

必修一 第1章 认识数据与大数据

一、重难点知识总结

✚ 数据的定义

- 1) 数据是对事物的符号记录, 描述了不同事物的特征, 是信息的载体。
- 2) 在计算机科学中, 数据是指所有能输入到计算机并被计算机程序处理的符号总称, 其表现形式可以是文字、图形、图像、音频、视频等。

✚ 数据的特征

- 1) 二进制: 在计算机中, 数据是以二进制的形式存储、加工。
- 2) 语义性: 数据可用于描述事物, 形式是多样的、可听的、可看到的、可感知的。
- 3) 加工处理: 数据是可加工, 可处理的。数据是信息的载体。

✚ 信息及信息的特征

- 1) 通俗的说信息是人们由客观事物得到的, 是用语言、文字、声音、图形、图像、视频动画、纸张等形式作为载体, 通过各种途径传播的内容。
- 2) 信息的特征: 信息具有载体依附性和可处理性、传递性、时效性、共享性, 价值性、价值相对性、真伪性等特征。

✚ 知识的概念

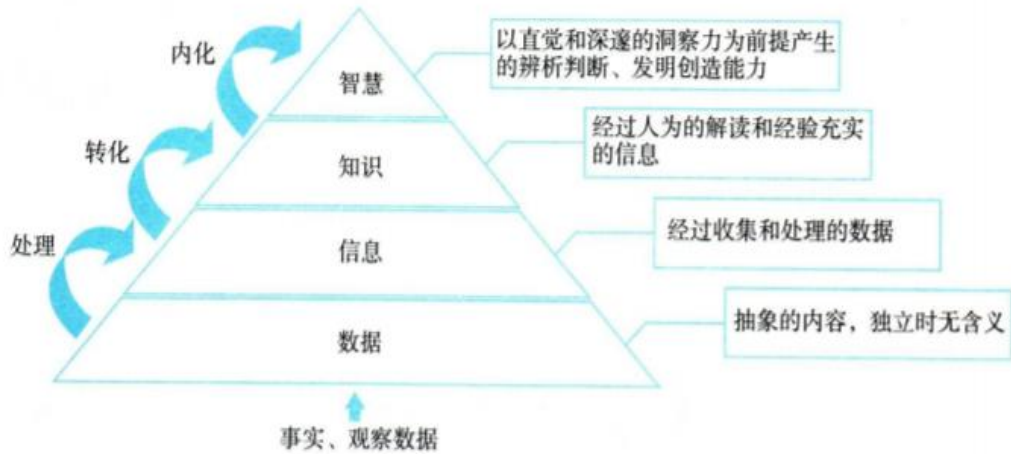
- (1) 知识是人类在社会实践中所获得的认识和经验的总和, 也是人类在实践中认识客观世界(包括人类自身)的成果, 它包括对事实、信息的描述及在教育 and 实践中获得的技能。
- (2) 知识是可以积累与传承的。

✚ 智慧的解释

- 1) 智慧是一种更高层次的综合能力, 主要表现为收集、加工、应用、传播知识的能力以及对事物发展的前瞻性看法。它是在知识的基础之上, 通过认识的累积, 而形成的事物的认识、远见, 体现为一种卓越的判断力。
- 2) 智慧是一种更高层次的综合能力。

✚ 数据、信息与知识、智慧的关系

- 1) 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的。
- 2) 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。
- 3) 人们综合运用知识和技能解决问题、发明创造或预测未来时, 知识就升华为智慧。
- 4) 数据、信息与知识的关系可以通过图表示。



通信系统中有两种主要的信息传输形式：一种是模拟信号；另一种是数字信号，它们在传输过程中是可以相互转换的

进制转换（字母 D 表示十进制，字母 B 表示二进制，字母 O 表示八进制，字母 H 表示 16 进制）

二进制：0, 1 八进制：0, 1, 2, 3, 4, 5, 6, 7

十进制：0, 1, 2, 3, 4, 5, 6, 7, 8, 9 十六进制：0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

十进制数转化为 R 进制数（R 为二进制，八进制，十六进制）：除 R 反向取余法。

R 进制转化为十进制数：按权展开求和法（二进制权为 2，八进制权为 8，十六进制权为 16）

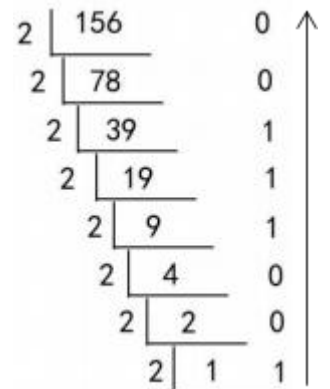
例：十进制 156 转化为二进制除 2 反向取余

$156D=100111000B$

二进制→十进制： $1011B=1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 11D$

二进制→八进制：（三位二进制数为一位八进制数） $101B=5O$ （O 是大写字母）

二进制→十六进制：（四个二进制数为一位十六进制数） $1001B=9H$



计算机中存储信息的单位

计算机中存储一个“0”或“1”占用 1 个二进制位（bit）。

比特（bit）是计算机中能处理或存储的最小单位；

字节（byte）：8 位为一个字节， $1\text{byte}=8\text{bit}$ ，字节是计算机信息存储的基本单位。（字节还可以用 B 表示）

单位换算关系： $1\text{TB}=1024\text{GB}$ ， $1\text{GB}=1024\text{MB}$ ， $1\text{MB}=1024\text{KB}$ ， $1\text{KB}=1024\text{B}$ ， $1\text{B}=8$ 位（bit）

数据编码

1. 字符编码

文字编码是指采用编码的方式将文字用“0”和“1”来表示。

常见的字符编码：

(1) ASCII 码: 英文字母 A 的编码为 01000001 (十进制数 65) 和 a 的编码为 01100001 (十进制数 97), 后面的字母依次往上加 1。0 的 ASCII 码十进制 48; 1 的 ASCII 码十进制 49。

(2) 汉字编码: GB2312 采用两个字节表示一个汉字。

(3) 条形码、二维码

条形码是将宽度不等的多个黑条和白条, 按照一定的编码规则排列, 用以表达一组信息的图形标识符。

二维条码/二维码是用某种特定的几何图形按一定规律在平面上(二维方向) 分布的黑白相间的图形记录数据符号信息。

条形码与二维码中的信息可以由专用设备扫描获取, 也可由智能手机扫描获取。

2. 声音编码

声音编码是按照一定的时间间隔采集声波的振幅, 并将其转换为二进制数序列, 即通过采样、量化和编码。

通常, 音频所占的存储容量取决于采样频率、量化位数、声道数和时长, 其计算公式为:

音频所占的存储容量=采样频率(Hz)*量化位数*声道数*时长/8 (Byte)

3. 图像编码

图像采样是从空间域上对信号进行分割, 分解出来的点称为像素, 并将每个像素的颜色量化为具体的数据。

像素: 是指在由一个数字序列表示的图像中的一个最小单位。

图像尺寸可以用像素个数来表示, 即“水平像素数”*“垂直像素数”

图像所占的存储容量=水平像素数*垂直像素数*量化位数/8 (Byte)

图像文件有多种存储格式, 常见的有 JPG/JPEG、PNG 和 BMP 等。

4. 视频编码

视频编码就是将视频信号经过视频采集卡转换成数字视频文件存储在数字载体——硬盘中。

视频画面是由以一定速度连续播放的一组静态图像组成的, 这些静态图像被称为帧。

常见的视频文件格式有 AVI、WMV 和 MP4 等。MPEG 是常见的视频压缩格式。

视频所占的存储容量=帧图像存储容量* 帧频* 时间 (Byte)

大数据的定义

大数据代表着信息量大、速度快种类繁多的信息资产, 需要特定的技术和分析方法将其转化为价值。

大数据的特征

大数据的特征可以用 4 个 V 来概括, 即数量(Volume)、速度(Velocity)、多样(Variety)和

价值(Value)。

- (1) 巨量性：大数据的数据量庞大,其存储单位需要用到 EB ,ZB 级别,且时刻在增长中。
- (2) 迅变性：速度快不仅指数据产生的速度快,还指数据处理的速度快。
- (3) 多样性：大数据来源多样,其数据往往多种形式共存。
- (4) 价值性：尽管大数据有很高的价值,但与其海量的数据量相比,则显得密度较低。

✚ 大数据思维

- (1) 大数据采用全样本数据分析。
- (2) 大数据不再追求个别数据的精确性,允许数据混杂。
- (3) 不一定强调事物之间的因果关系。

✚ 大数据技术

大数据技术指对巨量数据资源进行采集、提取、存储、分析和表达的技术,一般包括大数据采集、大数据预处理、大数据存储与管理、大数据分析与挖掘及大数据可视化 与应用等技术。

✚ 大数据对社会的影响

- (1) 大数据已经渗透到社会各个层面,不断提高工作效率与生活便利性。
- (2) 大数据让决策更精准。
- (3) 大数据带来了新的就业机会。
- (4) 大数据带来信息泄露、数据安全、个人隐私甚至伦理道德等方面的社会问题。

二、练习巩固

1、以下选项中属于知识的是 (B)

- A. 37 B. 气温 37°C, 夏天, 酷热
C. 气温 37°C D. 气温 37°C, 夏天, 酷热, 采取防暑降温措施

2、以下关于数据、信息和知识相互关系的理解, 不正确的是(B)。

- A. 数据是原始事物现象的符号记录
B. 信息等同于知识
C. 信息是经加工处理的、具有意义的数据
D. 知识是人们运用大脑对获取或积累的信息进行系统化地提炼、研究和分析的结果

3、一段时长为 10S 的 wav 单声道音频文件, 采样频率是 44.1khz, 量化位数为 16, 它所占的存储容量约为 (A)

- A. 861KB B. 6891B C. 1.68MB D. 6.89B

必修一 第2章 算法与程序实现

重难点知识总结

（一）解决问题的一般过程和用计算机解决问题

1. 解决问题的一般过程

分析问题——寻找解决问题的方法与途径——解决问题并验证结果

2. 用计算机解决问题的过程

提出问题——分析问题——设计方案——编程调试——解决问题

（二）算法的相关概念

1. 算法的基本概念

广义地讲，“算法”是指解决问题或完成任务的一系列步骤。在计算机科学领域，“算法”指的是用计算机解决问题的步骤，是为了解决问题而需要让计算机有序执行的、无歧义的、有限步骤的集合。

2. 算法的特征

（1）有穷性：处理步骤必须是有限的。

（2）可行性：每一步的操作与要求都是可行的，并且能在有限的时间内完成。

（3）确定性：每一步的执行描述必须是明确的。

（4）有0个或多个输入：算法被执行时，可以从外部获取数据，也可以包含在算法中。

（5）有1个或多个输出：算法必须有问题求解的结果，包含至少一个输出。

3. 算法的要素

（1）数据：用算法解决问题时，必须明确参与运算的初始数据、运算时产生的中间数据以及代表问题解决的结果数据。

（2）运算：在对数据进行运算时，必须明确每一步的运算是什么、对哪些数据进行运算等。

（3）控制转移：在算法执行过程中，有时需要根据数据或运算结果的特点进行不同的处理，需要运用控制转移来执行不同的操作。

4. 算法描述的常见方式

常见的算法描述方式有自然语言、流程图、伪代码、计算机程序设计语言等。

其中流程图描述算法直观易读、问题解决的步骤清晰简洁，更适合初学算法的人员使用。

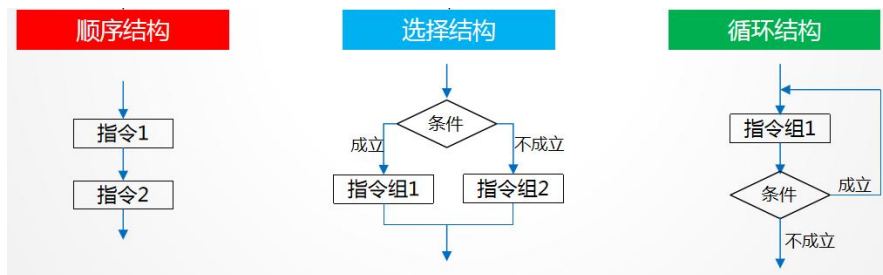
图形	名称	功能
	开始 / 结束符	表示算法的开始或结束
	输入 / 输出框	表示算法中数据的输入或输出
	处理框	表示算法中数据的运算处理
	判断框	表示算法中的条件判断
	流程线	表示算法中的流向
	连接点	表示算法中的转接

5. 算法的三种结构

(1) 顺序结构：在算法执行流程中，执行完一个处理步骤后，按照次序执行下一个步骤。顺序结构的每个步骤按照算法中出现的顺序依次执行；每个步骤一定会被执行一次，而且只执行一次。

(2) 分支结构：也称为选择结构。在算法执行流程中，先进行条件判断，再根据判断结果分别执行不同处理的控制结构。分支结构在执行时首先进行条件判断，根据条件满足与否来决定执行哪个分支。在一个分支结构中，必定有一个分支被执行，其他的分支则被忽略。

(3) 循环结构：算法执行过程中，在条件控制下，某些操作步骤需要重复执行（循环）的控制结构。循环结构的重复执行（循环）是在条件控制下的一种可控的重复，当需要重复处理的条件不满足（或满足）时，重复处理必须能及时结束。



(三) 程序设计基础知识

Python 中有很多内置模块。模块是类或函数的集合，用于处理一类问题，Python 可以使用很多已有模块，比如数学库 `math`、绘图库 `turtle` 等。

使用任何 Python 模块都必须先导入。模块的导入使用 `import` 语句。

`import math` 或 `from math import *`

2. python 的基础语句

(1) 数据类型

大类	数据类型	类型标识符	说明
数字	整型	int	整数类型理论上可以无限大
	浮点型	float	两种表示方式：小数形式或指数形式，比如 3.256 或 2.13e6、2.13E+6
	复数型	complex	复数的虚部以小写字母 j 或者大写字母 J 作为后缀。比如:5+3j
	布尔型	bool	逻辑值，真 (True) 或假 (False)
序列	字符串型	str	用单引号或双引号引起的一串字符。“red”、“你好”
	列表	list	一组数据的序列。用中括号 [] 括起来，数据元素用逗号分隔。比如：color=[“red”，“green”，“yellow”] 列表可以利用内置函数增加和删除数据。
	元组	tuple	一组数据的序列。用小括号 () 括起来，数据元素用逗号分隔。比如： color=(“red”，“green”，“yellow”) 元组元组是不可变序列，其中元素不可修改
映射	字典	dictionary	一组映射的集合。用大括号 {} 括起来，通过键(key)实现元素存取，每种映射映射之间用逗号分隔。 {key1:value1, key2:value2, ...} 比如： score={“math” :95, “english”: 98, “chi” :96}

(2) 输入和输出

输入语句的语法：input (“提示内容”)

input 用来接收键盘输入的内容。一般和赋值一起使用。

比如：x=input (“请输入你要计算的数字”)

输出语句的语法：print (输出内容 1, 输出内容 2……)

比如：print (“计算结果是：”, x**2)

提示：输入的内容默认是字符串类型，若需要算术运算则应进行类型转换。

语法：待转换类型（待转换内容）比如：`y=int(input("请输入你的出生年份："))`

(3) 常量、变量和表达式

常量：程序运行中，值不能被改变的量。

变量：程序运行中，值可以被改变的量。

变量命名时需要遵循一些基本规则：

1-变量名只能由数字,字母和下划线（英文）三种字符构成；

2-变量名不能用数字开头；

3-变量名区分大小写；

4-变量名不能和保留字相同；

这里特别需要指出两点：

1-Python 变量名支持中文字符，但考虑兼容性一般不建议使用；

2-Python 共有 35 个保留字（见下图），这些不可以被作为变量名使用。

and	elif	import	raise	global
as	else	in	return	nonlocal
assert	except	is	try	True
break	finally	lambda	while	False
class	for	not	with	None
continue	from	or	yield	async
def	if	pass	del	await

表达式：使用运算符连接的式子叫表达式。

运算类型	运算符	备注
算术运算	加减乘除：+ - * / 整除：// 求余：% 乘方：**	基本的算术运算
赋值运算	=	将右侧的值赋给左侧
关系运算	> < >= <= == !=	值为 bool 型 支持连写 a>b>c
逻辑运算	逻辑与：and 逻辑或：or 逻辑非：not	值为 bool 型 注意运算规则

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/368101034065007035>