越大。

\*3、声音存储量的计算：

声音存储量=采样频率\*量化位数\*声道数\*时间/8（字节B）

**注意：**    **单声道\*1**    **双声道（立体声）\*2**    **四声道\*4**

4、声音素材编辑的常用软件 Cool Edit、Wave Edit、GoldWave 等

5、一段未经压缩的视频文件存储量的计算：

视频文件存储量=一帧图像文件存储量×帧频×时间（帧频×时间=帧数）

### 1.4 信息及其特征

信息是客观世界各种事物的运动状态和变化的反映，是客观事物之间相互关系和相互作用的象征。

信息的基本特征：普遍性、传递性、共享性、依附性、真伪性、价值相对性。

## 第二章 知识与数字化学习

知识：知识是人们运用大脑对获取或积累的信息进行系统化的提炼、研究和分析的结果，能够精确的反应事物的本质。

数据是现实世界客观事物的符号记录；信息是经过加工处理的、具有意义的信息；知识是人们运用大脑对获取或积累的信息进行系统化的提炼、研究和分析的结果，能够精确的反应事物的本质；**智慧是为了达到预定目标而运用知识解决问题的创新思维能力。**

**数据、信息、知识、智慧**是逐渐递进的关系，前者是后者的基础和前提，后者是前者的抽象和升华。

数字化工具可大致分为：数据采集工具、信息获取工具、信息检索工具、数字化可视化表达工具、信息通信交流与传输工具、信息存储工具、信息加工工具、三维设计工具。

数字化工具与资源的优势：获取的便捷性、形式的多样性、资源的共享性、平台的互助内容的扩展性。

数字化学习的特点：问题化、合作性、个性化、创造性和再生性、开放性。

## 第三章 算法基础

3.1 体验计算机解决问题的过程：分析问题、设计算法、编写程序、调试运行。

3.2 算法及其描述：算法是指在有限步骤内求解某一问题所使用的一组定义明确的规则，通俗的说算法就是解决问题的方法和步骤；

算法的特征：有穷性、确定性、数据输入（一个算法必须有零个或多个输入）、数据输出（一个算法有一个或多个输出）、可行性；

算法描述的方法：自然语言描述法、流程图描述法、伪代码描述法。

计算机程序与程序设计语言：计算机程序是指为了得到某种结果而可以由计算机等具有信息处理能力的装置执行的代码化的命令。简而言之，计算机程序就是指计算机可以识别运行的指令集合。

常用的计算机主要包括：运算器、控制器、存储器、输入设备和输出设备五大基本部件。计算机内部采用二进制形式表示和存储指令或数据，把解决问题的程序和需要加工处理的原始数据是先转换成二进制数，并存入存储器中。

计算机程序设计语言：经历了机器语言、汇编语言到高级语言的发展历程。

机器语言：由 0 和 1 组成的二进制代码，由计算机直接识别，是第一代编程语言，也是编程的低级语言。

汇编语言：不能被计算机直接识别，介于机器语言和高级语言之间。

高级语言：不能被计算机直接识别和执行，必须经过编译程序或解释程序将其翻译成机器语言。例如：VB、Python、C、C++等。

## 第四章 程序设计基础

### 4.1 程序设计语言的基础知识

4.1.1 Python 的常量和变量：常量是指在程序运行过程中其值始终不发生变化的量，通常指固定的数值或字符串。例如：45, -30.2001, " Hello" 等。

变量是指在程序运行过程中其值可以发生变化的量。在程序设计语言中，变量可以用指定的名字来代表，即变量由变量的“标识符”和变量的“内容”（称“值”）两部分组成。

标识符有字母、数字、下划线组成，并且其第一个字符必须是字母或下划线，而不能是数字，如 s1, k, num 等。在标识符中，区分大小写。

### 4.1.2 Python 数据类型：数值、字符串、列表、元组、集合、字典等

【数值】整型 (int)、浮点型 (float)、复数 (complex)

【字符串】以单引号或双引号括起来的任意文本。

【列表】写在 “[ ]” 之间、用逗号隔开的元素列表，列表的数据项不需要具有相同的类型，与字符串的索引一样，列表索引从 0 开始。

【元组】元组与列表相似，不同之处在于元组列表不能修改。元组使用小括号，元素之间用逗号隔开。

【集合】是一个无序不重复序列。基本功能是进行成员关系测试和删除重复元素。使用大括号或者 set () 函数创建集合。

【字典】字典是无序的对象集合。

### 4.1.3 python 的运算符与表达式

1、算术运算符与算术表达式 (\*\*, //)

2、逻辑运算符与逻辑表达式 (and、or、not)

3、关系运算符与关系表达式 (> < <>(不等于) == ……)

4、赋值运算符与赋值表达式 (=、+=、-=、\*=、/=)

5、成员运算符 (in、not in)

6、运算符的优先级 (\*\*)\* / % // >+ - > ' >' > '==……' or' )

### 4.1.4 python 的函数

abx(x)、len(x)、min(x)、max(x)、int(x)、float(x)……

## 4.2 运用顺序结构描述问题求解过程

### 4.2.1 数据的输入与输出

【输入】input() 函数      【输出】print() 函数

### 4.2.2 顺序结构的应用：

```
n=int(input('输入购买笔记本 1 和笔记本 2 的本书均为：'))
a=3*n
b=5*n
y=(a+b)*(1-0.8)
print('购买笔记本 1 和购买笔记本 2 的本书均为：', 'n', '本')
print('可节省的费用为：', "%.2f" %y, '元')
```

## 4.3 运用选择结构描述问题求解过程

Python 中的选择结构使用 if 和 else 关键字来构造，语法如下：

if 条件：

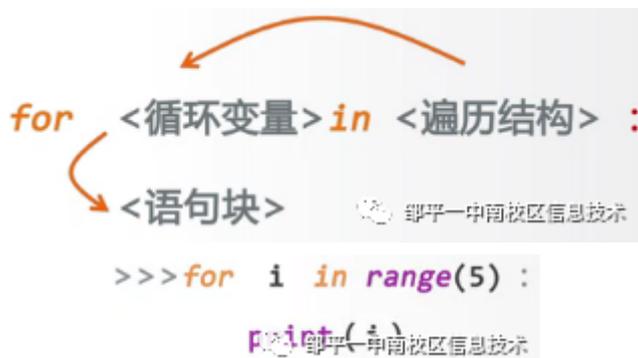
    条件为真时要执行的语句块

else：

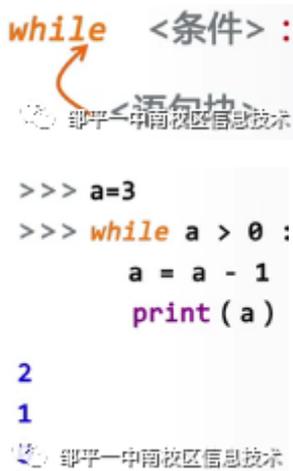
    条件为假时要执行的语句块

选择结构根据条件的判断结果来决定执行哪个语句块。在任何一次运行中，两个分支的语句块只执行其中的一个。不可能两个语句块同时执行。选择结构执行完毕，继续执行其后的语句。

## 4.4 运用循环结构描述问题求解过程



当条件成立就会执行循环，



## 第五章 数据处理和可视化表达

### 5.1 认识大数据：

大数据的特征：

1、从互联网产生大数据的角度来看，大数据具有“4V”特征：大量、多样、低价值密度、高速。

2、从互联网思维的角度来看，大数据具有三个特征：样本渐趋于总体、精确让位于模糊、相关性重于因果。

3、从大数据存储与计算的角度来看，大数据具有两个特征：分布式存储和分布式并行计算。

5.2 数据的采集：数据采集的方法：系统日志采集法、网络数据采集法、其它采集法。

5.3 数据的分析与可视化表达：数据分析一般包括特征探索、关联分析、聚类与分类、建立模型、模型评价等。

特征探索主要任务是对数据进行预处理，发现和处理缺失值、异常数据，绘制直方图，观察分析数据的分布特征。

关联分析就是分析并发现存在于大量数据之间的关联性或相关性，从而描述一个事物中某些属性同时出现的规律和模式。

聚类分析是一种探索性的分析。

数据分类是数据分析处理中最基本的方法。

## 第六章 人工智能及其应用

6.1 认识人工智能：人工智能是研究计算机模拟人的某些感知能力、思维过程和智能行为的学科。

典型的智能问答系统主要包括常见问题解答、问题理解、信息检索、文档库、答案抽取五大模块。

### 6.2 人工智能的应用

人工智能的诞生和发展：图灵-人工智能之父

人工智能发展历程经历三个阶段。

## 必修二 信息技术与社会

### 第一章 走进信息社会

1946年，世界上第一台早期计算机 ENIAC 在美国问世。

2006年，联合国将每年5月17日确定为“世界信息日”。

信息社会本质上以“信息活动为基础”。

计算机与网络通信技术是信息技术的核心技术。

### 第二章 信息系统的组成和功能

信息系统是一个由人、硬件、软件、网络和数据资源等构成的人机交互系统。

信息系统的功能：输入、处理、存储、控制、传输与输出。

计算机系统本身就是基本、重要的信息系统，也是其它信息系统的重要组成元素。

计算机的发展以硬件屋里器件的变革为主要特征，先后经历了电子管、晶体管、集成电路、大规模集成电路和超大规模集成电路。

一个完备的计算机系统应该包括计算机的硬件和软件。

计算机工作的基本原理：存储程序控制

信息论创始人香农。

### 第三章 信息系统的网络组建

1、信息系统与外部世界的联系方式的演变由三个阶段。人和计算机对话、计算机与计算机对话、将联网的终端将计算机扩展到物。

2、物联网：三个技术特征：全面感知（无线射频识别、传感器、定位器、二维码）、可靠传递、智能处理。

3、物联网的机构：感知层、网络层、应用层

4、物联网技术：传感技术、射频识别技术、嵌入式系统技术

5、计算机网络的功能：数据传输、资源共享、分布式处理。传输介质：有线和无线

单位换算：1 字节=8 位

实现不同网络之间互联的设备是路由器。

### 第四章 信息系统的软件与应用

1、信息系统的工作过程：C/S（客户机/服务器）、B/S（浏览器/服务器）

2、信息系统的开发过程：系统规划、系统分析、系统设计、系统实施、系统运行与维护

3、信息系统的软件分为：基础软件（操作系统(unix、linux、windows、ios、android 等)、数据库系统)和应用软件。

### 第五章 信息系统的安全风险

#### 知识产权

开发者设计开发的计算机软件，在开发完成之日起就受到法律的保护。作为软件的使用者，应树立法制观念，遵守相关的法律规定，自觉使用正版软件，抵制盗版及未授权的软件。

#### 信息的安全和保护

计算机病毒已经成为危害信息安全的最大隐患，必须加强对计算机病毒的防治。

**1 计算机病毒是人为编制的计算机程序，它能够通过修改程序，尽可能地把自身复制进去，进而传染给其它程序，破坏他人计算机系统。**

计算机病毒具有**寄生性、传染性、潜伏性、破坏性、爆发性**等特点。

#### 2 计算机病毒的防治

病毒的防治从预防、检查和杀毒三方面着手。比较有效的方法是安装杀毒软件并经常升级，定期或不定期地进行检查或杀毒，同时做好数据的备份工作。

### 补充知识

1. 二进制与十进制、十六进制的相互转换（正整数位）（注：在考试中可使用计算机转换）

a. 计算机中的信息都是以二进制的信息存储的。  
b. 二进制代码的特征：二进制数只由“0”与“1”两个数字组成，运算规则为“逢二进一”。

c. 数字后面加 B 表示二进制，D 表示十进制，H 表示十六进制

d. N 个二进制位能最多表示  $2^n$  个不同的数，能表示的最大十进制数是  $2^n-1$

①二进制转换为十进制：按权展开相加法

如  $(1101)_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 8 + 4 + 0 + 1 = 13$

②十进制转换为二进制：除二倒取余法

如  $173D = (10101101)_B$

方法如下：最后的余数从下往上取。

③二进制转换成十六进制：从二进制的低位开始，每四位二进制数转换成一位十六进制数。反之，每一位十六进制数转换成四位二进制数。

## 2. ASCII 码和汉字编码

计算机内的英文字符以二进制编码的形式来表示，其中使用最广的是 **ASCII 码**（即美国国家信息交换标准码）。标准的 ASCII 码用一个字节中的 7 个位来表示，可以表示 **128 种**编码，其中**数字、字母是按顺序依次排列的**。（ASCII 码表可参看书 P6）。汉字编码：汉字在计算机内也是以代码形式表示的，这些代码称为汉字编码。汉字的编码有输入码（智能、五笔等）、交换码、处理码、字型码。其中交换码又称为区位码。常见**汉字编码**有：**GB2312**（使用 2 个字节编码，包括一级汉字常用字和二级汉字非常用字）**一个 ASCII 码只占 1 个字节；一个汉字编码占 2 个字节**

## 3. 数据压缩

常见的多媒体数据压缩标准有：**静态图像压缩标准是 JPEG；动态图像（视频）压缩标准是 MPEG**[其中 **VCD 主要采用 MPEG-1 标准，数字电视、DVD 主要采用 MPEG-2 标准**]；**可视电话和电视会议的标准是 P\*64；音乐压缩标准是 MP3**。信息表达压缩的规范化是为了信息更好的交流。①为什么要压缩（目的）：为了**使存储容量更小**②为什么能压缩（原因）：数据本身存在着**冗余**（重复的数据），同时许多情况下媒体本身允许有少量失真。

多媒体数据中存在多种数据冗余：**空间冗余、视觉冗余、结构冗余、时间冗余**。

③**压缩方法分类**：**无损压缩**：借助压缩软件完成。（WinRAR、Winzip）使用者必须安装相应的压缩软件，俗称压缩包。**有损压缩**：保存成不同的扩展名文件实现数据压缩 如：bmp 变成 jpg，wav 变成 mp3，avi 变成 mpg 等

## 4. 文字识别与文字处理

汉字识别技术可分为印刷识别和手写体识别。

(1) **OCR 软件**：（功能：**识别图片中的文字**）

要对各种出版物中的文字进行编辑，可将其以**图片格式**保存在计算机中，利用 OCR 软件进行文字识别，转变为文本格式。有以下操作过程：

①**扫描件的获取**：用**扫描仪**扫描稿件，获取图像；

对图像进行预处理（可选），以提高识别的正确率。如去噪音、倾斜校正。

②**识别与转换**：

用 OCR 软件识别打开图像文件；（图像文件扩展名一般为 **bmp、jpg、tif** 等，不支持 gif）



(1) 图片包括图形和图像。图形是真实物体的模型化、抽象化和线条化的表现方式。而图像则真实地再现了一个物体的原形。

## (2) 位图与矢量图

(3) 图像文件常见扩展名：BMP、TIF、GIF 和 JPG。JPG 格式使用有损压缩，GIF 格式使用无损压缩，BMP 格式不压缩。**GIF 和 JPG 格式的文件由于使用了图像数据压缩技术，容量较少，所以是网页上常用的两种图像格式。**

(4) 黑白位图只有黑、白两种颜色，每个像素只需要一个位就可以表示；  
RGB 三原色，每一个像素的颜色值由 24 个数据位表示（即 24bit 或 3 个字节）  
RGB (0, 0, 0) 表示黑色；RGB (255, 255, 255) 表示白色；RGB (255, 0, 0) 表示红色，RGB (0, 255, 255) 表示由绿色和蓝色混合而成的青色。

**\* (5) 位图图像存储空间的计算公式是：水平像素\*垂直像素\*每个像素色彩所需位数/8（字节）**比如一幅 800\*600 的黑白图像，计算式子为 800\*600\*1/8（字节）。

## 8. 视频

(1) 视频是由连续的画面组成动态图像的一种表示方式。当图像以每秒 24 帧以上的速度播放时，由于人眼视觉的暂留，看到的就是连续变化的动作效果。

(2) 视频是由帧组成的。我国使用 PAL 制，每秒显示 25 帧；欧美国家使用 NTSC 制，每秒显示 30 帧。

(3) 目前被广泛使用的视频数据格式主要有：AVI、MPEG、MOV、DAT 和 Real 等格式。常见的视频素材处理软件有**豪杰超级解霸、会声会影和 Premiere** 等。

(4) 数字高清晰度电视是一种高标准的数字电视，其突出优点：图像质量高，节目容量大(节目容量可超过 500 套)，伴音质量好。

## 9. 网络知识点

考点 1：计算机网络的功能与组成

(一) 计算机网络的功能

数据通信、资源共享与分布处理。

(二) 计算机网络的组成

计算机网络=资源子网+通信子网

考点 2：计算机网络传输介质与连接设备

(一) 传输介质

无线介质：无线电波、红外线、微波、激光等。

有线介质：双绞线、同轴电缆、光缆等。

(二) 连接设备

考点 3：网络应用软件结构 B/S、C/S

B/S 结构：浏览器/服务器 (Browser/Server) 使用该结构，可以直接在浏览器中操作。

C/S 结构：客户机/服务器 (Client/Server) 结构，使用该结构，需要安装客户端软件。

考点 4：计算机网络拓扑结构及特点

拓扑结构	特点	结构图
<b>总线型网络拓扑结构</b>	1.各计算机地位平等 2.无中心控制节点 3.易于扩展、安装 4.费用低 5.局部节点出故障不影响整体	
<b>环型网络拓扑结构</b>	1.实时性好 2.可扩充性差 3.可靠性差，局部节点故障可导致全网瘫痪，故障检测困难	
<b>星型网络拓扑结构</b>	1.结构简单、易维护、扩充 2.电缆成本高 3.中心节点出故障，导致全网瘫痪	
<b>树型网络拓扑结构</b>	1.易于扩展，故障易于隔离、可靠性高 2.电缆成本高 3.对根节点依赖大	
<b>网型网络拓扑结构</b>	1.可靠性高 2.不易维护，线路成本高 3.性能好，路径选择复杂	

考点 5: TCP/IP 参考模型和 OSI 模型对比

OSI 参考模型	设备	传输数据单位	TCP/IP 参考模型	协议
应用层	网关		应用层	HTTP、FTP、Telnet、SMTP、POP3、DNS
表示层				
会话层				
传输层		数据段	传输层	TCP、UDP
网络层	路由器	数据包	网络层	IP、ARP、RARP、ICMP
数据链路层	交换机、网桥	帧	网络接口层	
物理层	中继器、集线器	比特		

考点 6: TCP/IP 协议组

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。

如要下载或阅读全文，请访问：

<https://d.book118.com/365123311114011300>