国家教材委员会专家委员会审核通过
2019

普通高中教科书

信息技术

必修 1

数据与计算

Shuju yu Jisuan



微信公众号：电子课本大全

广东教育出版社

前　言

信息技术作为当今先进生产力的代表，已经成为我国经济发展的重要支柱和网络强国的战略支撑。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。自电子计算机问世以来，信息技术沿着以计算机为核心、到以互联网为核心、再到以数据为核心的发展脉络，深刻影响着社会的经济结构和生产方式，加快了全球范围内的知识更新和技术创新，推动了社会信息化、智能化的建设与发展，催生出现实空间与虚拟空间并存的信息社会，并逐步构建出智慧社会。

信息技术与社会的交互融合引发了数据量的迅猛增长，数据对社会生产和人们生活的影响日益凸显。本教科书针对数据（包括大数据）在信息社会中的重要价值，分析数据与信息的关系，强调数据处理的基本方法与技能，发展同学们利用信息技术解决问题的能力。本教科书是信息技术课程后续学习的基础。

通过对本教科书的学习，同学们能认识到数据在信息社会中的重要价值，合理处理与应用数据，掌握算法与程序设计的基本知识，根据需要运用数字化工具解决生活与学习中的问题，认识到人工智能在信息社会中越来越重要的促进作用，逐步成为信息社会的积极参与者。

本教科书按“数据与信息”“数据处理与应用”“算法与程序实现”三部分内容展开，围绕信息技术学科核心素养，设计了“体验庆祝国庆多媒体作品的数据与信息处理”“运用数字化工具探究数理知识”“设计从A市到B市耗时最少的旅行路线方案”“设计购买纪念品的最佳方案”“网络购物平台客户行为数据分析和可视化表达”“剖析空调企业智能客服机器人”项目范例。教师围绕“情境→主题→规划→探究→实施→成果→评价”的项目范例主线开展教学活动，帮助同学们掌握本教科书的基础知识、方法与技能，增强信息意识，发展计算思维，提高数字化学习与创新能力，培养正确的信息社会价值观和责任感，从而促进同学们的信息素养提升。

本教科书要求同学们对现实世界中的真实性问题进行自主、协作、探究

微信公众号：电子课本大全

学习。同学们围绕“项目选题→项目规划→方案交流→探究活动→项目实施→成果交流→活动评价”的项目学习主线开展学习活动，体验“做中学、学中创、创中乐”的项目学习理念和“从实践入手、先学后教、先练后讲”的项目学习策略，将知识建构、技能培养与思维发展融入运用数字化工具解决问题和完成任务的过程中，从而促进信息意识、计算思维、数字化学习与创新、信息社会责任的信息技术学科核心素养达成。

本教科书设置了“项目范例”“项目选题”“项目规划”“方案交流”“探究活动”“项目实施”“成果交流”“活动评价”等学习栏目，指导同学们开展项目学习活动。其中，“项目范例”是教师通过“情境”“主题”“规划”“探究”“实施”“成果”“评价”等活动，引导同学们了解开展项目学习活动的全过程；“项目选题”是同学们从真实世界选择自己感兴趣的项目主题；“项目规划”是同学们根据项目选题，制订自己的项目方案；“方案交流”是同学们展示交流自己设计的项目方案，师生共同探究、完善其方案；“探究活动”是同学们通过“问题”“观察”“分析”“阅读”“思考”“交流”“实践”“实验”“体验”“调查”“讨论”“拓展”等活动，获取知识和技能的过程；“项目实施”是同学们运用在项目学习过程中所获得的知识和技能来完成项目方案；“成果交流”是教师组织同学们展示交流项目成果，共享创造、分享快乐；“活动评价”是教师组织同学们开展项目评价活动。

本教科书各章首页的导言，叙述了本章的学习目的与方式、学习目标与内容，让同学们对整章有个总体认识。每章设置了“本章扼要回顾”，通过知识结构图把每章的主要内容及它们之间的关系描述出来，这有助于同学们建立自己的知识结构体系。每章结尾的“本章学业评价”设计了基于学业质量水平的测试题，并通过本章的项目活动评价，让同学们综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。此外，本教科书为同学们提供了配套学习资源包，里面含有网络商城、购票网站等虚拟网站，智能客服机器人虚拟仿真系统，以及各Python程序设计的源代码等，为同学们提供数据采集、分析和可视化表达所需的实验数据和环境。当然，同学们还可以自己收集素材，让自己的项目学习作品更有特色。

微信公众号：电子课本大全

CONTENTS

目录

第一章 数据与信息 1

项目范例 体验庆祝国庆多媒体作品的数据与信息处理 2

1.1 数据及其特征 5

1.1.1 数据 5

1.1.2 数据的基本特征 7

1.2 数据编码 7

1.2.1 模拟信号与数字信号 7

1.2.2 编码的基本方式 10

1.3 信息及其特征 16

1.3.1 信息 16

1.3.2 信息的基本特征 16

第二章 知识与数字化学习 21

项目范例 运用数字化工具探究数理知识 22

2.1 知识与智慧 25

2.1.1 知识 25

2.1.2 智慧 27

微信公众号：电子课本大全

2.2 数字化学习与创新 29
2.2.1 数字化工具与资源及其优势 29
2.2.2 体验数字化创新学习过程 33

第三章 算法基础 39

项目范例 设计从A市到B市耗时最少的旅行路线方案 40
3.1 体验计算机解决问题的过程 43
3.1.1 人工解决问题的过程 43
3.1.2 计算机解决问题的过程 46
3.2 算法及其描述 48
3.2.1 算法 48
3.2.2 算法的描述 49
3.3 计算机程序与程序设计语言 53
3.3.1 计算机程序 54
3.3.2 计算机程序设计语言 55

第四章 程序设计基础 63

项目范例 设计购买纪念品的最佳方案 64
4.1 程序设计语言的基础知识 66
4.1.1 Python的常量和变量 67
4.1.2 Python的数据类型 68
4.1.3 Python的运算符与表达式 69
4.1.4 Python的函数 71
4.2 运用顺序结构描述问题求解过程 73
4.2.1 数据的输入与输出 73
4.2.2 顺序结构的应用 74

微信公众号：电子课本大全

4.3 运用选择结构描述问题求解过程 76
4.3.1 if语句的应用 76
4.3.2 if...elif...else语句的应用 80

4.4 运用循环结构描述问题求解过程 85
4.4.1 for循环的应用 85
4.4.2 while循环的应用 88
4.4.3 循环嵌套的应用 90
4.4.4 循环控制的应用 93

第五章 数据处理和可视化表达 97

项目范例 网络购物平台客户行为数据分析和可视化表达 98

5.1 认识大数据 101
5.1.1 大数据 101
5.1.2 大数据的特征 101
5.1.3 大数据对日常生活的影响 103

5.2 数据的采集 104
5.2.1 数据采集的方法和工具 104
5.2.2 数据的存储和保护 107

5.3 数据的分析 111
5.3.1 特征探索 112
5.3.2 关联分析 113
5.3.3 聚类分析 115
5.3.4 数据分类 116

5.4 数据的可视化表达 118
5.4.1 数据可视化表达的方式 119
5.4.2 数据可视化表达的工具 120

微信公众号：电子课本大全

第六章 人工智能及其应用 125

项目范例 剖析空调企业智能客服机器人 126

6.1 认识人工智能 129

6.1.1 人工智能 129

6.1.2 剖析领域知识智能问答机器人 130

6.1.3 人工智能的诞生与发展 132

6.2 人工智能的应用 134

6.2.1 智能制造 134

6.2.2 智能家居 136

6.2.3 智能教育 137

6.2.4 智能交通 137

6.2.5 智能安防 138

6.2.6 智能医疗 138

6.2.7 智能物流 140

附录1 部分术语、缩略语中英文对照表 145

附录2 项目活动评价表 146

微信公众号：电子课本大全

第一章
数据与信息

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。自电子计算机问世以来，信息技术沿着以个人计算机为核心、到以互联网为核心、再到以数据为核心的发展脉络，深刻影响着社会的经济结构和生产方式，加快了全球范围内的知识更新和技术创新。信息技术与社会的交互融合引发了数据量的迅速增长，数据对社会生产和人们生活的影响日益凸显，并具有重要价值。

本章将通过“多媒体作品的数据与信息处理”项目，进行自主、协作、探究学习，让同学们在具体感知数据与信息的基础上，描述数据与信息的特征，知道数据编码的基本方式，从而将知识建构、技能培养与思维发展融入运用数字化工具解决问题和完成任务的过程中，促进信息技术学科核心素养达成，完成项目学习目标。

- 数据及其特征
- 数据编码
- 信息及其特征

微信公众号：电子课本大全

项目范例　体验庆祝国庆多媒体作品的数据与信息处理

情境

自古以来，科学技术就以一种不可逆转、不可抗拒的力量推动着人类社会向前发展。纪录片《创新中国》集中呈现了党的十八大以来中国的最新科技成就，科技创新让人民生活更为便捷，让企业发展更具活力，让国家实力更加强大。2016年8月16日，世界首颗量子科学试验卫星“墨子号”发射升空；2016年9月25日，500米口径球面射电望远镜——“中国天眼”正式落成启用；2017年5月5日，装配超过20个系统、数百万个零件、攻克100多项核心关键技术的国产商用大飞机C919首飞成功；2019年1月3日，我国月球探测器“嫦娥四号”成功在月球背面着陆，这是人类探测器首次造访月球背面……一项项成就，都实现了一次次历史性的飞跃，创新科技铸造的中国正成为世界瞩目的国度（图1-1）。党的十九大让全世界都再次聚焦到中国新时代的宏伟蓝图上来。

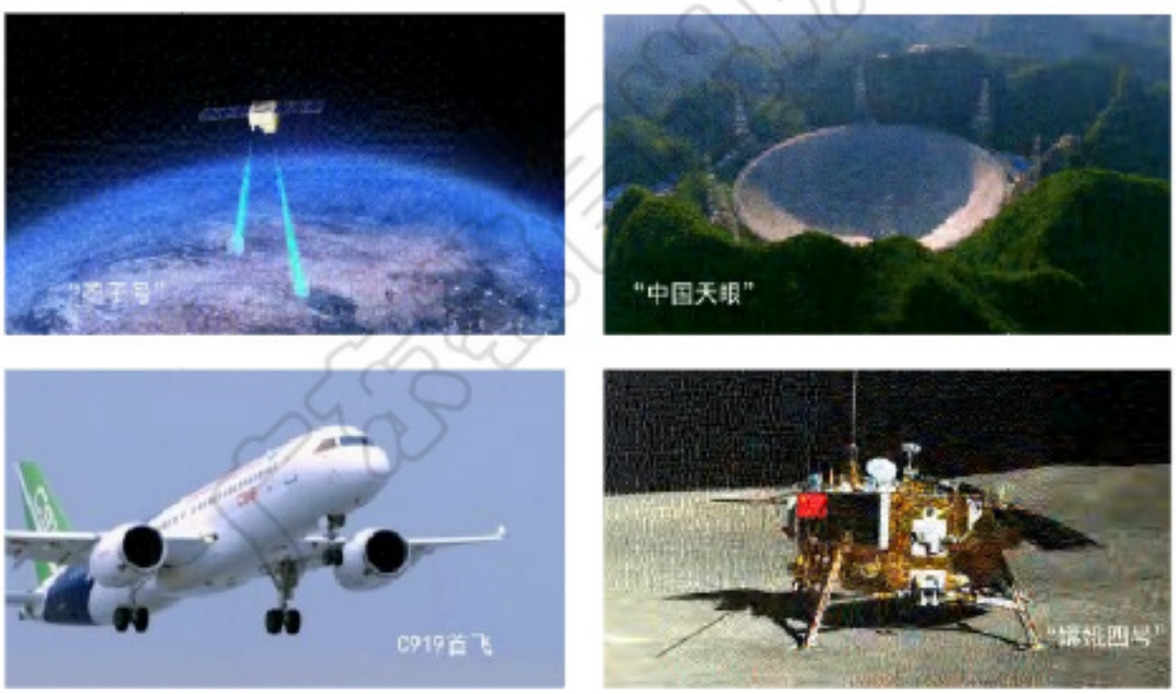


图1-1　创新科技铸造的中国正成为世界瞩目的国度

国庆节快到了，学校百灵创意社团通过网络、报纸和书籍进一步了解《创新中国》中介绍的伟大成就，并对获得成就的数据和信息进行相应的处理，创作庆祝国庆的多媒体作品。

主题

体验庆祝国庆多媒体作品的数据与信息处理

微信公众号：电子课本大全

规划

根据项目范例的主题，在小组中组织讨论，利用思维导图工具，制订项目范例的学习规划，如图1-2所示。

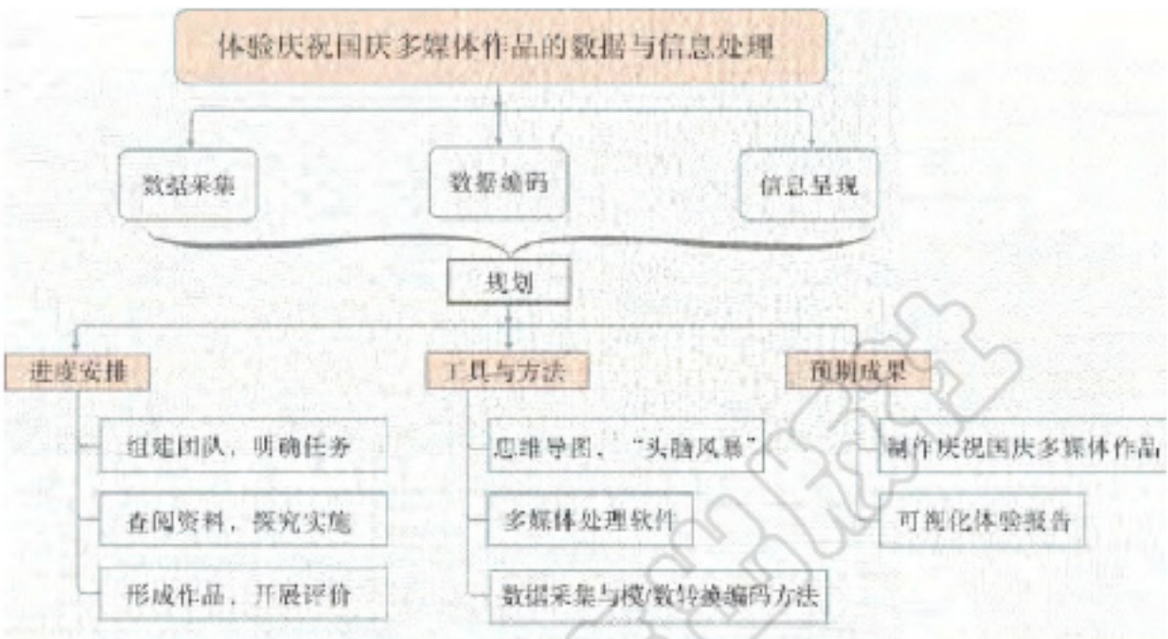


图1-2 “体验庆祝国庆多媒体作品的数据与信息处理”项目学习规划

探究

根据项目学习规划的安排，通过调查、案例分析、文献阅读和网上资料搜索，开展“体验庆祝国庆多媒体作品的数据与信息处理”项目学习探究活动，如表1-1所示。

表1-1 “体验庆祝国庆多媒体作品的数据与信息处理”项目学习探究活动

探究活动	学习内容		知识技能
数据采集	认识数据，了解各种数据的记录方式。	知道数据的概念，了解数据记录方式发展的基本历程，学会数据采集的基本方法和过程。	具体感知数据。 描述数据的基本特征。
	分析数据应用实例。	描述数据的特征。	
数据编码	比较模拟信号与数字信号。	知道模拟信号与数字信号的含义及其相互关系。	知道数据编码的基本方式。
	开展模拟信号的数字化实验活动。	体验数据编码的基本方法和过程，了解数据编码的作用。	
	列举常见数据的编码方式。	了解文本、图像等多种类型数据的编码方式。	

微信公众号：电子课本大全

（续表）

探究活动	学习内容		知识技能
信息呈现	列举信息的不同定义，分析事例中的信息。	知道信息的概念、信息的价值，了解信息呈现的基本方法和过程。	具体感知信息。 描述信息的特征。
	分析信息应用实例。	描述信息的基本特征，分析现实生活事例的信息特征。	



实施项目学习各项探究活动，进一步体验庆祝国庆多媒体作品数据与信息的处理。



在小组开展项目范例学习过程中，利用思维导图工具梳理小组成员在“头脑风暴”活动中的观点，建立观点结构图，运用多媒体创作工具（如演示文稿、在线编辑工具等），综合加工和表达，形成项目范例可视化学习成果，并通过各种分享平台发布，共享创造、分享快乐。例如，运用在线编辑工具制作的“体验庆祝国庆多媒体作品的数据与信息处理”可视化报告，可以在教科书的配套学习资源包中查看，其目录截图如图1-3所示。



图1-3　“体验庆祝国庆多媒体作品的数据与信息处理”可视化报告目录截图

微信公众号：电子课本大全

评价

根据教科书附录2的“项目活动评价表”，对项目范例的学习过程和学习成果在小组或班级上进行交流，开展项目学习活动评价。

项目选题

同学们以3~6人组成一个小组，选择下面一个参考主题，或者自拟一个感兴趣的主题，开展项目学习。

1. 体验庆祝国庆录音作品的数据与信息处理
2. 体验庆祝国庆摄影作品的数据与信息处理
3. 体验庆祝国庆视频作品的数据与信息处理

项目规划

各小组根据项目选题，参照项目范例的样式，利用思维导图工具，制订相应的项目方案。

方案交流

各小组将完成的方案在全班进行展示交流，师生共同探讨、完善相应的项目方案。

1.1 数据及其特征

互联网、大数据时代，人们可以通过计算机和移动终端，随时随地获取信息、了解社会。例如，《创新中国》系列纪录片对收集的大量数据［包括图形、图像、视频、音频、文本（文字、数字、数值、字符）］进行存储、挖掘和分析后，以多媒体记录新闻事件，满足人们对信息直观化、可视化的需求。

1.1.1 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，人们常使用符号来表达客观事物。例如，180 cm、70 kg、36.5 ℃可以分别表示人体的身高、体重、体温情况，200 m^2、600 m^3、120 km/h可以分别表示物体的面积、体积、速度情况。世界任何客观事物在人脑中的反映（意识）离不开符号的表达和基

微信公众号：电子课本大全

于符号的逻辑推理。

在计算机科学中，数据是对所有输入计算机并被计算机识别、存储和处理的符号的总称，是联系现实世界和计算机世界的途径。即现实世界的各种符号，包括图形、图像、视频、音频、文本（文字、数字、数值、字符）等数值性和非数值性符号，都可以被计算机转化为二进制数（如数字“180”在计算机中表示为10110100，字符“cm”在计算机中表示为0110001101101101），进行各种逻辑运算，处理成有意义的数据；然后将计算机运算的二进制数转换成现实世界的视、听、触等各种感觉的信息，如图1–4所示。



（a）　人使用符号表达客观事物（如体育运动符号）



（b）　计算机用二进制数表达客观事物（如表情的部分二进制码）

图1–4

在大数据时代，数据不仅是信息的载体，也是人们提取信息、做出决策的重要依据，成为人们认识和理解现实世界客观事物的重要资源。

探究活动

体验

请同学们赏析庆祝国庆多媒体作品。图1–5是该作品的截图。

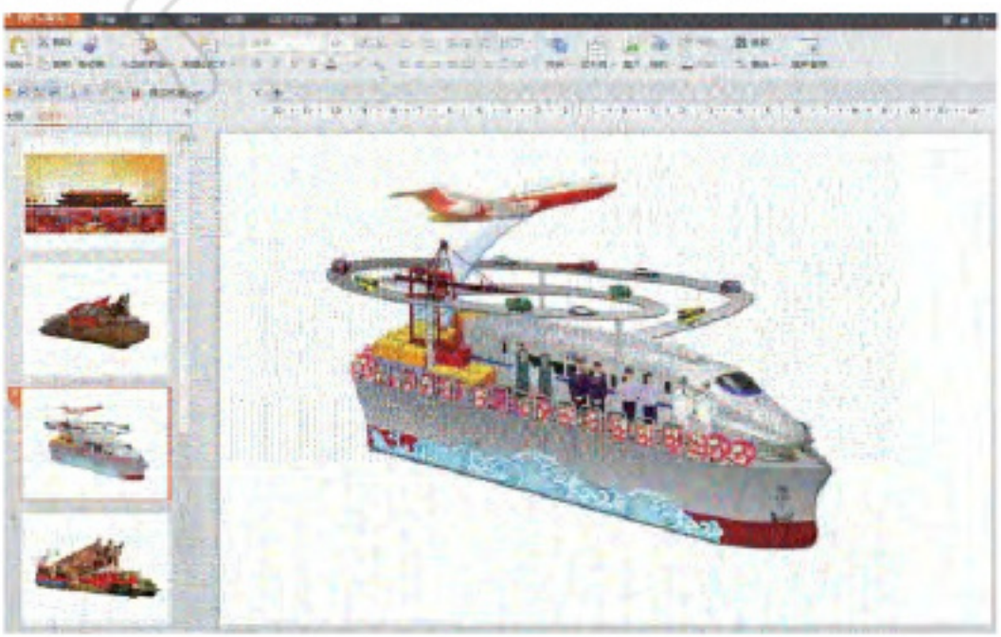


图1–5　庆祝国庆多媒体作品的截图

微信公众号：电子课本大全

1.1.2 数据的基本特征

人们在探索现实世界时，用数据记录自然现象与社会的发展。寻找数据的基本特征和规律，就是探索与揭示自然和社会的发展规律。在计算机科学中，数据具有如下基本特征：

（1）二进制。在计算机中，数据以二进制的形式存储、加工。

（2）语义性。语义是将数据符号解释为客观世界的事物。

（3）分散性。数据是分散的记录，分别记录不同客观事物的运动状态。

（4）多样性与感知性。数据记录的形式是多样的、可看的、可听的、可感知的，如图形、图像、视频、音频、文本等。

讨论

以小组为单位，查阅并分享体现数据基本特征的例子，同时讨论数据还有哪些特征。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，开展以下活动。

1. 完成相应作品录音、录像或扫描图像等数据的采集和分析。
2. 总结与记录数据采集和分析的方法。

1.2 数据编码

将现实世界的事物现象符号化、数据化，需要有一个理解、抽象、推理的过程，这个过程由计算机来完成。计算机将它们处理转换成0和1组成的二进制编码，进而采用基于二进制的算术运算和逻辑运算进行处理。

1.2.1 模拟信号与数字信号

众所周知，声、光信号是人的感官所能识别和接收的信号，而计算机是一种采用电信号进行运作的机器。要将人的感官所能识别的信号转换为计算机能够识别的信号，实现人

微信公众号：电子课本大全

机交互，就必须清楚计算机是如何处理这些信号的。

1. 模拟信号

模拟信号是指用连续变化的物理量所表达的信息。其信号的幅度、频率或相位随时间作连续变化，如声音信号、图形信号等。

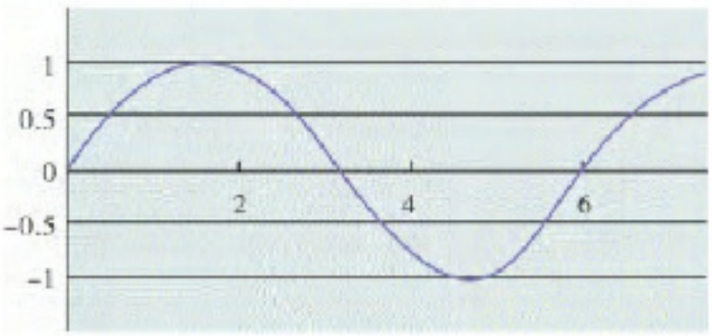


图1-6 单一频率波形示意图

模拟信号的波形可以是简单的，也可以是复杂的。例如，由单一钢琴定音器产生的声波，就是简单波形，仅仅包括一个单一频率（图1-6）；人类或管风琴的声音产生的声波，就是复杂波形，包含了许多不同频率的组合（图1-7）。

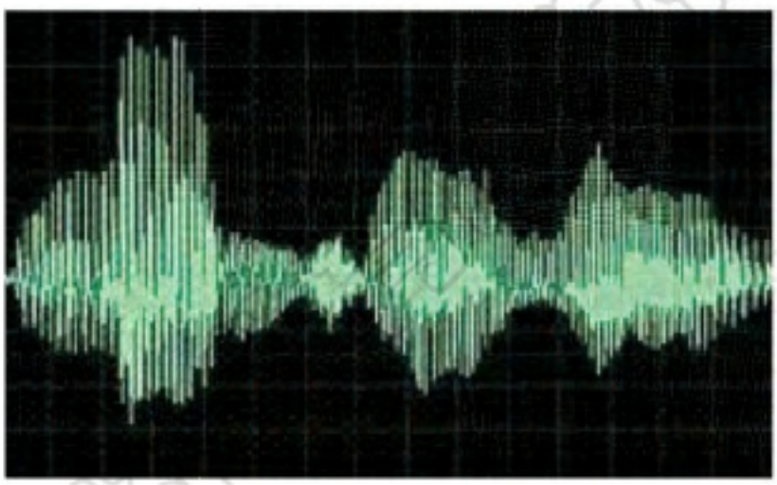


图1-7 多频率组合波形示意图

模拟信号是传导能量的一种方式。例如，声波通过空气等介质来向远处传送能量，在传播的过程中，能量会不断被损耗而逐渐衰减，如图1-8所示。

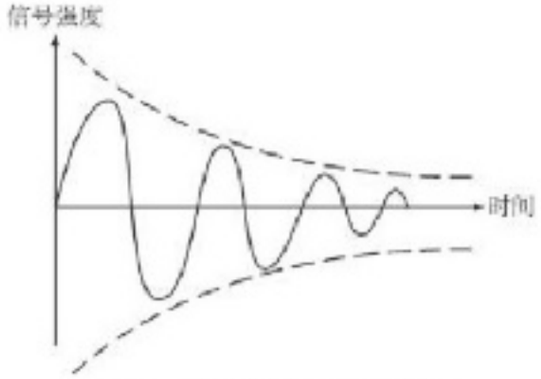


图1-8 模拟信号的衰减示意图

无论是有线相连的电话，还是无线发送的广播电视，都是通过模拟信号来传递信息的。

2. 数字信号

数字信号是离散时间信号的数字化表示。其信号的自变量、因变量都是离散的。例如，开关电路中输出电压和电流脉冲就是离散的信号，如图1-9所示。

微信公众号：电子课本大全

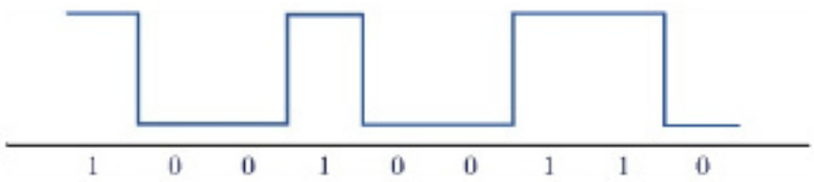


图1-9 离散的脉冲信号

在计算机中，数字信号的大小常用有限位的二进制数表示。例如，字长为2位的二进制数可表示00、01、10和11四种数字信号。

尽管基于数字信号的数据通信才出现几十年，但因其抵抗电路本身干扰和环境干扰的能力强，且数字信号有利于存储、加密与纠错，从而具有较强的保密性和可靠性。因此在现代技术的信号处理中，数字信号发挥的作用越来越大，覆盖的范围越来越广。

探究活动

分析

分析、比较模拟信号与数字信号在通信应用中的优劣，用简洁的文字将结论填入表1-2中。

表1-2 模拟信号与数字信号比较

名称	优点	缺点
模拟信号		
数字信号		
进行远距离传送时用哪种信号比较好？		

实验

在航海灯语系统中，灯语是一种通信手段，用灯光一明一暗的间歇做出长短不同的信号来传递信息。在古代，灯语对人们之间的交流起到很大的作用。随着科技的发展，灯语已经通过计算机网络实现了灯光信号的自动发送、获取和识别。

1. 实验名称：用手电筒传递信息。
2. 实验目的：了解灯语的使用场景，理解数据编码的必要性。

微信公众号：电子课本大全

3. 实验步骤：

（1）两人一组，一人控制手电筒开关，按照莫尔斯码发送一组字母，另一人观察手电筒发出的光信号，记录数据，并使用莫尔斯码译出这组字母。

（2）双方核对信息，多次重复实验。

（3）通过分析实验数据，总结该方法传递信息的利弊，提出改进方法。

（4）设计一套自己的灯语，用“密码”传送信息。

1.2.2 编码的基本方式

在现代技术的信号处理中，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经过编码进行存储和传输。

1. 文字编码

文字（字符）编码是效率相对较低的编码方式，有单字节码和双字节码两种。其中，ASCII码、莫尔斯码属于单字节码，国标码（GBK）、统一码（Unicode）属于双字节码。

（1）ASCII码。

ASCII码（American Standard Code for Information Interchange）是美国信息交换标准代码，用8位二进制码为所有的英文字母（大小写52个）、阿拉伯数字（10个）和常用的不可见控制符（33个）以及标点符号、运算符号等（33个）建立了转换码，将符号转换为“0”和“1”构成的编码。从表1-3可知，英文字母A和a的编码分别为01000001（十进制数65）和01100001（十进制数97）。

表1-3　ASCII码表

二进制	0000	0001	0010	0011	0100	0101	0110	0111
0000	NUL	DLE	空格	0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w

微信公众号：电子课本大全

（续表）

1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SOH	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL



在计算机中，英文字母、数字、符号等都可用ASCII码来表示，如大写字母“C”的编码为8位二进制数01000011，数字“9”的编码为00111001等。可是，中文却难以用ASCII码来表示。如果要设计一套汉字的编码系统，可以如何进行？仿照ASCII编码系统的方法，描述大致的设计思路，并在小组内交流。

仿照ASCII码，描述设计思路：

（2）国标码。

由于ASCII码只包含英文、数字和一些控制符号，并不能表示汉字，因此，我国设计了用于处理汉字的简体中文的GB码和用于繁体中文的BIG5码（大五码）。

1980年发布的《信息交换用汉字编码字符集》（GB 2312）一共收录了7445个字符，包括6763个汉字和682个其他符号。GB 2312主要分为两部分：编号127之前的符号跟ASCII码所表示的意义相同，属于单字节码；编号127之后的符号统一用两个字节表示，包含了几乎所有的简体中文字（中文编码高字节的最高位不为0）。

1995年公布的《汉字内码扩展规范》（GBK 1.0）收录了21 886个符号，分为汉字区和图形符号区，汉字区包括21 003个字符。

微信公众号：电子课本大全

2. 图像编码

图像编码是指在满足一定保真度的条件下，对图像数据进行变换、编码和压缩，以较少比特数表示图像或图像中所包含的信息的技术。

我们以位图（Bitmap）为例来说明图像的编码原理与方法，以及图像文件的大小计算方法。

（1）位图图像编码。

位图，最小单位为光栅点（或称像素），因而也叫作点阵图（或像素图）。

位图采用位映射存储格式，即将每一个像素映射为一个数据，存放在以字节为单位的矩阵中。例如，在如图1-10所示的黑白图像中，共有32×2个像素，如果将黑、白像素分别映射为1和0，就表示为11111111，00000110，10000110，11111111，10000001，01100011，01111100，10000001，这就是该黑白图像的二进制编码。通常图像编码采用16进制编码，因而图1-10的编码为“FF，06，86，FF，81，63，7C，81”。



图1-10　每一个像素只有两种可能性的黑白图像

（2）位图文件大小。

在计算机二进制数系统中，每个0或1就是一个位（bit，数据存储的最小单位），8个位就称为一个字节（Byte）。

黑白图像，每一个像素有2种可选颜色（黑、白），称为1位图像。因此，图1-10中所包含的图形数据为32×2÷8=8个字节。

16色图像，每一个像素有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）；24位图像的可选颜色更丰富，为2^{24}种。

实际上，一个位图文件除了包含图形数据，还包括文件头、位图信息头、颜色信息、图形数据等几部分。

①文件头：包含文件的类型、大小和位图起始位置等信息，共14个字节。

②位图信息头：用于说明位图的尺寸等信息，占40个字节。

③颜色信息：用于说明位图中的颜色，有若干个表项，每一个表项定义一种颜色。当图像量化位数为1、4、8时，分别有2、16、256种颜色，每个颜色表项占4个字节；当图像量化位数为24时，没有颜色表项。

④图形数据：记录位图的每一个像素值，其记录顺序先从左到右，再从下到上。

一般情况下，位图文件所占用的空间可按以下公式计算：

文件的大小=文件头+信息头+颜色表项+图像分辨率×图像量化位数÷8

其中，图像分辨率=图像x方向的像素数×图像y方向的像素数。图像量化位数，也叫图像深度，是指图像中每个像素点记录颜色所用二进制数的位数。

微信公众号：电子课本大全

3. 声音编码

对声音进行数据编码，必须经过前期的数据采样和数据量化。

（1）采样。采样就是把输入的模拟信号按适当的时间间隔得到各个时刻的样本值，使其转换为时间上离散、幅度上连续的脉冲信号。

根据奈奎斯特（Harry Nyquist，物理学家，1889—1976）采样定理，如果以一定时间间隔对某个信号$f(t)$进行采样，并且采样频率高于该信号最高频率的两倍，则采样值包含了原信号的全部信息，如图1-11所示。

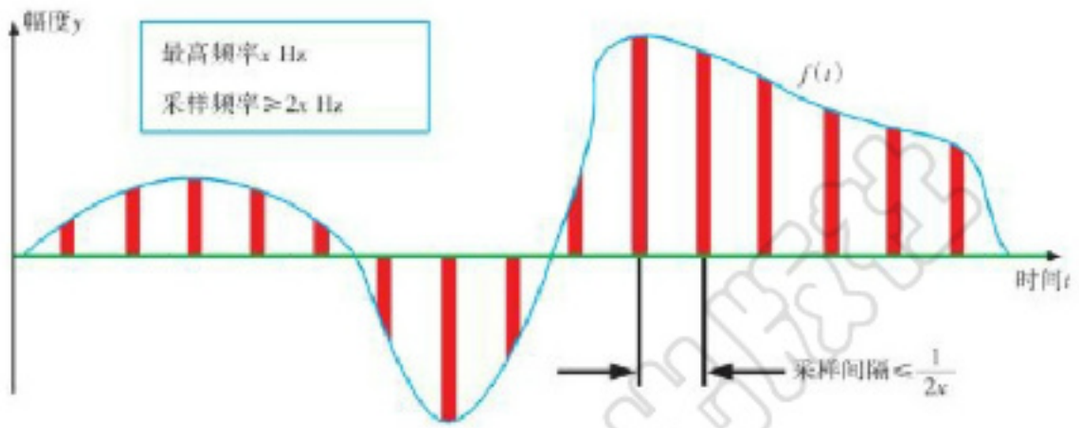


图1-11　奈奎斯特采样定理示意图

对于音频信号，常用的采样频率有三种：44.1 kHz、22.05 kHz和11.025 kHz。

（2）量化。量化是把样值信号的无限多个可能的取值，近似地用有限个数的数值来表示。首先是将采样信号幅度划分为若干量化等级（国标声音量化等级分为256个，即2^8个），然后将采样后的信号幅度与所划分的各个量化等级进行比较，向下取最接近的量化等级的数值，如图1-12所示。

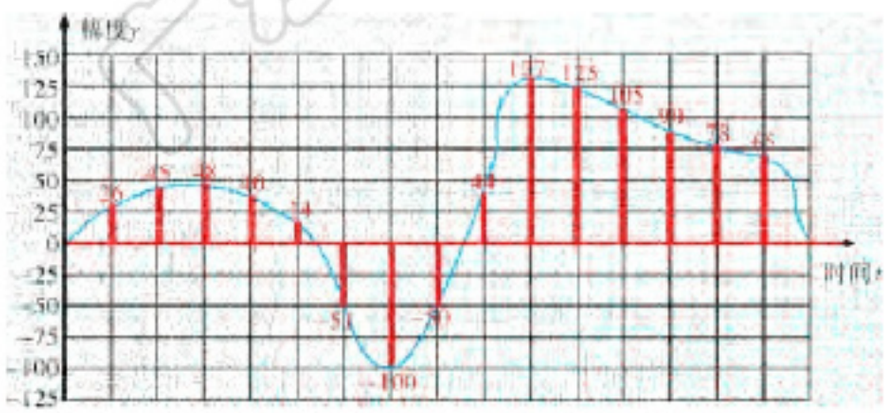


图1-12　脉冲幅度调制信号的量化

（3）编码。编码是将量化后的采样值用二进制数码表示，并转换为由二进制编码0和1组成的数字信号。模拟信号采样后可用8位二进制数表示，最高位表示符号，正数为0，负数为1。例如，采样值48的二进制编码为00110000，-100的二进制编码为11100100，如表1-4所示。

微信公众号：电子课本大全

表1-4　声音信号的数据编码

模拟量	26	45	48	40	14
编码	00011010	00101101	00110000	00101000	000010110
模拟量	-51	-100	-50	44	127
编码	10110011	11100100	10110010	00101100	01111111
模拟量	125	105	90	78	65
编码	01111101	01101001	01011010	01001110	01000001

编码时采用的二进制位数越多，数据量越大，占用的存储空间也越大。以声音信息为例，其存储空间遵循如下公式：

声音存储空间=采样频率×量化位数×声道数×时间÷8

例如，采样频率为44.1 kHz、量化位数为16位的立体声，1秒声音所需字节数为44.1×1000×16×2×1÷8=176 400（B）。

编码后的信号可以实现无差错的数据传输，如图1-13所示。



图1-13　数字信号的传送



各小组讨论如何减少声音信号还原时的失真，提高声音的保真度。提出改进方案后，继续实验探索，直到得出结论。



1. 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。
2. 若将上述图像另存为256色位图图像，则文件大小为多大？
3. 试上网查询有关视频编码的知识，了解视频是如何被编码的。

实际上，不仅文字、图像、声音等类型的数据可以被编码，视频等数据也都可以被编码成数据文件。

程序设计，简称编程，也属于编码，是运用程序设计语言，实现人机交互，提高计算机应用效率，为实现操作程序化要求所进行的编码。有关程序设计的内容将在后续章节中学习。总之，编码是数据存储及其表现的基础，也是人们认识世界的基础。

微信公众号：电子课本大全

拓展

布尔代数与图灵机

现实世界的任何事物，如果用计算系统进行识别、存储和计算，首先需要将其符号化。所谓符号化，是指将现实世界的任何事物用符号表达，进而进行基于符号的计算的一种思维方式。将事物表达为不同的符号，便可采用不同的工具（或数学方法）进行计算；将符号赋予不同事物的意义，则能计算不同的现实世界问题。

所谓逻辑，是指事物因果之间所遵循的规律，是现实中普遍的思维方式。布尔在汲取亚里士多德和莱布尼茨形式逻辑符号化思想的基础上，创立了用0和1表达、处理现实中命题判断与推理（1表示真/0表示假）的基于二进制逻辑的代数系统，为电子计算机的开关电路设计提供了重要的数学方法和理论基础。

二进制由0和1两个数码组成，逢2进1，借1当2，第i数位的权值为2^i。之所以青睐二进制，是因为：①二进制算术运算规则简单；②二进制算术运算可以与逻辑运算实现统一，或者说可以用逻辑运算实现算术运算；③能表示两种状态的元器件容易找到，如继电器开关、灯泡、二极管或三极管等。因此，计算机硬件存储和处理的是二进制数。

20世纪30年代，图灵提出了图灵机模型，直观且形象地说明了通用计算机器的工作机理，建立了指令、程序及通用机器执行程序的理论模型，奠定了计算理论的基础。图灵认为，所谓计算就是计算者（人或机器）对一条两端可无限延长的纸带上的一串0或1执行指令，一步一步地改变纸带上的0或1，经过有限步骤，最后得到一个满足预先规定的符号串的变换过程，如图1-14所示。数据被制成一串0和1的纸带送入机器中，作为输入，如0001000001000011…机器可对输入纸带执行的基本动作包括翻转0为1，或翻转1为0，前移一位，停止。机器对基本动作的执行是由指令来控制的，机器是按照指令的控制选择执行哪个动作，指令也可以用0和1来表示：01表示翻转0为1（当输入为1时不变），10表示翻转1为0（当输入0时不变），11表示前移一位，00表示停止。输入如何变为输出的控制可以用指令编写一个程序来完成，如01111011011011100。机器能够读取程序，按程序中的指令顺序读取指令，读一条指令执行一条指令，由此实现自动计算。因此可以说，图灵机就是一个最简单的计算机模型，图灵机将控制处理的规则用0和1表达，将待处理的信息及处理结果也用0和1表达，处理即是对0和1的变换，这样可以用机械或电子系统实现。

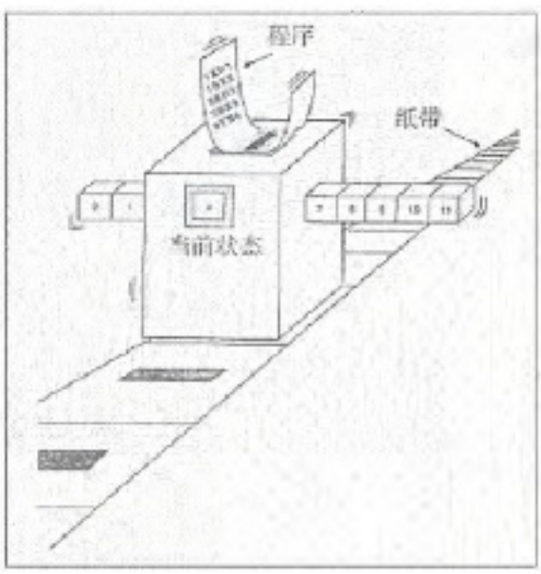


图1-14 图灵机模型

微信公众号：电子课本大全

项目实施

各小组根据项目选题及拟定的项目方案，结合本节所学知识，掌握数据的编码过程。

1. 分析常见编码方式与相应作品数据的编码过程。
2. 总结与记录数据编码的方法和步骤。

1.3 信息及其特征

在互联网时代，数据正以越来越快的速度不断增长。移动电话、社交媒体和用于医疗诊断的影像技术等新业务，每天都会产生大量新数据。从这些数据中分析并提炼出有用的信息，将为人们的日常生活带来便利，并对推进科技、生产和社会的发展有重大意义。

1.3.1 信息

信息是经过加工处理的、具有意义的数据。信息是对客观世界中各种事物的运动状态和变化的反映，是客观事物之间相互联系和相互作用的表征。人们通过获得、识别自然界和人类社会的不同信息来区别不同事物，得以认识和改造世界。

探究活动

交流

结合生活实际中的例子，在小组内交流信息和数据的区别与联系。

1.3.2 信息的基本特征

信息在人类社会活动中发挥着重要的作用，因而受到人们高度重视。那么，信息具有哪些特征？对人类活动又产生了什么影响？

（1）普遍性。信息不是虚无缥缈的事物，只要有事物运动，有数据存在，就会有信息存在。信息普遍存在于自然界和人类社会发展的始终，可以被人们感知、获取、传递和利用。

微信公众号：电子课本大全

（2）传递性。信息的传递打破了时间和空间的限制。例如，我们可以借助计算机、网络、手机等工具了解发生在世界各地的许多事情。

（3）共享性。信息作为一种资源，通过交流可以在不同个体或群体间共享。信息交流与实物交流有本质上的不同：实物交流，一方有所得，另一方必有所失。例如，两个人交换手中的一个苹果，每人得到的是对方的那一个苹果。而信息交流则不然，两个人交换则都拥有了两份信息。因此，人们悟出以下道理：分享、沟通与行动是将知识转化为智慧的关键。

（4）依附性和可处理性。各种信息必须依附一定的图像、文字、声音等符号才能够表现出来，为人们所接收，并按照某种需要进行处理和存储。信息经过分析和处理，往往会产生新的信息，使信息得到增值。

（5）时效性。信息不是一成不变的东西，会随着客观事物的变化而变化。如果信息不能反映事物的最新变化状态，它的效用就会降低，一旦超过其“生命周期”，信息就失去原有的效用，成为历史记录。

（6）真伪性。人们接收到的信息，并不都是对事物的真实反映，因此，信息具有真伪性。产生伪信息的原因很多：有些是出于某种目的，被人故意采用窜改、捏造、欺骗、夸大等手段人为制造的；有些是由于人们的认识能力或表达能力不足导致的；有些则是传递过程中的失误造成的。因而，对信息的真伪性需要加以鉴别。

（7）价值相对性。信息作为一种特殊的资源，具有相应的使用价值，能够满足人们某些方面的需要。但信息使用价值的大小是相对的，取决于接收信息者的需求及其对信息的理解、认识和利用的能力。

拓展

香农与信息论

关于文字、数字、图形、声音的知识已有几千年历史了。但是它们的总称是什么，如何统一地计量，直到19世纪末还没有被正确地提出来，更谈不上如何去解决了。20世纪初期，随着电报、电话、照片、电视、无线电、雷达等技术的发展，如何计量信号中信息量的问题被隐约地提上日程。

克劳德·艾尔伍德·香农（Claude Elwood Shannon，1916—2001）是美国数学家、信息论的创始人。他先后发表了论文《通信的数学原理》和《噪声下的通信》，系统论述了信息的定义、怎样数量化信息、怎样更好地对信息进行编码等通信的基本问题，给出了信息熵的概念和通信系统的模型，提出了信息量的数学表达式，并解决了信道容量、信源统计特性、信源编码、信道编码等一系列基本技术问题，为信息论和数字通信奠定了基础。香农被誉为“信息论之父”。

香农在引入熵（指体系的混乱的程度）的概念进行信息的定量计算的时候，明确地把信息量定义为随机不确定性程度的减少。这就表明了他对信息的理解，即信息是用来减少随机不定性的东西，或信息是确定性程度的增加。

微信公众号：电子课本大全

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，了解信息及其基本特征，进一步完善该项目方案中的各项学习活动，并参照项目范例的样式，撰写相应的项目成果报告。

成果交流

各小组运用数字化学习工具，将所完成的项目成果，在小组或班级上进行展示与交流，共享创造、分享快乐。

活动评价

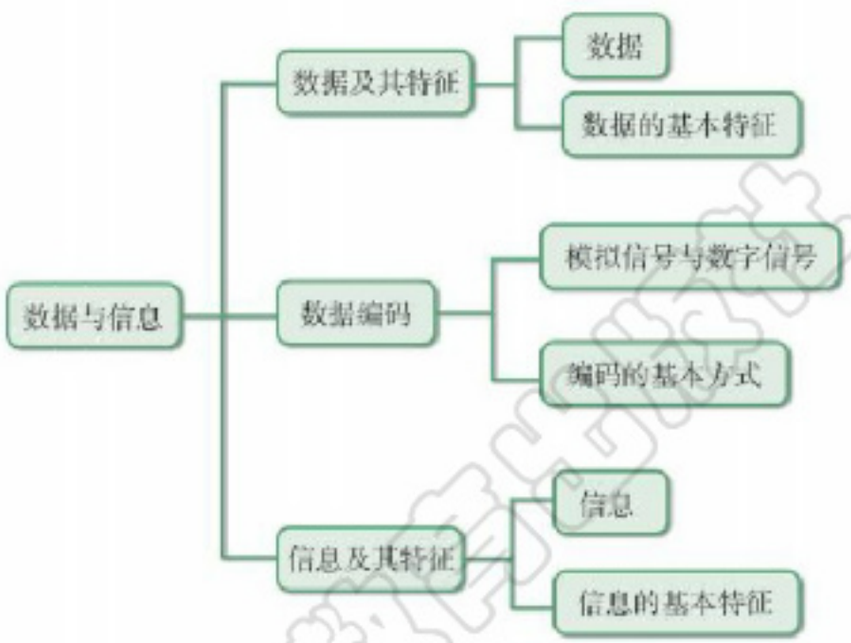
各小组根据项目选题、拟订的项目方案、实施情况以及所形成的项目成果，利用教科书附录2的“项目活动评价表”，开展项目学习活动评价。

广东教育出版社

微信公众号：电子课本大全

本章扼要回顾

同学们通过本章学习，根据“数据与信息”知识结构图，扼要回顾、总结、归纳学过的内容，建立自己的知识结构体系。



回顾与总结

微信公众号：电子课本大全

本章学业评价

同学们完成下列测试题（更多的测试题可以在教科书的配套学习资源包中查看），并通过“本章扼要回顾”以及本章的项目活动评价，综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。

1. 单选题

（1）数据是现实世界原始事物现象的（　　），是信息的载体。

A. 符号记录　　B. 还原　　C. 反映　　D. 提升

（2）下列编码方式中，属于图像编码的是（　　）。

A. ASCII码　　B. GBK　　C. 图像压缩　　D. 统一码

（3）有人写了一个数制转换的式子：70–25=41。若要使该等式成立，则70、25、41有可能分别为（　　）。

A. 十进制、十进制、十六进制　　B. 十进制、十六进制、八进制

C. 十六进制、八进制、十进制　　D. 八进制、八进制、十六进制

2. 思考题

在信息时代如何判别信息的真伪？谈谈你的方法或经验。

3. 情境题

数据压缩。

RLE（Run-Length Encoding，行程长度编码）是一种编码方式，也是一种简单的无损压缩法，压缩和解压缩效率高。

RLE压缩法：利用控制字节的最高位来标识是否进行了压缩。当最高位为1时，则后面7位表示的是数据的重复次数；当最高位为0时，则后面7个位的数是被压缩的数据。

例如，如果有15个十六进制数都是30，那么，压缩后的编码为8F 30。

说明：十六进制数8F转换为二进制是10001111，最高位是1，后面的7位0001111表示字节的重复次数15；十六进制数30转换为二进制是00110000，最高位是0，30是被重复的数据。采取这种方法，一个字符只有重复两次以上，才能被压缩。即使一个数据只重复3次，也可以获得30%的压缩比。

问题1：如果经过RLE方法压缩后的数据为8A 6F，那么解压缩前的数据是什么？

问题2：根据RLE压缩方法，下列字符数据经过压缩后的数据是什么？

20，30，30，30，35，35，…，35，64。其中，十六进制数30共有3个，35共有60个。

微信公众号：电子课本大全

第二章
知识与数字化学习

在人类文明的历史长河中，无论是描述自然现象的实验科学、以牛顿运动定律和麦克斯韦方程等为代表的理论科学，还是模拟复杂现象的计算科学以及数据密集型科学，数据和处理数据的工具都始终驱动着人类文明和社会向前发展。

本章将通过“运用数字化工具探究知识”项目，进行自主、协作、探究学习，让同学们体验数字化学习过程，理解数据、信息与知识的相互关系，感受利用数字化工具和资源的优势，从而将知识建构、技能培养与思维发展融入运用数字化工具解决问题和完成任务的过程中，促进信息技术学科核心素养达成，完成项目学习目标。

- 知识与智慧
- 数字化学习与创新

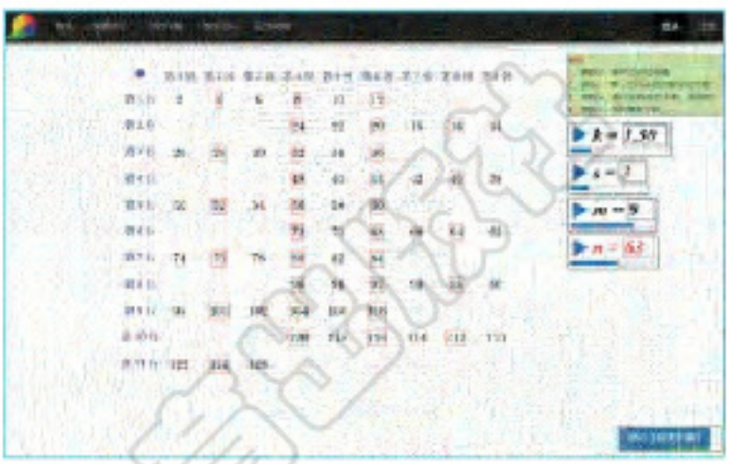
微信公众号：电子课本大全

项目范例　运用数字化工具探究数理知识

情境

未来教室里，教师们运用Python、思维导图、网络画板（图2-1）等各类数字化学习工具进行个性化教学，使教学更加形象、直观、易懂；同学们利用数字化学习工具进行个性化学习，不仅可以开阔视野、开发思维、启迪智慧，而且可以直观地感受知识形成的过程，提高学习效率。

网络画板



思维导图



图2-1　未来教室里的数字化学习工具

主题

运用数字化工具探究数理知识

规划

根据项目范例的主题，在小组中组织讨论，利用思维导图工具，制订项目范例的学习规划，如图2-2所示。

微信公众号：电子课本大全

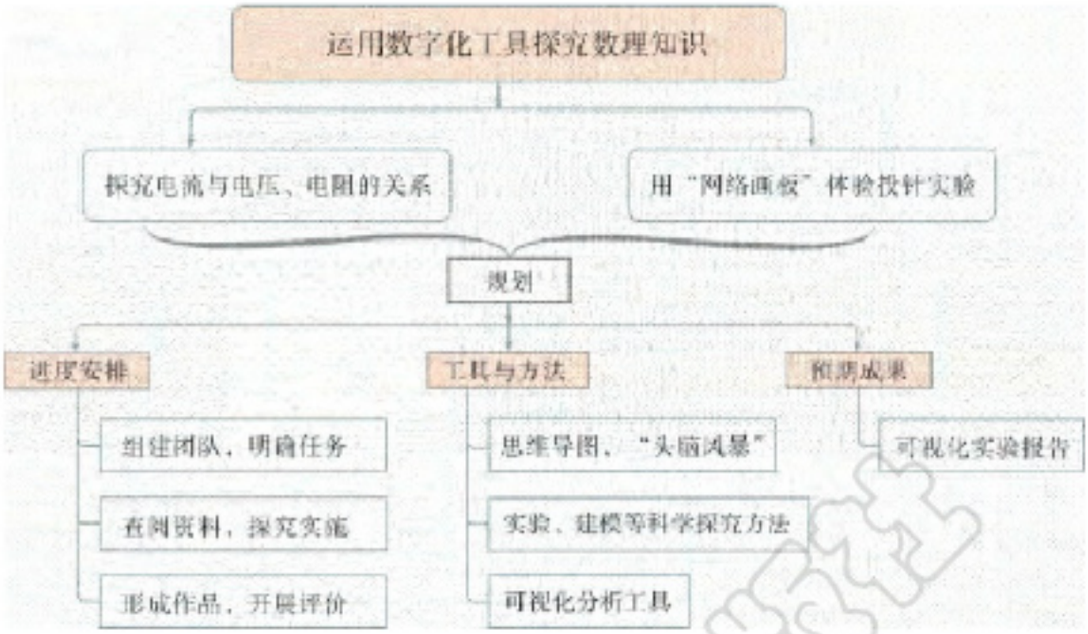


图2-2 “运用数字化工具探究数理知识”项目学习规划



根据项目学习规划的安排，通过调查、案例分析、文献阅读和网上资料搜索，开展“运用数字化工具探究数理知识”项目学习探究活动，如表2-1所示。

表2-1 “运用数字化工具探究数理知识”项目学习探究活动

探究活动	学习内容		知识技能
探究电流与电压、电阻的关系	收集实验数据。	根据观察的现象，选择实验仪器，合理设计实验。	体验数字化学习过程，熟练运用数字化工具收集和储存数据。 理解数据、信息与知识的相互关系。 认识到数据对人们日常生活的影响。
		操作仪器收集实验数据，并记录数据。	
	分析数据，建立模型。	根据记录的实验数据，运用Python的绘图功能，制作散点图。	
		通过观察、分析、猜想和试探，选择趋势线的类型，形成探索性模型。	
	实验验证模型。	再次进行实验，收集数据验证探索性模型。	
用“网络画板”体验投针实验	利用数字化工具进行数学实验。	可以根据问题，选择适当的数字化学习工具。	感受利用数字化资源和工具的优势。
		利用数字化工具，开展学习活动。	

微信公众号：电子课本大全

实施

实施项目学习各项探究活动，进一步运用数字化工具探究数理知识。

成果

在小组开展项目范例学习过程中，利用思维导图工具梳理小组成员在“头脑风暴”活动中的观点，建立观点结构图，运用多媒体创作工具（如演示文稿、在线编辑工具等），综合加工和表达，形成项目范例可视化学习成果，并通过各种分享平台发布，共享创造、分享快乐。例如，运用在线编辑工具制作的“运用数字化工具探究数理知识”可视化报告，可以在教科书的配套学习资源包中查看，其目录截图如图2–3所示。



图2–3　“运用数字化工具探究数理知识”可视化报告目录截图

评价

根据教科书附录2的“项目活动评价表”，对项目范例的学习过程和学习成果在小组或班级上进行交流，开展项目学习活动评价。

项目选题

同学们以3~6人组成一个小组，选择下面一个参考主题，或者自拟一个感兴趣的主题，开展项目学习。

微信公众号：电子课本大全

1. 运用数字化工具探究化学知识
2. 运用数字化工具探究生物知识
3. 运用数字化工具探究地理知识



各小组根据项目选题，参照项目范例的样式，利用思维导图工具，制订相应的项目方案。



各小组将完成的方案在全班进行展示交流，师生共同探讨、完善相应的项目方案。

2.1 知识与智慧

在人类的生产、生活和研究活动中，人们会不断学习、总结、发现认识自然、探究规律的方法。“做实验→获取观察数据→分析处理数据→推理建立数学模型→实验验证模型→形成知识→应用知识解决问题”是一条非常有效的知识发现路径。

2.1.1 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地提炼、研究和分析的结果，能够精确地反映事物的本质。它来源于实践，经过加工提炼，又高于实践。





欧姆在研究电流与电压、电阻关系的实验中利用了控制变量法，即保持电压不变，探究电流和电阻的关系；或保持电阻不变，探究电流和电压的关系。

下面我们运用Python工具，利用控制变量法采集实验数据，研究当导体电压保持不变时，通过导体的电流随导体电阻变化的规律。

微信公众号：电子课本大全

1. 实验仪器

电压表、电流表、电池、开关、导线、滑动变阻器和10个不同阻值的电阻。

2. 实验步骤

（1）收集实验数据。

①连接电路图如图2-4所示。连接电路时注意：接线时开关要断开；闭合开关前，滑动变阻器的滑片要滑到最大阻值处。

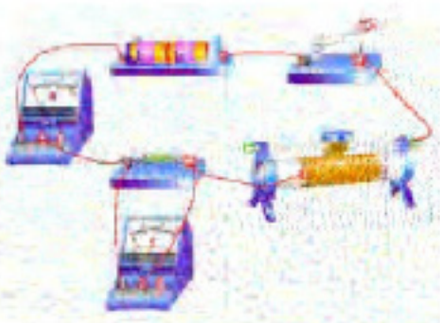


图2-4　欧姆定律实验简化示意图

②把电阻R_1（记下阻值）接入电路，移动滑动变阻器的滑片，使电压表示数U=3 V，记录电流表示数I_1。

③把电阻R_1换成电阻R_2（记下阻值），再次移动滑动变阻器的滑片，使电压表示数U=3 V，记录电流表示数I_2。

④如此类推，把电阻R_9换成电阻R_{10}，再次移动滑片，使电压表示数U=3 V，记录电流表示数I_{10}。

所得到实验数据如表2-2所示。

表2-2　欧姆定律实验的观察数据

电阻R/Ω	5	10	15	20	25	30	35	40	45	50
电流I/mA	586	291	198	143	101	98	72	68	59	55

（2）运用Python工具处理数据，绘制图像。

①绘制散点图。可以从教科书配套学习资源包中安装Python解释器Jupyter notebook，运行配套学习资源包“第二章\课本素材\绘制散点图”程序，把表2-2中的数据输入程序中，执行程序后得到如图2-5所示的电流I与电阻R关系数据散点图。

②在“绘制散点图”程序中加上指令“plt.plot（x0,y0）”，执行程序后，得到如图2-6所示的电流I与电阻R关系数据曲线图。

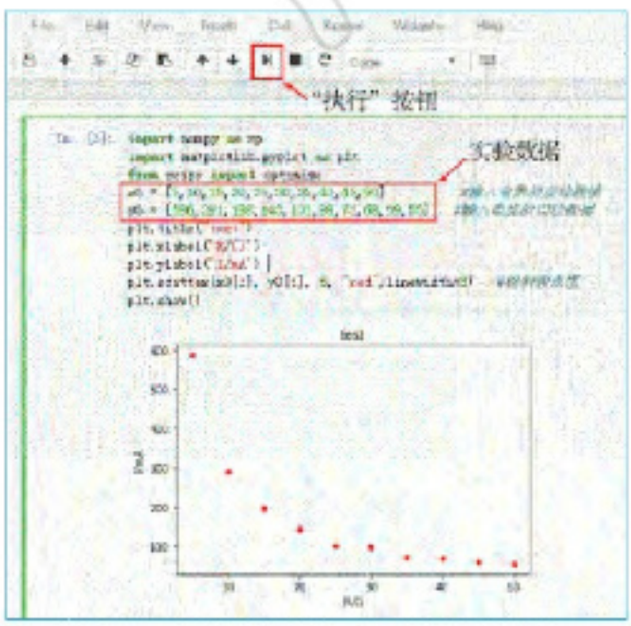


图2-5　电流I与电阻R关系数据散点图

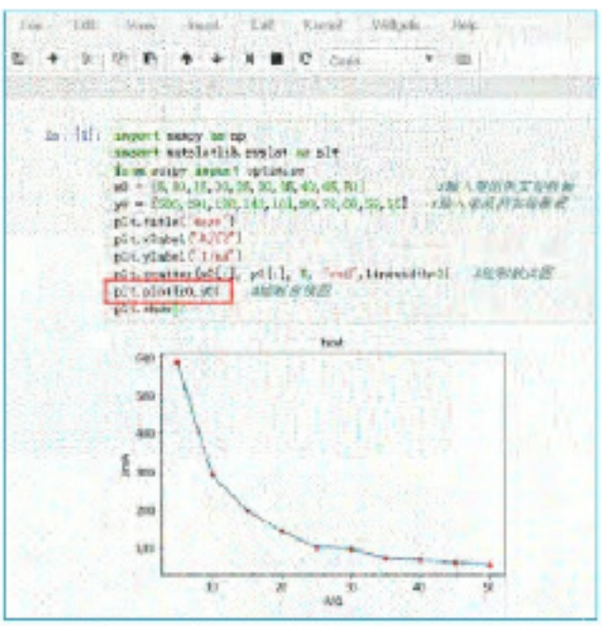


图2-6　电流I与电阻R关系数据曲线图

微信公众号：电子课本大全

③观察电流I与电阻R关系数据曲线图后，可以猜想电流I与电阻R的关系是反比例关系，于是按$I=\frac{U}{R}$的关系对数据做曲线拟合。打开配套学习资源包“第二章\课本素材\曲线拟合”程序，执行程序后得到如图2-7所示的电流I与电阻R关系数据曲线拟合图。其中，U=2905.67607341 mV与电压表的示数基本吻合，即$I=\frac{2905.67607341}{R}$。

④经过推理和实验验证，电流I与电阻R反比例关系成立，即$I=\frac{U}{R}$。

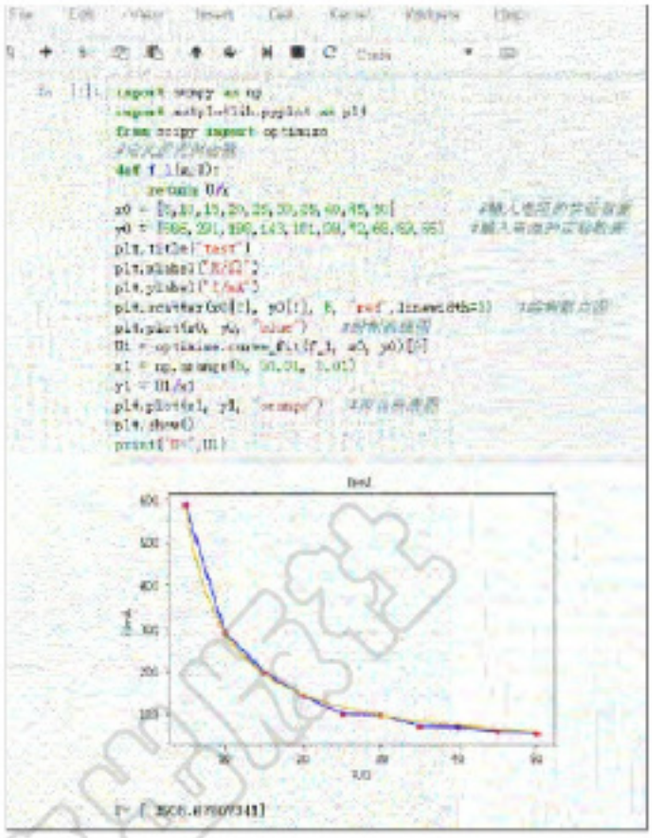


图2-7 电流I与电阻R关系数据曲线拟合图

从以上实验数据的采集、分析、建模和验证过程可以知道，知识的发现不仅可以通过实验、观察和总结为特征的方法获取，还可以通过直接观察获取数据，对数据进行分析，进而发现规律，获得知识。

2.1.2 智慧

在日常生活中，我们要调节台灯的亮度、电视机音量的大小，都要应用欧姆定律的知识，设计一个可变电阻，去调节它们的变化。

探究与应用欧姆定律，揭示了数据、信息、知识和智慧的相互关系（图2-8）：数据是现实世界客观事物的符号记录；信息是经加工处理的、具有意义的数据；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，能够精确地反映事物的本质；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。

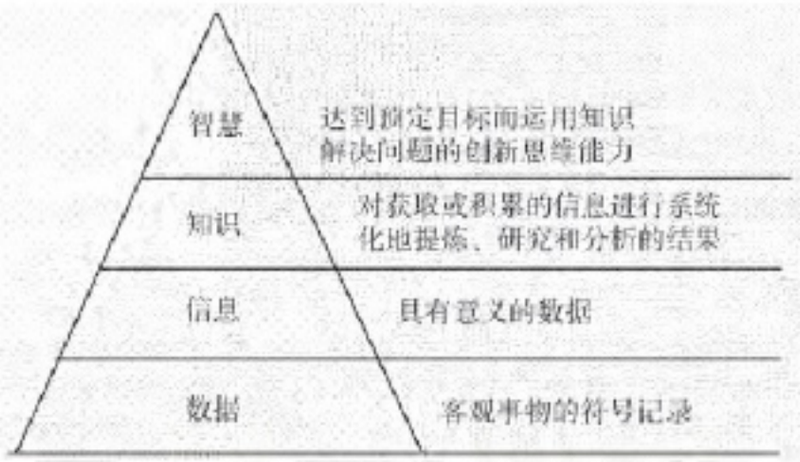


图2-8 数据、信息、知识和智慧的相互关系

微信公众号：电子课本大全

数据、信息、知识和智慧是逐渐递进的概念，前者是后者的基础和前提，后者是前者的抽象与升华。从数据到信息的转化，主要理解数据间的相互联系；从信息到知识的转化，主要理解信息间的相互联系，发现和抽象其规律；从知识到智慧的转化，主要理解和应用知识原理与法则解决现实世界的问题。



举例说明数据、信息、知识和智慧的相互关系。



科学研究的第三种方法——计算方法

科学研究主要依靠理论方法和实验方法。理论方法是指以数学学科为代表，以推理和演绎为特征的方法。研究人员通过构建分析模型和理论推导进行规律预测与发现。实验方法是指以物理学科为代表，以实验、观察和总结为特征的方法。研究人员通过直接观察获取数据，对数据进行分析，进而发现规律，获得知识。

随着移动互联网和物联网的飞速发展，现实世界的各种事物都变得可感知、可度量，从而生成数量庞大的数据或数据群，基于庞大数据形成的仿真系统成为可能。因此，依靠计算发现和预测规律成为与理论方法和实验方法并存的科学研究的第三种方法。研究人员通过建立仿真的分析模型和有效的算法，利用计算工具来进行规律发现和预测。

依靠计算方法发现和预测规律已成为不同学科的科学家进行研究的重要方法。例如，生物学家利用计算方法研究生命体的特性，化学家利用计算方法研究化学反应的机理，建筑学家利用计算方法来研究建筑结构的抗震性，社会学家利用计算方法研究社会群体网络的各种特性等。

由此，计算方法与各学科结合形成了所谓的计算科学，如计算物理学、计算化学、计算生物学、计算经济学等。

著名的计算机科学家、1972年图灵奖得主Edsger Dijkstra说：“我们所使用的工具影响着我们的思维方式和思维习惯，从而也深刻影响着我们的思维能力。”

研究人员在利用计算方法进行创新研究的同时，也在不断地研究新型的计算方法。这种结合不同专业的新型计算方法的研究需要专业知识与计算思维的结合。1998年，John Pople便因成功地研究出量子化学综合软件包Gaussian而获得诺贝尔奖，Gaussian已成为研究化学领域许多课题的重要的计算方法。另一个典型的计算方法是求解应力或疲劳等结构力学、多物理场耦合的有限元分析方法。利用先进的计算方法的虚拟现实、增强现实等技术也在各个领域创造出令人惊喜的视觉效果。

微信公众号：电子课本大全

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，了解科学研究的思维方法与步骤。

1. 体验实验数据收集、分析、建模和验证的过程。
2. 总结归纳科学研究的思维方法与步骤。

2.2 数字化学习与创新

数字化学习与创新是指个体通过评估并选用常见的数字化资源与工具，有效地管理学习过程与学习资源，创造性地解决问题，从而完成学习任务，形成创新作品的能力。具备数字化学习与创新能力的学生，能够认识数字化学习环境的优势和局限，适应数字化学习环境，养成数字化学习与创新的习惯；掌握数字化学习系统、学习资源与学习工具的操作技能，用于开展自主学习、协同工作、知识分享与创新创造，助力终身学习能力的提高。

2.2.1 数字化工具与资源及其优势

1. 数字化工具与资源

数字化工具是指能够采集、获取、检索、表示、传输、存储和加工多媒体数字化资源的设备装置。按其功能不同，数字化工具可大致分为以下几种：

（1）数据采集（模数转换）工具。如声卡、视频卡、数码照相机、数码录像机、手写板、扫描仪等。

（2）信息获取工具。如台式电脑、笔记本电脑、平板电脑和手机等。

（3）信息检索工具。如搜索引擎和电子词典等。

（4）数字化可视化表达工具。如思维导图和演示文稿等。

（5）信息通信交流与传输工具。如电子邮件、博客、社交软件和视频会议系统等。

（6）信息存储工具。如光盘、U盘、硬盘和服务器等。

（7）信息加工工具。如编程软件、多媒体制作软件、网站制作工具和信息系统制作工具等。

（8）三维设计工具。如3D制作软件、3D打印机、虚拟现实和增强现实制作系统等。

数字化资源是指经过数字化工具处理，可以在多媒体计算机上或网络环境下运行的，且可以实现共享的多媒体材料。按其呈现方式不同，数字化资源可分为数字图形、图像、

微信公众号：电子课本大全

视频、音频、文本，多媒体应用软件、网站、信息系统（数字图书馆、科技馆、博物馆等），计算机模拟、数据文件、数据库等。

2. 数字化工具与资源的优势

（1）获取的便捷性。

利用数字化工具与资源学习的学习者可以不受时空和传递呈现方式的限制，通过多种设备，使用各种学习平台获得高质量课程相关信息，可以实现信息的传送、接收、共享、组织和储存。

（2）形式的多样性。

数字化工具与资源以电子数据的形式表现信息内容，其主要的媒体呈现形式有文本、图像、声音、动画、视频等，极大地丰富了信息内容的表现力。除此之外，其友好的交互界面、超文本结构极大地方便了学习者的学习，虚拟仿真的应用也更有助于学习者对知识的记忆与理解。

（3）资源的共享性。

数字化工具与资源具有共享性属性，可以不断地反复利用、复制、传递和再生，为广泛的使用者所享有，而信息量不变。

（4）平台的互动性。

数字化工具与资源和传统的学习工具与资源相比较，其最大的优势在于互动性。无论是通过网络媒介进行的学习方式，还是通过光盘等进行的学习方式，其双向交流的互动方式正得到越来越多学习者的喜爱。一方面，学习者可以通过网络上的交流工具，实现与教师或学习者之间的交互；另一方面，学习者还可以从学习软件的数据库中寻求问题的答案，同时也可将软件数据库自行更新。

（5）内容的扩展性。

数字化工具与资源的扩展性主要表现为可操作性和可再生性。在可操作性方面，数字化学习的过程，既是将课程内容进行数字化处理的过程，同时又是利用共享的数字化资源与课程教学进行融合的过程。这些数字化学习内容能够被评价、被修改和再生产，它允许学习者和教师用多种先进的数字化工具对其进行运用与再创造。在可再生性方面，经数字化处理的课程学习内容能够激发学习者主动学习的兴趣。

探究活动

交流

和同学们一起交流自己使用过的数字化工具与资源及其优势。

调查

以小组为单位，调查各学科教师在教学中所运用到的数字化资源类型，并填入表2-3中。

微信公众号：电子课本大全

表2-3 各学科教师在教学中所运用到的数字化资源

学科	多媒体课件	数字视频	数字音频	电子书	软件	其他
语文						
数学						
英语						
物理						
信息技术						
生物						
地理						
其他						

3. 数字化学习的特点

数字化学习是伴随计算机多媒体技术、互联网通信技术的发展而产生的，是信息社会的重要特征。与传统的学习方式不同，数字化学习具有问题化、合作性、个性化、创造性和再生性、开放性等特点。

（1）问题化。产生学习的根本原因是问题，数字化学习特别强调问题在学习活动中的重要作用。在数字化学习的过程中，学习者一方面参与决定学习问题（或主题），并围绕问题开展学习，将问题看作学习的动力、起点和贯穿学习全过程的主线；另一方面通过学习生成问题、提出问题、分析问题和解决问题。

（2）合作性。数字化学习平台为学习者提供了一个信息交流、资源共享和协作学习的环境。学习者可以在任何时间通过网络向教师提出问题和请求指导；也可以和其他学习者进行学习交流，而且这种交流是双向的，多对多的。数字化学习能使学习者从个体封闭的学习和认知中走出来，学会合作学习。

（3）个性化。数字化学习能满足以学习者为中心的、灵活的和个性化学习的需要。在传统课堂教学中，针对每一名学习者较难展开因材施教与个性化学习。而数字化学习，使这种教与学的方式成为可能。学习者可以根据自己的特点自行安排学习进度，从互联网上选择自己需要的资源，按照适合自己的方式进行学习。在整个学习过程中，学习者通过数字化学习进行主动的意义建构，从而获取知识，真正成为学习的主人，如图2-9所示。

微信公众号：电子课本大全



图2-9　个性化学习

（4）创造性和再生性。数字化学习具有创造性和再生性。互联网采用超文本链接的形式联结信息，非线性、跳跃性是其重要特征。网络课程的组织方式符合人们的思维规律，有利于人的发散性思维的发展，也有利于丰富想象力的培养和创新意识、创新思维的发展。

数字化的课程学习内容使学习过程变得不再枯燥，学习者更愿意参与和沉浸到学习中去。学习者不再是被动地接收信息，而是采用新颖、熟练的数字化加工方法，进行知识的整合、再创造，并作为学习者的学习成果。数字化学习的可再生性，不仅能很好地激发学习者的创造力，而且为学习者创造力的发挥提供了更大的可能。

（5）开放性。在社会已进入移动互联时代的今天，数字化学习资源的全球共享，虚拟课堂、虚拟学校的出现，使得人们可以随时随地通过互联网进入数字化的学习环境，实现碎片化学习。学习空间正变得无围墙、无界限。同时，个体的学习也不再是阶段性的，而是终身的；数字化学习为构建学习型社会提供了有力的保障，推动了终身教育的发展。

深入了解数字化学习的各种特点，既有利于我们准确地把握和理解数字化学习的基本内涵，也有利于我们根据数字化学习的特点，有效地运用数字化学习工具开展自主学习、协同工作与知识分享，适应数字化学习环境。



1. 收集身边的、生活中常见的数字化学习事例，分析人们是如何利用数字化学习为生活、学习服务的。

2. 分析传统学习与数字化学习两种方式的区别，填写表2-4。

微信公众号：电子课本大全

表2-4 传统学习与数字化学习两种方式的比较

比较内容	学习方式	
	传统学习	数字化学习
学习问题		
学习效率		
学习时间及地点		
情感交流的效果		
……		

2.2.2 体验数字化创新学习过程

在信息时代，学习者不仅可以向教师与课本学习，而且可以高效利用数字化工具与资源的优势，在数字化学习环境中，与同伴和教师高效开展自主、协作、探究学习，创造性地完成具体的学习任务。

体验

投针实验

蒙特卡洛方法是计算机模拟的基础，起源于18世纪数学家蒲丰的投针实验。投针实验是指在平面上画有一组间距为d的平行线（图2-10），将一根长度为l（$l \leqslant d$）的针任意掷在这个平面上，求此针与平行线中任一条相交的概率。这是一个数学上的概率问题。当人们反复进行大量随机投针实验后，发现了一个独特的现象，即针与平行线中任一条相交的概率是一个常数。那这个常数的倒数是什么呢？

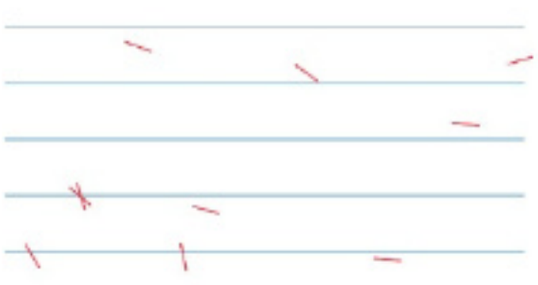


图2-10 投针实验

下面我们利用开放共享的“网络画板”平台模拟投针实验，体验使用计算机模拟随机实验处理确定性数学问题的思维方法。我们设定平行线的间距d=1，针长l=0.5，抛掷次数a从100到5000，k为记录针与平行线中任一条相交的次数，探究针与平行线中任一条相交的概率的倒数$P=\frac{a}{k}$的变化。具体操作步骤如下：

（1）在线打开“网络画板”平台，点击“开始作图”按钮，进入画板的作图界面。

（2）点击“文件”菜单的“打开本地文件”按钮，打开并运行配套学习资源包中“第二章\课本素材\投针实验.npc”文件，如图2-11所示。

微信公众号：电子课本大全

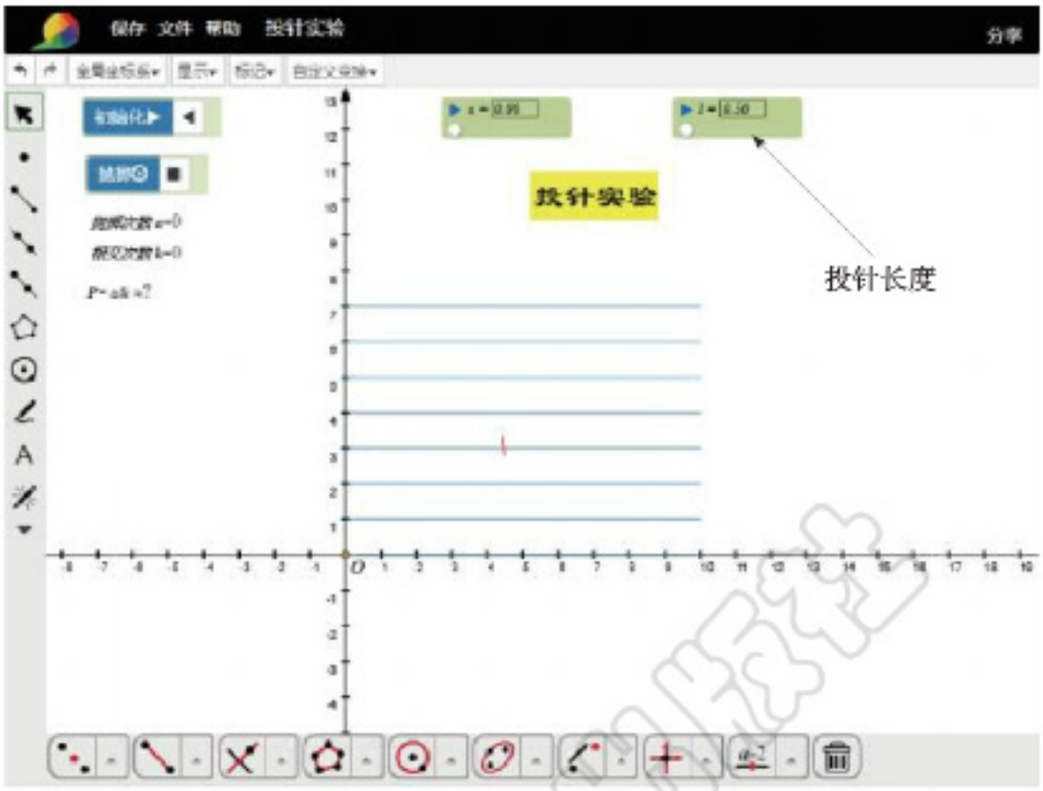


图2–11　投针实验界面图

（3）点击“抛掷”按钮，开启或停止“抛掷”，观察P值的变化，重复多次操作，可发现$P \approx \pi$的近似值，如图2–12所示。

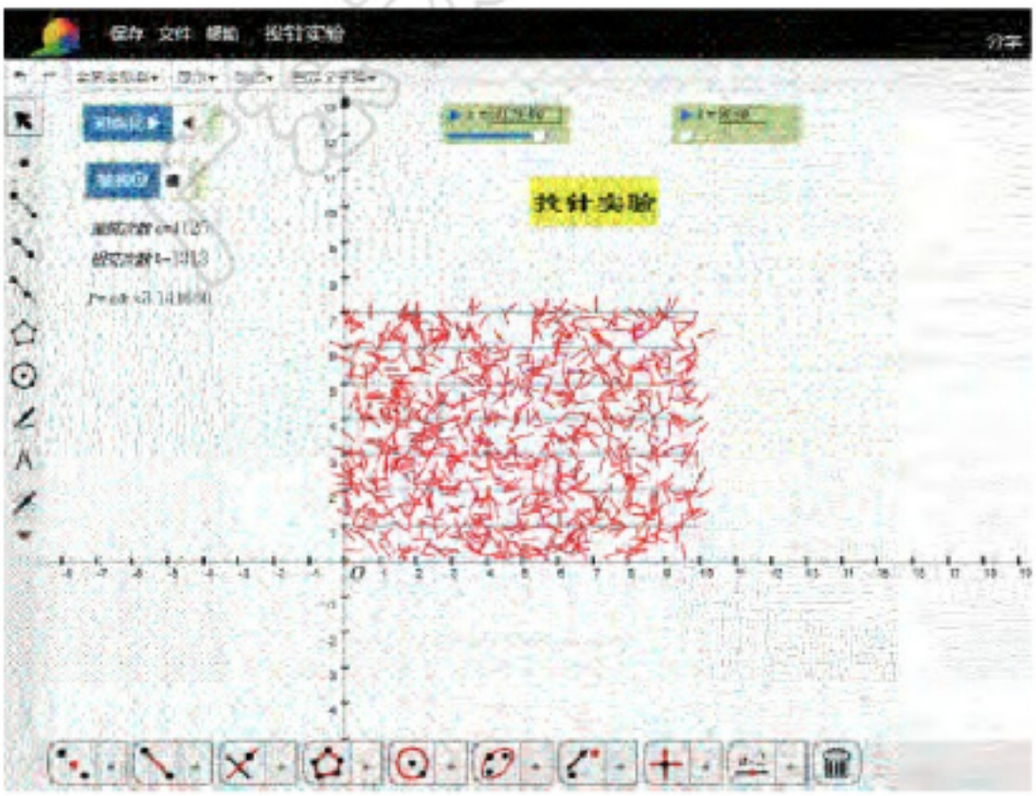


图2–12　投针实验过程图

微信公众号：电子课本大全

交流

1. 查阅初中数学教材或上网查阅有关资料，讨论针与平行线中任一条相交的概率的倒数$P=\frac{a}{h}\approx\pi$的推理过程，感受数字化模拟实验的过程，体验数字化学习工具的优势。

2. 以小组为单位，运用思维导图软件绘制思维导图，分类列举在项目实施过程中所涉及的数字化工具，生成JPG文件，并在班级开展交流与评价。

因数字化学习环境和资源利用方式不同，学习工具、途径和方法多样，数字化学习模式也不尽相同。因此，只有牢固树立数字化学习的意识，学会数字化学习的技术手段和方法，养成数字化学习的习惯，才能促进终身学习，实现自主发展。

分析

1. 利用数字化学习的特点，运用数字化学习方式，针对小组选定的项目，各小组广泛收集与项目相关的研究资料。资料的来源可以是报纸杂志、书籍、广播、电视等，也可以是学习网站、教育资源平台、论坛、电子邮件及搜索引擎等。

2. 按照一定规则将研究素材分门别类进行整理，便于开展研究性学习。

讨论

思维导图既可以手绘，也可以使用软件绘制。对于绘制思维导图的软件，目前比较知名的有XMind、Mindmanager、FreeMind等。现以小组为单位，自行安装应用XMind、Mindmanager、FreeMind三种软件绘制思维导图，小组讨论比较三者之间的异同，并填写表2-5。

表2-5 三种思维导图软件比较

比较项目	软件		
	XMind	Mindmanager	FreeMind
是否开源			
支持平台			
模板功能			
导出格式			

微信公众号：电子课本大全

（续表）

比较项目	软件		
	XMind	Mindmanager	FreeMind
优点			
不足之处			
综合评价			

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，了解数字化工具与资源及其优势，进一步完善该项目方案中的各项学习活动，并参照项目范例的样式，撰写本组的项目成果报告。

成果交流

各小组运用数字化学习工具，将所完成的项目成果，在小组或班级上进行展示与交流，共享创造、分享快乐。

活动评价

各小组根据项目选题、拟订的项目方案、实施情况以及所形成的项目成果，利用教科书附录2的“项目活动评价表”，开展项目学习活动评价。

微信公众号：电子课本大全

本章扼要回顾

同学们通过本章学习，根据“知识与数字化学习”知识结构图，扼要回顾、总结、归纳学过的内容，建立自己的知识结构体系。



回顾与总结

微信公众号：电子课本大全

本章学业评价

同学们完成下列测试题（更多的测试题可以在教科书的配套学习资源包中查看），并通过“本章扼要回顾”以及本章的项目活动评价，综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。

1. 单选题

（1）下列工具中，更适用于免费在线存储学习资源的是（　　）。

A. 微信　　B. 云盘　　C. 博客　　D. 微博

（2）以下关于数据、信息和知识相互关系的理解，不正确的是（　　）。

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）下列属于数字化学习工具的是（　　）。

A. 视频　　B. 音频　　C. 多媒体课件　　D. 思维导图软件

2. 思考题

用思维导图描述数据、信息、知识和智慧之间的关系。

3. 情境题

（1）在公交车或地铁上，我们经常看到人们使用手机、平板电脑等智能设备，或阅读电子书籍，或看新闻，或发微博、微信等，充分利用碎片化时间进行阅读……

任务：针对这种情况，对比纸质读物，请说明数字化阅读的利弊。

（2）新学期开始，某一中学投入资金配置了演播室。接下来，校团委招募了十几名同学，成立了“校园电视台”学生社团，定期制作播报校园新闻。由于演播室的设备尚未配备齐全，为便于主持人顺利报道新闻，需要设计一个自动播放文本的程序，实现现场提词器的功能，如右图所示。

任务：应用数字化学习的方法，完成上述任务，并简述其制作过程。



现场提词器

微信公众号：电子课本大全

第三章
算法基础

数据在信息社会中具有重要价值，掌握数据处理的基本方法与技能具有重要意义。随着数据的快速增长，掌握利用计算机和算法更高效地分析和解决海量数据问题的方法对人类社会发展具有重要作用。

本章将通过“设计从A市到B市最佳的旅行路线方案”项目，进行自主、协作、探究学习，让同学们体验计算机解决问题的过程，理解和概述算法的概念与特征，学会运用恰当的描述方法和控制结构表示简单算法，懂得描述程序设计语言产生与发展的过程，了解不同种类程序设计语言的特点，从而将知识建构、技能培养与思维发展融入运用数字化工具解决问题和完成任务的过程中，促进信息技术学科核心素养达成，完成项目学习目标。

- 体验计算机解决问题的过程
- 算法及其描述
- 计算机程序与程序设计语言

微信公众号：电子课本大全

项目范例　设计从A市到B市耗时最少的旅行路线方案

情境

中国桥、中国路、中国车、中国港、中国网，一个个奇迹般的工程，让人民切切实实地感受到生活的便捷、美好和幸福。截至2017年10月，中国高速公路里程13.1万千米，位居世界第一，2020 年将达15万千米；高铁里程2.2万千米，位居世界第一（图3-1）；城市轨道交通4153千米，位居世界第一；光缆线路3041万千米，位居世界第一；世界前10的斜拉桥，中国占7座；世界前10的悬索桥，中国占6座；吞吐量世界前10的港口，中国占7席。



图3-1　中国高铁

当从A市到B市没有直达的交通工具时（不考虑水上交通工具），人们可以利用铁路公司、汽车客运公司和航空公司公布的信息，设计出耗时最少的旅行路线。

主题

设计从A市到B市耗时最少的旅行路线方案

规划

根据项目范例的主题，在小组中组织讨论，利用思维导图工具，制订项目范例的学习规划，如图3-2所示。

微信公众号：电子课本大全

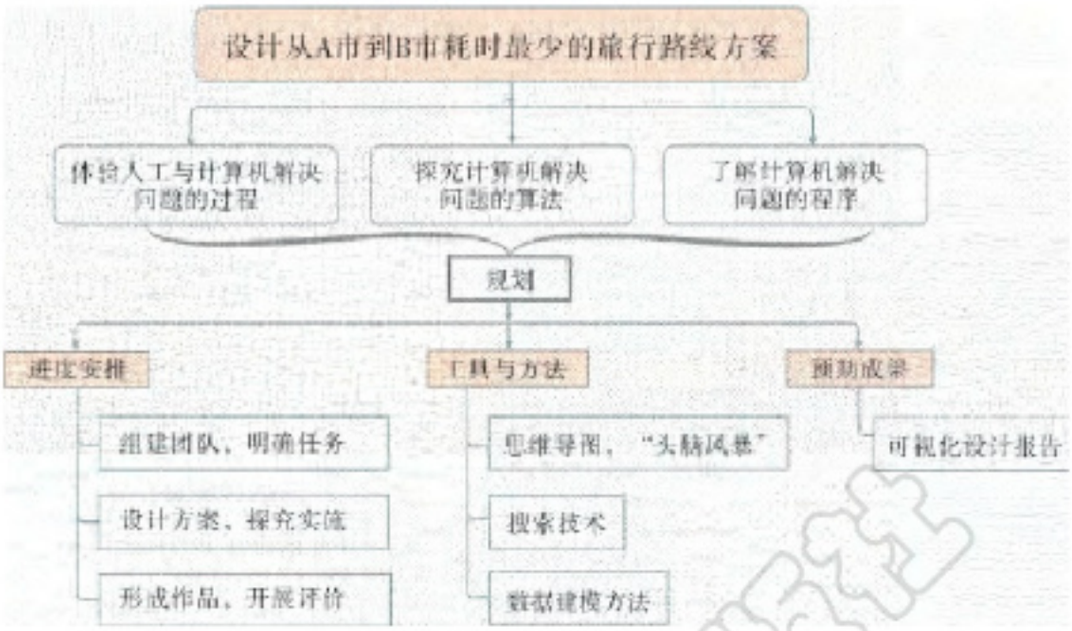


图3-2 “设计从A市到B市耗时最少的旅行路线方案”项目学习规划



根据项目学习规划的安排，通过调查、案例分析、文献阅读和网上资料搜索，开展“设计从A市到B市耗时最少的旅行路线方案”项目学习探究活动，如表3-1所示。

表3-1 “设计从A市到B市耗时最少的旅行路线方案”项目学习探究活动

探究活动	学习内容		知识技能
体验人工与计算机解决问题的过程	设计从A市到B市耗时最少旅行路线的问题分析。	人工解决问题的过程。 计算机解决问题的过程。	体验计算机解决问题的过程。
探究计算机解决问题的算法	设计从A市到B市耗时最少旅行路线的算法。	算法的概念与特征。 算法的描述方法。	理解和概述算法的概念与特征。 运用恰当的描述方法和控制结构表示简单算法。
了解计算机解决问题的程序	编写求解从A市到B市耗时最少的旅行路线问题的程序。	计算机程序。 程序设计语言的产生与发展。	描述程序设计语言产生与发展的过程。 了解不同种类程序设计语言的特点。

微信公众号：电子课本大全

实施

实施项目学习各项探究活动，进一步设计耗时最少的旅行路线。

成果

在小组开展项目范例学习过程中，利用思维导图工具梳理小组成员在“头脑风暴”活动中的观点，建立观点结构图，运用多媒体创作工具（如演示文稿、在线编辑工具等），综合加工和表达，形成项目范例可视化学习成果，并通过各种分享平台发布，共享创造、分享快乐。例如，运用在线编辑工具制作的“设计从A市到B市耗时最少的旅行路线方案”可视化报告，可以在教科书的配套学习资源包中查看，其目录截图如图3-3所示。



图3-3　“从A市到B市耗时最少的旅行路线方案”可视化报告目录截图

评价

根据教科书附录2的“项目活动评价表”，对项目范例的学习过程和学习成果在小组或班级上进行交流，开展项目学习活动评价。

项目选题

同学们以3~6人组成一个小组，选择下面一个参考主题，或者自拟一个感兴趣的主题，开展项目学习。

1. 设计从A市到B市交通费最少的旅行路线方案
2. 设计从A市到B市路程最短的旅行路线方案
3. 设计从A市到B市交通费大约相同而耗时最少的旅行路线方案

微信公众号：电子课本大全

项目规划

各小组根据项目选题，参照项目范例的样式，利用思维导图工具，制订相应的项目方案。

方案交流

各小组将完成的方案在全班进行展示交流，师生共同探讨、完善相应的项目方案。

3.1 体验计算机解决问题的过程

在现实生活中，我们经常需要对数据进行统计、分析。当数据量不多时，我们可以采用人工方法来处理；然而，当数据量变多时，我们运用计算机来解决问题将是一种更高效、更便捷的方法。

3.1.1 人工解决问题的过程

采用人工方法来解决问题，首先需要明确所要解决的问题和给出的条件，然后再根据已有的经验和知识确定解决问题的方法，从而解决问题。

探究活动

思考

如何设计从A市到B市耗时最少的旅行路线方案呢？假如我们从铁路公司、各航空公司和汽车客运公司网站得知，直达B市的交通工具只有火车和汽车两种，出发地有B_1，B_2，…，B_k市（没有A市），从A市出发到B_1，B_2，…，B_k市的交通工具有飞机、火车和汽车三种，这样从A市经B_1，B_2，…，B_k市到B市的交通情况如图3-4所示。

由于从A市到B_1，B_2，…，B_k市有不同的交通工具，每一种交通工具又有不同的班次，因此从A市出发到中转城市B_1，B_2，…，B_k市就有M_1，M_2，…，M_k种班次。同样，

微信公众号：电子课本大全

从中转城市B_1，B_2，…，B_k市到B市也有不同的交通工具，每一种交通工具有不同的班次，因此从中转城市B_1，B_2，…，B_k市到B市就有N_1，N_2，…，N_k种班次。于是从A市经B_1，B_2，…，B_k市到B市的交通班车（班机）数共有：

$$S=M_1\times N_1+M_2\times N_2+\cdots+M_k\times N_k$$

寻找从A市到B市耗时最少的旅行路线问题就转化为在S种联运班次中找到一种耗时最少的联运班次。这样就需要遍历每一个班次进行比较。若用人工方式找出能够中转且等待时间和行驶时间最少的班次，工作量将极其浩大！

假设从A市到B市的中转城市只有B_1，B_2市，从A市经B_1，B_2市到B市的交通情况如表3-2和表3-3所示。

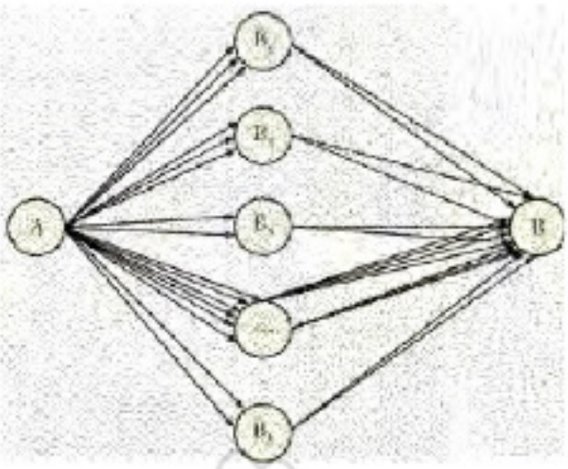


图3-4　从A市经B_1，B_2，…，B_k市到B市的交通情况

表3-2　从A市经B_1市到B市的交通情况

从A市到B_1市的交通工具	发车时间	行驶时间/时	从B_1市到B市的交通工具	发车时间	行驶时间/时
飞机	9:00	3	飞机	无	无
	12:00	2.5			
	16:00	2			
火车	8:00	12	火车	9:00	1.5
	10:00	11		11:00	1.33
	12:00	9.6		15:00	1
	13:00	9.5		18:00	2
汽车	无	无	汽车	8:00	1.8
				9:00	2
				13:00	2.2
				15:00	2.66
				18:00	1.66

微信公众号：电子课本大全

表3-3 从A市经B_2市到B市的交通情况

从A市到B_2市的交通工具	发车时间	行驶时间/时	从B_2市到B市的交通工具	发车时间	行驶时间/时
飞机	9:00	3	飞机	无	无
	11:00	2.5			
	15:00	2.5			
火车	8:30	12.6	火车	8:00	1.5
	11:20	11		10:00	1.3
	14:00	10.6		14:00	1
	16:00	11.6		19:00	2
汽车	8:30	22.6	汽车	8:00	1.8
	11:20	21		9:00	2.1
	14:00	19.6		13:00	2.5
	16:00	19.6		16:00	2.2
	18:00	20		19:00	1.7

从以上两表可知，从A市经B_1市到B市的联运班次有$7\times9=63$（班），从A市经B_2市到B市的联运班次有$12\times9=108$（班），合计为$S=63+108=171$（班）。然后在171班次中找到能够中转且等待时间加上行驶时间最少的联运班次，如图3-5所示。

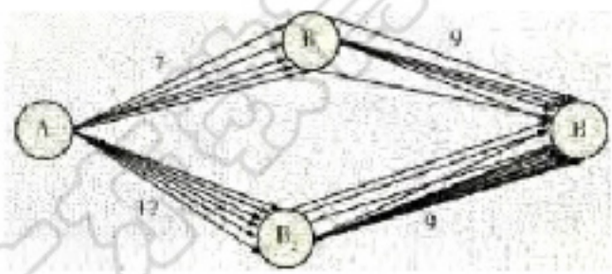


图3-5 从A市到B市的班次状况



根据表3-2和表3-3给出的已知条件，可以采用以下的思路求解耗时最少的联运班次问题：

（1）找出能够中转的从A市经B_1市到达B市的联运班次，并计算所用的时间。

（2）找到能够中转的从A市经B_1市到达B市的联运班次中耗时最少的联运班次。

（3）找出能够中转的从A市经B_2市到达B市的联运班次，并计算所用的时间。

（4）找到能够中转的从A市经B_2市到达B市的联运班次中耗时最少的联运班次。

（5）取两条线路中耗时最少的联运班次为最佳旅行路线。

上述问题中，假如中转城市很多，交通班次也很多，找出耗时最少路线的工作量会非常大，若用人工穷举遍历，其效率就会很低。

微信公众号：电子课本大全

3.1.2 计算机解决问题的过程

当数据量很大，人工处理效率很低时，我们可以借助计算机，通过编写计算机程序解决问题。编写计算机程序解决问题要经过分析问题、设计算法、编写程序、调试运行程序等若干个步骤。

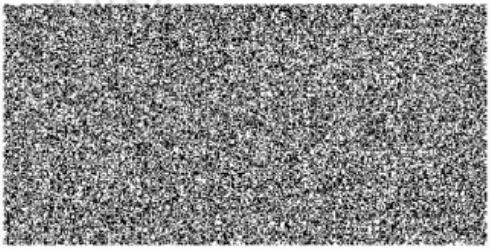
1. 分析问题

在利用计算机解决问题之前，我们首先要分析问题的需求情况、已知条件和需要解决的问题。

例如，在从A市到B市耗时最少的旅行路线问题中，在不知道有多少个中转城市和每个城市有多少班车（或飞机）的情况下，我们可以利用大数据挖掘技术中的爬虫程序（参见配套学习资源包“第三章\课本素材\程序3-1”）到铁路网站、各航空公司和汽车客运公司网站获取从A市经中转城市B_1，B_2，…，B_n市到达B市的交通班次信息，再经过数据清洗后，形成结构化的数据存储为Excel文件（例如，班次查询结果的截图如图3-6所示，详细文件可参见配套学习资源包“第三章\课本素材\Excel”文件夹）。

序号	A	B	C	D	E
1	交通工具	出发地	出发时间	到达地	到达时间
2	飞机	A	9:00	B_1	12:00
3	飞机	A	12:00	B_1	14:30
4	飞机	A	16:00	B_1	18:00
5	火车	A	8:00	B_1	20:00
6	火车	A	10:00	B_1	21:00
7	火车	A	12:00	B_1	21:36
8	火车	A	13:00	B_1	22:30

清洗后的数据表B_1 Sheet1



清洗后的数据表B_1 Sheet2

图3-6 班次查询结果的截图

2. 设计算法

问题分析清楚后，需要给出解决问题的详细方法和步骤，这一过程称为设计算法。

例如，对于从A市到B市耗时最少的旅行路线问题，根据获取的从A市到B市的中转城市

微信公众号：电子课本大全

B_1，B_2，…，B_k的班次，以及各城市各交通班次的发车时间和行驶时间等信息，采用以下的思路找出耗时最少的联运班次问题：

（1）分别找出能够中转且中转等待时间不小于1小时的从A市经B_1，B_2，…，B_k市到达B市的联运班次，并计算所用的时间。

（2）分别找到能够中转的从A市经B_1，B_2，…，B_k市到达B市的联运班次中耗时最少的联运班次，共k条线路。

（3）取k条线路中耗时最少的联运班次为最佳旅行路线。

3. 编写程序

有了清晰可操作的算法描述，就可以选择一种计算机语言工具来编写程序，实现算法。一般来说，只要算法确定，对计算机程序设计语言的选择没有特别的限定，通常根据问题的特性和编程人员对语言的熟悉程度来选定编写程序。

例如，用Python语言编写从A市到B市耗时最少的旅行路线问题的算法的程序可参见配套学习资源包“第三章\课本素材\程序3-1”。其中，找出能够从A市经B_i（i=1，2，…，k）市到达B市的中转联运班次，并计算所用的时间以及找到耗时最少的联运路线的关键程序段如下。

```
m=99
for i in range(1,rs1):
    t14 = table_1.cell(i,4).value
    t12 = t14-table_1.cell(i,2).value
    for j in range(1,rs2):
        t22 = table_2.cell(j,2).value
        if t22-t14>=1/24 :
            m1=t12+(t22-t14)+(table_2.cell(j,4).value-t22)
            if m>m1:
                m=m1
                r1=i
                r2=j
```

4. 调试运行程序

程序编写完成以后，再通过键盘把程序输入计算机中运行，检查程序能否按预想的效果执行，这一过程称为程序的调试运行。计算机只能识别程序设计语言中所规定的语法规则，如果编写程序时与规则不一致，哪怕是一个标点符号出错，也会因程序出错而中断运行。此时，我们可以根据计算机提示的出错信息修改程序，重新调试运行。由于Python是解释程序，因此它的调试是在运行过程中逐行进行的。

当程序能够顺利运行以后，我们还需要对程序运行的结果进行检查。因为如果程序语句符合语法规则，而程序中却有逻辑或计算方法等错误，计算机是检查不出来的。因此，如果结果不合理，还要对程序甚至算法进行修改，直到程序的功能符合设计要求为止。

微信公众号：电子课本大全

实践

打开配套学习资源包“第三章\课本素材\程序3-1”，调试并运行程序3-1，找出从A市到B市耗时最少的旅行路线问题的结果，如图3-7所示。

从A地出发经B_1到达B地，最少耗时为：4.333333333333332小时。具体行程请查看文件ZHXC.XLS。

（a）运行结果



	A	B	C	D	E
1	交通工具	出发地	出发时间	到达地	到达时间
2	飞机	A	10:00:00	B_1	12:20:00
3	汽车	B_2	13:40:00	B	14:20:00

（b）ZHXC.XLS文件截图

图3-7　从A市到B市耗时最少的旅行路线问题的程序运行结果

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，体验计算机解决问题的过程。

1. 体验运用计算机解决问题经历的问题描述、数据抽象和结构分析、模型建立、算法设计、程序编写、程序调试和测试验证等过程。

2. 总结归纳运用计算机解决问题的方法和步骤。

3.2 算法及其描述

3.2.1 算法

1. 算法

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。通俗地说，算法就是用计算机求解某一问题的方法，是能被机械地执行的动作或指令的有穷集合。

探究活动

观察

若要求方程$6x+5y+4z=50$的正整数解的个数t，则解决问题的算法步骤如下：

①$t=0$；

微信公众号：电子课本大全

②$x=1$；

③$y=1$；

④$z=1$；

⑤如果满足式子$6x+5y+4z=50$，则解的个数加1（即$t=t+1$，表示右边式子的值赋值给左边式子），并输出这个解（即输出t，x，y，z的值）；

⑥$z=z+1$；

⑦如果$z\leqslant 12$则转步骤⑤，否则继续步骤⑧；

⑧$y=y+1$；

⑨如果$y\leqslant 10$则转步骤④，否则继续步骤⑩；

⑩$x=x+1$；

⑪如果$x\leqslant 8$则转步骤③，否则继续步骤⑫；

⑫结束。

2. 算法的特征

算法作为能确实解决某个问题的策略，具有五个方面的重要特征：

（1）有穷性。一个算法在执行有穷步之后必须结束，即一个算法所包含的计算步骤是有限的。例如，在上面的算法中，x的值从1开始穷举，重复执行语句，直到$x>8$时终止执行。

（2）确定性。算法执行的每一个步骤必须有确切的定义，不能出现模棱两可的情况。例如，上面算法步骤⑤就明确规定：当满足式子$6x+5y+4z=50$时，则解的个数加1（即$t=t+1$），并输出这个解。

（3）数据输入。一个算法必须有零个或多个数据输入，以刻画运算对象的初始情况。例如，在上面的算法中，就没有数据输入。

（4）数据输出。一个算法有一个或多个数据输出，以反映对输入数据加工后的结果，没有输出的算法是毫无意义的。例如，在上面的算法中，有两个输出，即步骤⑤的个数t和具体解（x，y，z的值）。

（5）可行性。算法中执行的任何计算步骤都可以被分解为基本的可执行的操作步骤，即每个计算步骤都可以在有限时间内完成。例如，上面的算法中每一步都是可以在有限时间内完成的。

3.2.2 算法的描述

算法是对解题过程的精确描述，且需要使用某种方法将其表示出来。

1. 描述算法的常用方法

描述算法的常用方法有自然语言描述算法、流程图描述算法和伪代码描述算法。

（1）用自然语言描述算法。

用自然语言描述算法，就是用人们日常所用的语言，如汉语、英语等来描述算法。例如，从A市到B市耗时最少的旅行路线问题的算法描述，即使用了自然语言。

使用自然语言描述算法比较容易掌握，但也存在明显的缺点。例如，当算法中含有多分支或循环操作较多时，使用自然语言很难将其清晰地表示出来；并且由于自然语言的歧

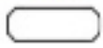
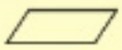
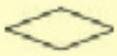
微信公众号：电子课本大全

义性，也容易导致算法执行的不确定性。

（2）用流程图描述算法。

用流程图描述算法是用程序框图来描述算法的一种表示方法。使用流程图描述算法，可使算法的流程描述得清晰、简洁。流程图的基本图形及其功能如表3-4所示。

表3-4　流程图的基本图形及其功能

图形	名称	功能
	开始/结束	表示算法的开始或结束。
	输入/输出	表示算法中变量的输入或输出。
	处理	表示算法中变量的计算与赋值。
	判断	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。

例如，用流程图描述求方程$6x+5y+4z=50$的正整数解的算法，如图3-8所示。

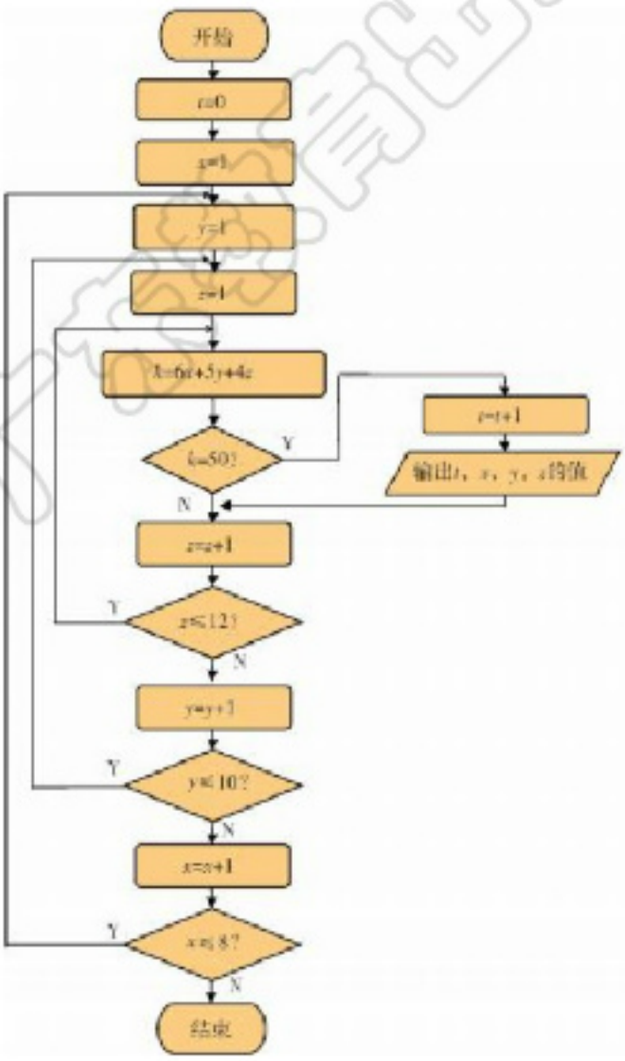


图3-8　求方程正整数解的流程图

微信公众号：电子课本大全

（3）用伪代码描述算法。

用伪代码描述算法就是用介于自然语言和计算机语言之间的文字和符号来描述算法。它不用图形符号，书写方便，格式紧凑，易于理解，便于向计算机程序设计语言过渡。

例如，用伪代码描述求解方程6x+5y+4z=50的正整数解的算法如下：

```
t=0
for x in range(1,9):
    for y in range(1,11):
        for z in range(1,13):
            if(x*6+y*5+z*4==50)
             { t=t+1;
             给出解的个数t及对应的三个整数x,y,z}
```



1.根据求解不定方程的知识，简化方程6x+5y+4z=50中变量x、y、z的筛选范围，设计并交流更加简洁高效的算法。

2.各小组交流三种算法描述方法的优势和不足，并完成表3-5。

表3-5 三种算法描述方法的比较

算法描述的方法	优势	不足
自然语言描述法		
流程图描述法		
伪代码描述法		



在《几何原本》一书中，欧几里得阐述了关于求两个正整数的最大公约数的过程，这就是著名的欧几里得算法——辗转相除法，其具体过程如下：

设给定的两个正整数为m和n，求它们的最大公约数的步骤为：

①以m除以n，令所得的余数为R。

②若$R=0$，则输出结果n，算法结束；否则，继续步骤③。

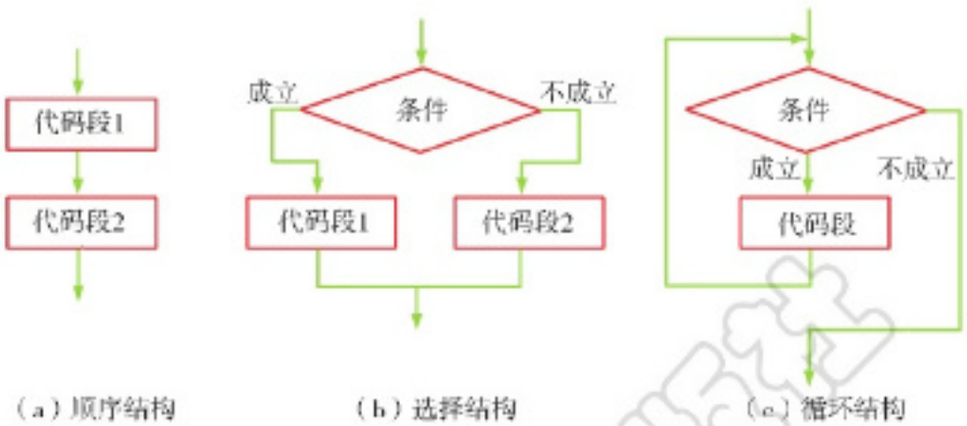
③令$m=n$，$n=R$，并返回步骤①继续进行。

用流程图将上述算法表示出来，试探索欧几里得算法在现实生活中有哪些应用，举出两个应用实例。

微信公众号：电子课本大全

2. 三种基本控制结构

前面的算法描述中，我们用到了顺序结构、选择结构和循环结构这三种基本控制结构（其流程图如图3–9所示），而任何复杂的算法都可以用这三种基本控制结构组合来表示。



（a）顺序结构　　（b）选择结构　　（c）循环结构

图3–9　三种基本控制结构

这三种基本控制结构的主要作用是：

（1）顺序结构表示程序中的各步操作按出现的先后顺序执行。

（2）选择结构表示程序的处理步骤出现了分支，需要根据某一特定的条件选择其中的一个分支执行。选择结构有单选择、双选择和多选择三种。

（3）循环结构表示程序反复执行某个或某些操作，直到判断条件为假（或为真）时才可终止循环。

使用三种基本控制结构的组合来描述算法，可以改善算法的清晰度，提高算法的可读性，原因如下：

（1）以控制结构为单位，只有一个入口和一个出口，各单位之间接口简单，比较容易独立地理解每一单位。

（2）缩小了算法的静态描述与动态执行过程之间的差异，使得两者容易对应，易于理解。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，开展以下活动。

1. 完成相应问题的算法设计及其描述。
2. 总结归纳所采用的方法和步骤。

微信公众号：电子课本大全

3.3 计算机程序与程序设计语言

目前，我们已对计算机解决问题的基本过程和算法有了一定的了解。通过对本节任务的探究学习，我们将更深切体会计算机程序在解决问题过程中的重要作用。

探究活动

思考

在完成问题分析和算法设计两个环节之后，接下来就要开始编写计算机程序对数据进行统计分析，进而形成解决问题的方案。各小组交流什么是计算机程序语言。

1. 围绕交流的问题展开“头脑风暴”活动，确定认识计算机程序需要解决的问题和知识，具体可参考表3-6。

表3-6 认识计算机程序的问题列表及分工表

班级：_____ 小组编号：_____ 组长：_____

问题	信息的获取过程				
	选择信息来源	确定信息获取方法	评价信息	分工及时间安排	认知程度（认识/了解/描述/掌握）
问题1：什么是计算机程序？					
问题2：程序设计语言有哪些？					
问题3：					
问题4：					

微信公众号：电子课本大全

2. 表3-6中所要解决的每一个问题都可以制作成各类相关图表，如思维导图、概念图或者时间线表等。参考表3-7，思考其他问题的列表形式。

表3-7　高级程序设计语言的发展时间线表

时间	高级语言名称	功能简述
1954年	Fortran	Fortran，亦译为福传，是英文“FORmula TRANslator”的缩写，意思是“公式翻译器”，是世界上最早出现的计算机高级程序设计语言，广泛应用于科学和工程计算领域。Fortran语言以其特有的功能在数值、科学和工程计算领域发挥着重要作用。
1958年		
1959年		
……		

3.3.1　计算机程序

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指令序列，或者可被自动转换成代码化指令序列的符号化指令序列或者符号化语句序列。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事先转换成二进制数，并存入存储器中。计算机的工作过程实际上是周而复始地获取指令、执行指令的过程，如图3-10所示。

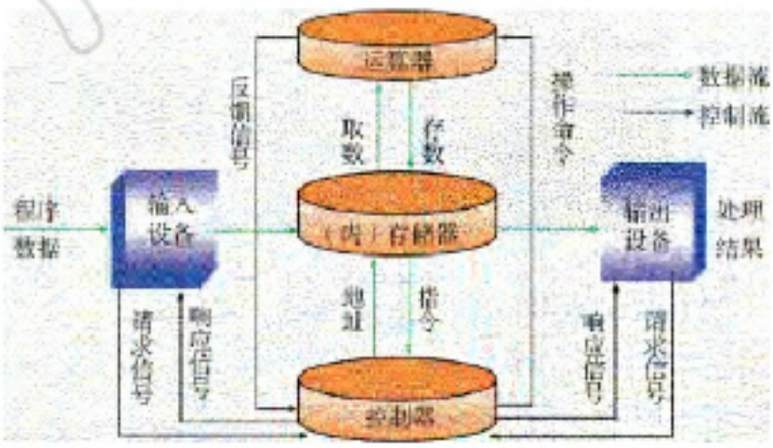


图3-10　计算机程序运行图

微信公众号：电子课本大全

3.3.2 计算机程序设计语言

在用计算机解决问题时，用自然语言、流程图或是伪代码所描述的解决问题的算法都不能被计算机直接执行，还必须将算法按照一定的规则编写成计算机能够识别和运行的程序。而人们编写程序时需要遵循的规则就是计算机语言规则。计算机程序设计语言，是指一组用来定义计算机程序的语法规则，通常简称为“编程语言”。它是一种被标准化的交流技巧，用于向计算机发出指令。正确地使用计算机程序设计语言，能让程序员准确地定义计算机所需要使用的数据，并精确地定义在不同情况下所应执行的命令。

计算机程序设计语言的发展，经历了从机器语言、汇编语言到高级语言的发展历程。

1. 机器语言

目前，计算机采用的物理器件主要是电子元件，但由于电子元件的物理特性，计算机只能识别“0”和“1”组成的二进制数。因此，二进制是计算机语言的基础。计算机发明之初，人们只能使用计算机的基础语言（图3-11）。因此，早期的程序设计语言是由“0”和“1”所表示的二进制代码指令组表示的。这样的语言是计算机能直接接收和执行的，通常被称为“机器语言”。机器语言是第一代计算机语言。

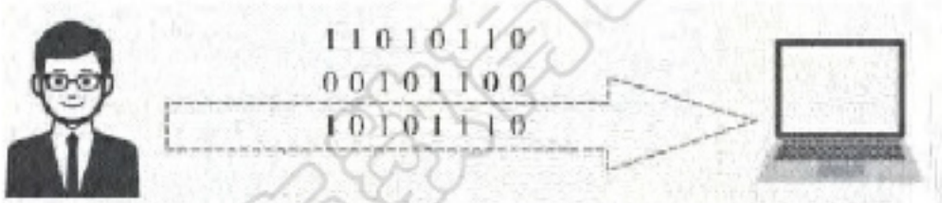


图3-11 机器语言示意图

不同计算机的机器指令系统会有所不同，但每一条机器指令都包含两个主要部分：操作（指出计算机应做什么）和被操作的对象（指出处理的数据或它的地址）。在这些指令的控制下，计算机可以实现最基本的算术运算和逻辑运算。例如，表3-8是用Intel 80386完成“9+8”加法运算的机器指令。

表3-8 用Intel 80386机器指令完成“9+8”的加法运算表

指令序号	机器指令	指令功能
1	10110000 00001001	把加数9送到累加器AL中。
2	00000100 00001000	把累加器AL中的内容与另一个数8相加，结果存在累加器AL中（即完成9+8的运算）。
3	11110100	停止操作。

微信公众号：电子课本大全

早期人们与计算机的通信都是通过机器语言编写的程序完成的，程序中的所有语句都被分解成一系列由0和1组成的代码记录在穿孔带上，如图3-12所示。



图3-12　穿孔纸带

这种机器语言所编写的程序难以被理解，程序设计任务也非常繁重，而且在程序出现错误需要修改时，效率更是低下。除此之外，由于每台计算机的指令系统往往各不相同，所以在一台计算机上执行的程序，如果想要在另一台计算机上执行，就必须另编程序，这就造成了重复工作。由此可见，机器语言与人类的自然语言之间存在着巨大的鸿沟。但由于使用的是针对特定型号计算机的语言，因此运算效率也是所有语言中最高的。

2. 汇编语言

为了让使用机器语言编写的程序更容易被理解，人们使用了一种类似英文缩略词且带有助记性符号的语言，来替代一个特定的指令二进制串，每条指令都和一条机器指令相对应，只是指令码和操作数都采用符号形式，这种程序设计语言就被称为汇编语言，即第二代计算机语言。例如，指令码用“ADD”代表加法，用“MOV”代表数据传递等。这样一来，人们就会比较容易读懂并理解程序，纠错及维护也会变得更加方便了。例如，表3-9列出了用汇编语言来实现“9+8”加法运算的有关指令。

表3-9　用汇编语言来实现“9+8”运算的有关指令表

指令序号	汇编语言指令	指令功能
1	MOV AL,9	把加数9送到累加器AL中。
2	ADD AL,8	把累加器AL中的内容与另一个数8相加，结果存在累加器AL中（即完成9+8的运算）。
3	HTL	停止操作。

但是，计算机是不能直接认识这些符号的，计算机还需要一个专门的语言翻译器，负责将程序中的每条语句都翻译成用二进制数表示的机器语言（图3-13）。



图3-13　汇编语言被翻译成机器语言示意图

微信公众号：电子课本大全

汇编语言同样与特定的机器及特定的微处理器相关，这也导致了汇编语言的移植性不好。这一点与机器语言相比差异不大，然而汇编语言程序的效率仍然很高。针对计算机特定硬件而编制的汇编语言程序，能准确发挥计算机硬件的功能和特长，其编写的程序不仅精练、质量高，而且易于理解，所以至今在一些领域仍是一种常用而强有力的软件开发工具。

3. 高级语言

人们在使用机器语言和汇编语言这两种语言与计算机交流的过程中，依然存在很大的障碍，而且对于程序的理解和调试仍然十分困难。于是，高级语言应运而生。高级语言接近于数学语言和人的自然语言，并且不再过度地依赖某种特定的机器或环境。第一种高级语言是Fortran语言，它主要用于科学和工程计算。在Fortran语言之后，出现了许多高级语言，如ALGOL、COBOL、Basic、LISP、SNOBOL、Pascal、C、PROLOG、Ada、VC、VB、Delphi、Java、Python等。例如，在计算“9+8”的问题时，若使用高级语言Python，则只需要语句“print(9+8)”就能完成，既简单又易于理解。

用高级语言编写的程序也不能直接被计算机所识别和执行，必须经过编译程序或解释程序将其翻译成机器语言（图3-14）。所谓编译程序是指计算机把高级语言程序的每一条语句都编译成机器语言，并保存成二进制文件后才运行的程序。解释程序是指计算机在执行高级语言程序时，逐条语句解释成机器语言并立即执行的程序。

由于高级语言抽象度高，源代码无须与硬件、系统底层操作对应，所以移植性非常好，理想的情况下甚至不必为不同的系统平台或者机器改动源代码。



图3-14 高级语言被翻译成机器语言示意图

高级语言的出现为计算机的应用开辟了广阔的前景，它的发展也经历了从早期语言到结构化程序设计语言、从面向过程到非过程化程序设计语言的过程。所谓非过程化程序设计语言，就是面向应用，即只需要告诉程序要干什么，程序就能自动生成算法，自动进行处理，这是高级语言的下一个发展目标，将给人们解决问题带来更大的便利。



以小组为单位开展交流活动，试用自己的语言表述程序与程序设计语言之间的关系。

微信公众号：电子课本大全

讨论

1. 一个高质量的程序应具备哪些条件？

2. 学习了本节的有关内容后，我们是否能够用生活中熟悉的事物来形容或比喻程序及其功能？

3. 高级语言的翻译程序有两种类型：编译程序和解释程序。试通过网络检索相关知识，比较编译型语言和解释型语言的优势和不足，并填写在表3-10中。

表3-10　编译型语言和解释型语言的比较表

语言种类	优势	不足
编译型语言		
解释型语言		

4. 高级语言的编译过程和解释过程是不同的，请通过查阅相关资料，将两个不同的执行过程用图表形式呈现出来，完成图3-15。



（a）编译过程　　（b）解释过程

图3-15　编译过程和解释过程

微信公众号：电子课本大全

5. 以程序设计语言的发展过程为主题，预测程序设计语言将来的发展趋势，就其便捷性及其功能等方面，大胆展望未来程序设计语言的特点，自拟调查提纲，写一篇调查报告。

拓展

自动计算思想发展回顾

一般而言，计算与自动计算要解决以下四个问题：①数据的表示；②数据的存储及自动存储；③计算规则表示；④计算规则的执行及自动执行。

算盘是一种计算工具，算盘上的珠子可以表示和存储数，计算规则是一套口诀，按照口诀拨动珠子可以进行四则运算。然而所有的操作都要靠人的大脑和手完成，因此算盘被认为是一种计算辅助工具，不能被归入自动计算工具范畴。

1642年，法国科学家帕斯卡发明了著名的帕斯卡机械计算机，首次确立了计算机器的概念。该机器用齿轮来表示和存储十进各数位上的数字，通过齿轮比来解决进位问题。低位上的齿轮每转动10圈，高位上的齿轮只转动1圈。机器可自动执行一些计算规则，“数”在计算过程中自动存储。帕斯卡机的意义是让人们认识到“用纯机械装置可代替人的思维和记忆”，开辟了自动计算的道路。如图3-16所示是一台手摇计算器。

1674年，莱布尼茨提出了“二进制”数的概念。1847年，英国的数学家布尔发表著作《逻辑的数学分析》和《思维规律的研究——逻辑与概率的数学理论基础》，创立了一门全新的学科——布尔代数，为百年后出现的数字计算机的开关电路设计提供了重要的数学方法和理论基础。



图3-16 手摇计算器

1938年，香农发表了著名的论文《继电器和开关电路的符号分析》，首次用布尔代数进行开关电路分析，并证明布尔代数的逻辑运算可以通过继电器电路来实现，明确地给出了实现加、减、乘、除等运算的电子电路的设计方法。这篇论文成为开关电路理论的开端。

1936年，阿兰·图灵提出了抽象计算模型——“图灵机”。一切可计算的函数都可用图灵机计算，反之亦然。

1940年，“控制论之父”维纳提出了计算机五原则：①不是模拟式，而是数字式；②由电子元件构成，尽量减少机械部件；③采用二进制，而不是十进制；④内部存放计算表，计算控制规则；⑤在计算机内部存储数据。

微信公众号：电子课本大全

1945年，美国宾夕法尼亚大学“莫尔小组”汲取前人关于计算机的结构、存储程序和二进制编码等研究思想，设计制造了世界上的一台早期计算机ENIAC（图3-17）。ENIAC是可编程的通用计算机，可以解决不同的问题。



图3-17　计算机ENIAC

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，安装相应的计算机程序语言软件，了解计算机程序设计语言的特点，进一步完善该项目方案中的各项学习活动，并参照项目范例的样式，撰写相应的项目成果报告。

成果交流

各小组运用数字化学习工具，将所完成的项目成果，在小组或班级上进行展示与交流，共享创造、分享快乐。

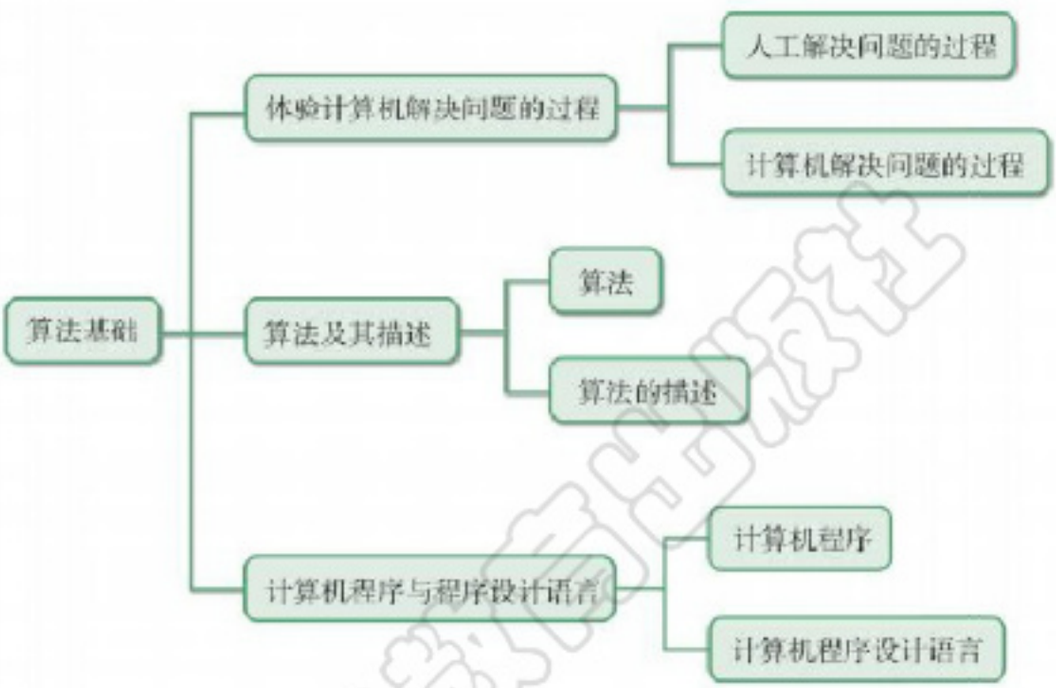
活动评价

各小组根据项目选题、拟订的项目方案、实施情况以及所形成的项目成果，利用教科书附录2的“项目活动评价表”，开展项目学习活动评价。

微信公众号：电子课本大全

本章扼要回顾

同学们通过本章学习，根据“算法基础”知识结构图，扼要回顾、总结、归纳学过的内容，建立自己的知识结构体系。



回顾与总结

微信公众号：电子课本大全

本章学业评价

同学们完成下列测试题（更多的测试题可以在教科书的配套学习资源包中查看），并通过“本章扼要回顾”以及本章的项目活动评价，综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。

1. 单选题

（1）人们利用计算机解决问题的基本过程为（　　）。

①调试运行程序　②分析问题　③设计算法　④问题解决　⑤编写程序

A. ①②③④⑤　　B. ②④③⑤①　　C. ④②③⑤①　　D. ②③⑤①④

（2）下面关于算法的描述，正确的是（　　）。

A. 算法不可以用自然语言描述

B. 算法只能用流程图来描述

C. 一个算法必须保证它的执行步骤是有限的

D. 算法的流程图表示法有零个或多个输入，但只能有一个输出

（3）计算机能直接识别、理解执行的语言是（　　）。

A. 汇编语言　　B. Python语言　　C. Basic语言　　D. 机器语言

2. 思考题

写出“求一个任意实数C的算术平方根G（精确到0.0001）”的算法设计思想。

3. 情境题

某自动售货机（如右图）是由微处理器控制的。张明同学在自动售货机选择了一种价值5元的饮料，然后在收款口放入10元。自动售货机接收货币后，在商品取出口给出了张明所要的饮料，并找了5元。

（1）说明自动售货机中微处理器在这次自动售货的大致过程。

（2）如果张明同学在自动售货机购物没有成功，请列出没有成功的三种可能原因。



自动售货机

微信公众号：电子课本大全

第四章
程序设计基础

在运用计算机解决实际问题的过程中，如何把求解问题的算法告知计算机，就需要有一种能够实现人机交互的语言，而这种语言就是计算机程序设计语言。

本章将通过“设计购买物品的最佳方案”项目，进行自主、协作、探究学习，让同学们了解Python程序设计语言的基础知识，掌握程序的顺序结构、选择结构和循环结构三种控制结构，学会使用程序设计语言实现简单算法，体验程序设计的基本流程，掌握程序调试与运行的方法，从而将知识建构、技能培养与思维发展融入运用数字化工具解决问题和完成任务的过程中，促进信息技术学科核心素养达成，完成项目学习目标。

- 程序设计语言的基础知识
- 运用顺序结构描述问题求解过程
- 运用选择结构描述问题求解过程
- 运用循环结构描述问题求解过程

微信公众号：电子课本大全

项目范例　设计购买纪念品的最佳方案

情境

如今，网上购物已经成为人们生活中的一部分。通过网购，人们不仅可以方便地在各家电商中获取商品的相关信息，而且可以对商品进行比质比价，从而以优惠的价格购得心仪的商品（图4–1）。为了购得既经济适用又美观大方的商品，人们就要根据需求设计好购买方案。



图4–1　网上购物

主题

设计购买纪念品的最佳方案

规划

根据项目范例的主题，在小组中组织讨论，利用思维导图工具，制订项目范例的学习规划，如图4–2所示。

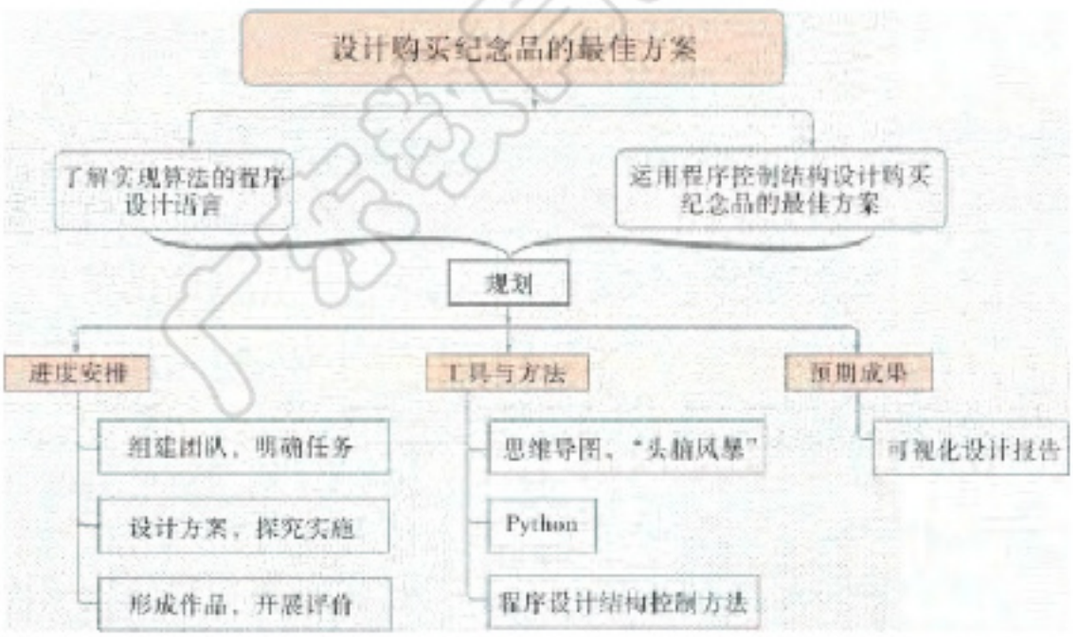


图4–2　“设计购买纪念品的最佳方案”项目学习规划

探究

根据项目学习规划的安排，通过调查、案例分析、文献阅读和网上资料搜索，开展“设计购买纪念品的最佳方案”项目学习探究活动，如表4–1所示。

微信公众号：电子课本大全

表4-1 “设计购买纪念品的最佳方案”项目学习探究活动

探究活动	学习内容	知识技能
了解实现算法的程序设计语言	Python的常量和变量。	了解程序设计语言的基础知识。
	Python的数据类型。	
	Python的运算符和表达式。	
	Python的函数。	
运用程序控制结构设计购买纪念品的最佳方案	运用顺序结构设计最佳方案。	掌握程序的顺序结构、选择结构和循环结构。 学会使用程序设计语言实现简单算法。 体验程序设计的基本流程。 掌握程序调试与运行的方法。
	运用选择结构设计最佳方案。	
	运用循环结构设计最佳方案。	
	综合运用三种结构设计最佳方案。	



实施项目学习各项探究活动，进一步设计购买纪念品的最佳方案。



在小组开展项目范例学习过程中，利用思维导图工具梳理小组成员在“头脑风暴”活动中的观点，建立观点结构图，运用多媒体创作工具（如演示文稿、在线编辑工具等）综合加工和表达，形成项目范例可视化学习成果，并通过各种分享平台发布，共享创造、分享快乐。例如，运用在线编辑工具制作的“设计购买纪念品的最佳方案”可视化报告，可以在教科书的配套学习资源包中查看，其目录截图如图4-3所示。



图4-3 “设计购买纪念品的最佳方案”可视化报告目录截图

微信公众号：电子课本大全

评价

根据教科书附录2的“项目活动评价表”，对项目范例学习过程和学习成果在小组或班级上进行交流，开展项目学习活动评价。

项目选题

同学们以3~6人组成一个小组，选择下面一个参考主题，或者自拟一个感兴趣的主题，开展项目学习。

1. 设计购买电信服务的最佳方案
2. 设计购买家用电器的最佳方案
3. 设计购买图书的最佳方案

项目规划

各小组根据项目选题，参照项目范例的样式，利用思维导图工具，制订相应的项目方案。

方案交流

各小组将完成的方案在全班进行展示交流，师生共同探讨、完善相应的项目方案。

4.1 程序设计语言的基础知识

在运用计算机程序求解问题时，需要有一种能够实现人机交流的语言，即程序设计语言。人们通过这种语言把求解的方法和步骤告诉计算机，计算机则按照程序的流程自动求解问题。

例如，求解购买纪念品的最佳方案问题，需要知道各商品的单价，才能设计出最佳方案。现在设总预算为S，需要购买10个不同品种的商品数量为x_1，x_2，…，x_{10}，则求解问题的思路如下：

（1）上网搜索各个商品的价格。

（2）求出各个商品的最低单价a_1，a_2，…，a_{10}。

微信公众号：电子课本大全

（3）令$E=S-(a_1\times x_1+a_2\times x_2+\cdots+a_{10}\times x_{10})$，求$x_1$，$x_2$，…，$x_{10}$的值（其中$x_1$，$x_2$，…，$x_{10}$为正整数），使$E$最小（$E\geqslant 0$）。

在编写上述求解算法程序前，首先应掌握程序设计的一些基础知识，包括程序的常量和变量、数据类型、运算符与表达式及函数等。我们以Python语言为例。

4.1.1 Python的常量和变量

数据总是以常量或变量两种形式出现在程序中。

1. 常量

常量是指在程序运行过程中其值始终不发生变化的量，通常是固定的数值或字符串。例如，45，-30.2001，'Hello'，'Good'等都是常量。常量可以在程序中直接使用，如语句“x=30*40”表示将30乘40的结果赋值给x，其中30和40都是常量，可以直接在程序中使用以表示数值30和40。

2. 变量

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”（又称“名字”）和变量的“内容”（又称“值”）两部分组成。

Python程序设计语言规定标识符由字母(A～Z，a～z)、数字(0～9)、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，pai等。在标识符中，字母大小写是有区别的，如“value”与“Value”是两个不同的标识符。

在Python中，如果使用非法标识符，在运行程序时会有相关报错信息。例如，以下是报错非法标识符的程序：

```
>>> a1=3
>>> a1
3
>>> 1a=3        ←——— 不能以数字开头
SyntaxError: invalid syntax
```

变量的内容在程序运行过程中是可以变化的。例如，一个变量的名字为E，其内容可以为50，也可以为70。变量就像一个房间一样，变量名相当于房间的房间号，内容相当于居住于房间的不同的人员等。

需要注意的是，Python是动态语言，动态语言的变量不需要进行定义数据类型，其数据类型可以不固定，是动态的，需要根据其赋予的值来确定其数据类型。与之对应的是静态语言，静态语言在定义变量时必须指定变量类型，如果赋值的时候类型不匹配，就会报

微信公众号：电子课本大全

错。因此，和静态语言相比，动态语言更灵活。同样一个变量是可以重复赋值的，而且可以被赋予不同类型的值，如表4-2所示。同时，Python支持多种数据类型，在计算机内部，可以把任何数据都看成一个“对象”，而变量就是在程序中用来指向这些数据对象的，对变量赋值就是把数据和变量关联起来。

表4-2　变量动态赋值代码示例

变量动态赋值代码	结果
a=5 print(a) a='ABC' print(a)	5 ABC

4.1.2　Python的数据类型

计算机能处理的远不只数值数据，还可以是文本、图形、音频、视频、网页等各种各样的数据。不同的数据，需要定义不同的数据类型。Python能够直接处理的常用数据类型有Number（数值）、String（字符串）、List（列表）、Tuple（元组）、Sets（集合）、Dictionary（字典）等。表4-3中列出了Python常用的基本数据类型。

表4-3　Python常用的基本数据类型

数据类型	类型说明
Number（数值）	Python主要支持以下三种不同的数值类型： 整型（int），通常被称为整型或整数。在Python 3.x版本中，整型数没有限制大小。 浮点型（float）。浮点型由整数部分与小数部分组成，浮点型也可以使用科学计数法表示，如2.5e2 =2.5 × 10^2=250。 复数（complex）。复数由实数部分和虚数部分构成，可以用a+bj或者complex(a, b)表示，复数的实部a和虚部b都是浮点型。
String（字符串）	字符串是以单引号或双引号括起来的任意文本，如'abc' "xyz"等。字符串的索引从0开始，-1 为从末尾开始的位置。
List（列表）	列表是写在方括号“[ ]”之间、用逗号分隔开的元素列表。例如，list1=['Good', 'Run', 1997, 2000]和list2=[1, 2, 3, 4, 5, 6, 7]。 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。 Python已经内置确定列表的长度以及确定最大和最小的元素的方法。

微信公众号：电子课本大全

（续表）

数据类型	类型说明
Tuple（元组）	Python 的元组与列表类似，不同之处在于元组的元素不能修改。元组使用小括号，创建元组只需要在括号中添加元素，并使用逗号隔开。例如，tup1=('Good','Run',1997,2000)和tup2=(1,2,3,4,5)。元组与字符串类似，可以被索引且下标索引从0开始，-1为从末尾开始的位置，也可以进行截取。
Sets（集合）	集合是一个无序不重复元素的序列。基本功能是进行成员关系测试和删除重复元素。可以使用大括号{ }或者set( )函数创建集合。创建一个空集合必须用 set( ) 而不是{ }，因为{ }是用来创建一个空字典的。
Dictionary（字典）	字典是无序的对象集合。与列表的区别在于：列表是有序的对象集合。字典当中的元素是通过键来存取的，而不是通过偏移存取。字典用{ }标识，它是一个无序的“键（key）：值（value）”对集合。键必须使用不可变类型。在同一个字典中，键必须是唯一的。

探究活动

交流

各小组讨论、交流相应项目选题实施过程中可能用到的数据类型。

讨论

各小组对“合理定义数据类型可以优化程序的运行”这一观点进行交流和讨论，并举例说明。

4.1.3 Python的运算符与表达式

在Python程序设计语言中对常量或变量进行运算或处理的符号称为运算符，参与运算的对象称为操作数。常用的运算符有算术运算符、逻辑运算符、关系运算符、赋值运算符、成员运算符等。由运算符、操作数连接而成的式子称为表达式。操作数可以是常量、变量、函数等。

1. 算术运算符与算术表达式

算术运算符包括“+”“-”“*”“/”“%”“**”和“//”，分别表示加、减、乘、除、求余、幂运算和整除。例如，“7%3=1”表示7除以3的余数，即1；“10**20”表示10的20次方，即10^{20}。由算术运算符连接的表达式称为算术表达式。例如，“a+3*c”“（x+y）*n-7”等。

微信公众号：电子课本大全

2. 逻辑运算符与逻辑表达式

逻辑运算符包括“and”（与）、“or”（或）、“not”（非）共三个。由逻辑运算符连接而成的表达式称为逻辑表达式。逻辑表达式的值为True或False，可对应于整数1或0直接参与其他运算。逻辑表达式的运算顺序要根据逻辑运算符的优先级而定。假设变量a为0，b为1，c为2，则对应的逻辑表达式及其值如表4-4所示。

表4-4　逻辑表达式及对应运算结果

运算符	逻辑表达式	描述	示例
and	x and y	“与”。如果x为 False，x and y 返回 False，否则它返回 y 的计算值。	(a and b) 返回0。 (c and b) 返回1。
or	x or y	“或”。如果x是 True，它返回 x 的值，否则它返回 y 的计算值。	(a or b) 返回1。 (c or b) 返回2。
not	not x	“非”。如果 x为 True，返回 False；如果 x 为 False，它返回 True。	not(a and b) 返回 True。

3. 关系运算符与关系表达式

判断数据大小关系的运算符称为关系运算符，关系运算符有“>”（大于）、“>=”（大于等于）、“<”（小于）、“<=”（小于等于）、“==”（等于）、“!=”（不等于）。

由关系运算符连接而成的表达式称为关系表达式。关系表达式的值为True或False，对应于整数1或0直接参与其他运算。例如，关系表达式“5>=3”的值为True，“9<8”的值为False。

4. 赋值运算符与赋值表达式

在Python中对变量的赋值通过赋值运算符“=”来完成。赋值号不是等号，具有方向性，由右边赋给左边，且赋值号左边只能是变量，而不能是常量或表达式。Python的赋值运算符有多种，具体描述如表4-5所示。

表4-5　Python语言的赋值运算符

赋值运算符	描述	示例	赋值运算符	描述	示例
=	直接赋值。	$x=y$。	/=	除法赋值。	$x/=y$相当于$x=x/y$。
+=	加法赋值。	$x+=y$相当于$x=x+y$。	%=	求余赋值。	$x\%=y$相当于$x=x\%y$。
-=	减法赋值。	$x-=y$相当于$x=x-y$。	**=	指数幂赋值。	$x**=y$相当于$x=x**y$。
*=	乘法赋值。	$x*=y$相当于$x=x*y$。	//=	整除赋值。	$x//=y$相当于$x=x//y$。

5. 成员运算符

除了前面的运算符，Python用成员运算符判断序列中是否有某个成员。Python的成员运算符描述如表4-6所示。

微信公众号：电子课本大全

表4-6 Python语言的成员运算符

成员运算符	描述	示例
in	x in y。如果x是序列y的成员，则计算结果为True，否则为False。	3 in [1,2,3,4]，计算结果为True。 5 in [1,2,3,4]，计算结果为False。
not in	x not in y。如果x不是序列y的成员，则计算结果为True，否则为False。	3 not in [1,2,3,4]，计算结果为False。 5 not in [1,2,3,4]，计算结果为True。

6. 运算符的优先级

当一个表达式中有多种运算符时，运算的顺序将根据运算符的优先级由高到低进行运算。如果是同级运算则按从左到右的顺序进行。Python常用运算符的优先级如表4-7所示。

表4-7 Python常用运算符的优先级

优先级	运算符	功能及说明	优先级	运算符	功能及说明
1	**	指数运算。	6	=、%=、/=、//= -=、+=、*=、**=	赋值运算。
2	*、/、%、//	乘、除、求余、整除。	7	is、is not	标识运算符。
3	+、-	加运算、减运算。	8	in、not in	成员运算符。
4	>、>=	大于、大于等于。	9	not	逻辑非运算。
	<、<=	小于、小于等于。	10	and	逻辑与运算。
5	==、!=	等于、不等于。	11	or	逻辑或运算。



各小组思考程序设计中的运算符与表达式跟数学中的运算符与表达式的异同。

4.1.4 Python的函数

函数是Python语言程序的基石，是组织好的、可重复使用的、用来实现单一或相关联功能的代码段。Python 默认带有一些常用的函数，如求绝对值函数、求和函数、求最大值和最小值函数等。Python中常用的函数如表4-8所示。

表4-8 Python常用函数

函数名称	使用说明
abs(x)	返回x的绝对值。
len(x)	计算变量x的长度，但x必须是可以计算长度的类型。

微信公众号：电子课本大全

（续表）

函数名称	使用说明
min(x)	返回变量x的最小值。
max(x)	返回变量x的最大值。
int(x)	把变量x转换成整型。
float(x)	将变量x转换成浮点型。
sum(x)	计算变量x中元素的总和。
bin(x)	把数值x转换为二进制数字。
bool(x)	x如果是False、None或是空值就返回False。
chr(x)	返回整数x对应的ASCII字符。
divmod(a,b)	返回a/b的商和余数，以元组的方式返回。
all(x)	x中的所有元素都是True才会返回True。
any(x)	x中的所有元素中只要有一个是True就会返回True。
input(x)	显示出信息x，并要求输入数据。
print( )	输出函数。

拓展

Python程序设计开发环境

Python是一种面向对象的解释型计算机程序设计语言。Python语言的安装包内置了开发环境IDLE。为了快捷、方便地掌握Python语言编程，这里介绍另一个开源的Python发行版本Anaconda及其程序设计开发环境Jupyter notebook，它包含了Python语言完整的内核、功能强大的标准库和第三方扩展库等。

按照教科书配套学习资源包“关于Python语言编程环境及各章程序使用说明”的指引方法，获取并运行文件安装Anaconda。安装成功后，双击“Anaconda Navigator”图标，选择“Jupyter notebook”，即可进入Python程序设计开发环境，其界面如图4-4所示。



图4-4　Jupyter notebook 界面

微信公众号：电子课本大全

Python程序设计开发环境Jupyter notebook的使用操作，可查看教科书的配套学习资源包。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，开展以下活动。

1. 分析相应的程序设计语言的基本知识。
2. 分析所选程序设计语言的特点。

4.2 运用顺序结构描述问题求解过程

将程序的基本结构按一定规律组成并对算法进行描述，是保证和提高算法的质量，以及编写优质程序的基础。按语句的先后顺序执行的结构称为“顺序结构”。顺序结构是结构化程序的基本结构之一。

探究活动

问题

项目小组成员在某网购站点上获取到如下信息：笔记本1的单价是3元，笔记本2的单价是5元，如果两种笔记本同时购买，价格可以打八折。计算一下，购买笔记本1和笔记本2各n本，可以优惠多少元？（以元为单位，四舍五入到小数点后两位。）

4.2.1 数据的输入与输出

在程序设计中，数据的输入输出是最基本的操作之一。在Python中，数据的输入输出通过函数来实现。

1. 数据的输入

在Python程序设计语言中，数据输入通过input( ) 函数来实现，input( ) 的一般格式为：

```
<变量>=input('提示信息：')
```

该函数返回输入的对象，可输入数字、字符串和其他任意类型对象。在Python 3.x中，不论用户输入什么类型的数据，input( )函数的返回结果都是字符串，如果需要可将其转换为相应的类型再处理。

微信公众号：电子课本大全

2. 数据的输出

在Python 3.x中，数据输出可通过print( )函数来实现，print( )的一般格式为：

```
print( [object, …][, sep=' '][,end='\n'] )
```

说明：

（1）方括号中的项是可选的，可以省略，如省略则取系统的默认值。

（2）object是要输出的对象，可以是常量、变量或表达式等。

（3）sep后面的空格（可以指定为其他字符）表示每个输出对象之间的分隔符，如果缺省的话，默认值是一个单个的空格。

例如，“print(4,5,6,sep='***')”，则输出结果为“4***5***6”。

（4）end后面的字符串含义为输出文本尾的一个字符串，如果缺省的话，默认值是一个\n换行符。如果设为其他字符，如end=' '，则输出当前行的所有内容后，在末尾加一个空格，不换行接着输出下一个print( )的输出对象。

例如，print( )输出示例程序如下：

```
print("hello",end=' ')
print("world")
```

输出为“hello world”。

（5）print( )函数支持参数格式化，与C语言的printf类似。

4.2.2　顺序结构的应用

顺序结构是最常用、最简单的程序结构，其语句执行顺序是自上而下依次执行的。如第三章第二节“三种基本控制结构”所述，只要按照解决问题的顺序写出相应的语句，便可设计出程序的顺序结构。顺序结构的流程图如图4–5所示。

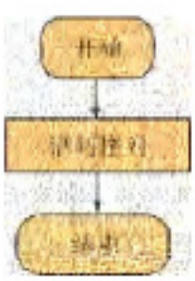


图4–5　顺序结构的流程图



各小组编程解决本节探究活动“购买笔记本的费用优惠”的问题。

微信公众号：电子课本大全



按照“3.1.2 计算机解决问题的过程”，解决上述问题的方法有多种，其中一种方法步骤如下：

（1）分析问题。

先求出不使用优惠而单独购买所需费用，再求出使用优惠方案购买所需费用，两者相减即为节省的费用。

（2）设计算法。

根据上面的分析，设计算法如下：

①输入n的值；

②令$a=3\times n$；

③令$b=5\times n$；

④令$y=(a+b)\times(1-0.8)$；

⑤输出结果y；

⑥结束。

其顺序结构流程图如图4-6所示。

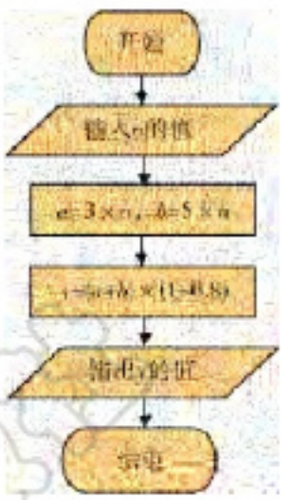


图4-6 实现购买笔记本优惠问题的顺序结构流程图

（3）编写程序。

根据上面设计的算法，可以编写出该问题的程序代码（见配套学习资源包“第四章\课本素材\程序4-1”）如下：

程序4-1

```
n=int(input("输入购买的笔记本1和笔记本2的本数均为："))
a=3*n
b=5*n
y=(a+b)*(1-0.8)
print("购买笔记本1和笔记本2的数量均为：",n,"本")
print("可节省的费用为：","%.2f"%y,"元")
```

（4）调试程序。

将程序4-1的程序代码输入计算机，然后运行程序，输入n的值为50，得到结果如下所示。经验算，结果正确。

```
输入购买的笔记本1和笔记本2的本数均为：50
购买笔记本1和笔记本2的数量均为：50本
可节省的费用为：80.00元
```


微信公众号：电子课本大全

交流

各小组讨论、交流以下问题：

（1）解决“购买笔记本的费用优惠”问题的其他方法。

（2）顺序结构的特点。

（3）Python程序设计语言中，还提供了位运算。查找相关资料，与项目小组成员分享、交流关于位运算的用法。

4.3 运用选择结构描述问题求解过程

4.3.1 if语句的应用

在解决问题的过程中，常常需要对事物进行判断和选择。在程序设计中，主要使用条件语句（if语句）来实现判断和选择。if语句在使用中包括单分支if语句和双分支if语句。

1. 单分支if语句

（1）单分支if语句的格式：

```
if(条件表达式):
    语句A
```

在使用单分支if语句时需要注意：①if（条件表达式）后面加冒号“：”；②当语句A是由多个语句构成的复合语句时，所有语句严格遵循相同的缩进。

（2）单分支if语句的执行过程。

如果条件表达式为真，则执行语句A，然后继续往下执行；如果条件表达式为假，则不执行语句A，然后继续往下执行。其执行过程如图4-7所示。

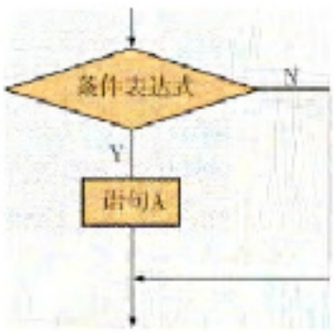


图4-7　if语句的执行过程

问题

项目小组成员在某购物网站上了解到某种笔记本搞优惠活动，规定一次购买6本以上（含6本）10本以下（含10本）打九折，一次购买11本以上（含11本）打八折。怎样根据笔记本的单价和购买数量编写程序计算总费用？

思考

上述问题给出了什么条件，需要解决什么问题，解决问题的方法和步骤是怎样的？

微信公众号：电子课本大全



解决上述问题的方法有多种，其中一种步骤如下：

（1）分析问题。

设项目小组购买笔记本的数量为s本，笔记本的单价为p元，折扣系数为j，总费用为t。根据题意，有：

如果$s<6$，则$j=1.0$；

如果$6\leqslant s\leqslant 10$，则$j=0.9$；

如果$s\geqslant 11$，则$j=0.8$；

又$t=s\times p\times j$，求出t的值即为所求的总费用。

（2）设计算法。

根据上面对问题的分析，解决问题的算法描述如下：

①输入s，p的值；

②如果$s<6$，则$j=1.0$；

③如果$6\leqslant s\leqslant 10$，则$j=0.9$；

④如果$s\geqslant 11$，则$j=0.8$；

⑤$t=s\times p\times j$；

⑥输出t的值；

⑦结束。

算法的流程图如图4-8所示。

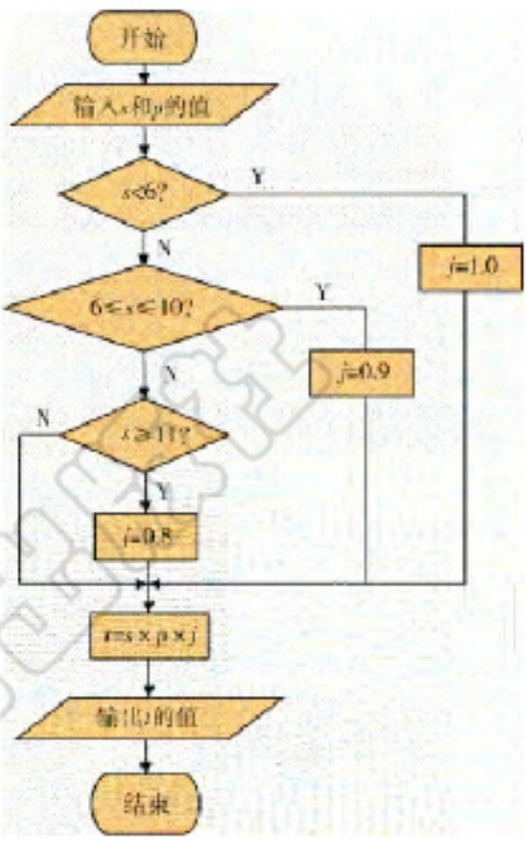


图4-8 网购笔记本的金额计算流程图

（3）编写程序。

根据上面设计的算法，编写出解决该问题的程序（见配套学习资源包“第四章\课本素材\程序4-2”）如下：

程序4-2

```
s=input("请输入购买的笔记本的数量（多少本）：")
s=int(s)
p=float(input("请输入笔记本的单价（每本多少元）："))
if s<6:
    j=1.0
if s>=6 and s<=10:
    j=0.9
if s>=11:
    j=0.8
t=s*p*j
print("总费用为：","%.1f"%t,"元")
```


微信公众号：电子课本大全

（4）调试程序。

将程序4-2输入计算机进行调试运行，并输入s的值为100，p的值为3.2，得到结果如下：

```
请输入购买笔记本的数量（多少本）：100
请输入笔记本的单价（每本多少元）：3.2
总费用为：256.0元
```

2. 双分支if语句

（1）双分支if语句的格式：

```
if (条件表达式):
    语句A
else:
    语句B
```

在使用双分支if语句时需要注意：①if（条件表达式）后面加冒号“：”；②当语句A和语句B是由多个语句构成的复合语句时，所有语句严格遵循相同的缩进。

（2）双分支if语句的执行过程。

如果条件表达式为真，则执行语句A，然后继续往下执行。如果条件表达式为假，则执行语句B，然后继续往下执行。其执行过程如图4-9所示。

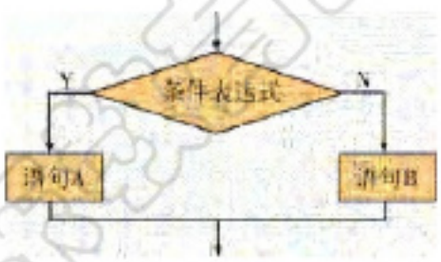


图4-9　双分支if语句执行过程



项目小组成员在某购物网站上看到三种笔记本的优惠价各不相同，笔记本1、笔记本2和笔记本3每本分别优惠价为a元、b元和c元，编程输出其优惠价的最大值。



“求网购笔记本的费用”和“求网购笔记本优惠的最大值”有何区别？在“求网购笔记本优惠的最大值”中，求最值的方法有哪些，该如何编程实现？



解决上述问题的方法有多种，其中一种的方法步骤如下：

微信公众号：电子课本大全

（1）分析问题。

对于输入的三个数a，b，c，设其中的最大数为m，先找出a和b中较大的数赋给m，然后将m与c比较，从而找出三个数中最大的数。

（2）设计算法。

根据上面的分析，解决问题的算法描述如下：

①输入三个数a，b，c；

②如果$a>b$，则$m=a$；否则$m=b$；

③如果$c>m$，则$m=c$；

④输出m的值；

⑤结束。

（3）编写程序。

根据前面设计的算法，编写出解决该问题的程序（见配套学习资源包“第四章\课本素材\程序4-3”）如下：

程序4-3

```
print("请输入三种笔记本的优惠价（元）：")
a,b,c=input().split()
a=float(a)
b=float(b)
c=float(c)
if a>b:
    m=a
else:
    m=b
if c>m:
    m=c
print("优惠价最大的为：",m,"元")
```

（4）调试程序。

将程序4-3输入计算机进行调试运行，并输入a，b，c的值分别为32.5，63.4，78，结果如下：

```
请输入三种笔记本的优惠价（元）：
32.5 63.4 78
优惠价最大的为：78.0元
```

3. if语句的嵌套

if语句允许嵌套，即if语句中的语句A和语句B还可以是if语句。

微信公众号：电子课本大全

4.3.2　if...elif...else语句的应用

应用if条件语句可以快速地实现程序的分支选择，但当分支比较多时，用if...elif...else语句来实现比较简便。

（1）if...elif...else语句格式如下：

```
if (表达式1):
    语句1
elif (表达式2):
    语句2
    ……
elif (表达式N):
    语句N
else:
    语句N+1
```

（2）if...elif...else语句的执行过程。

if...elif...else语句的执行过程如图4-10所示。

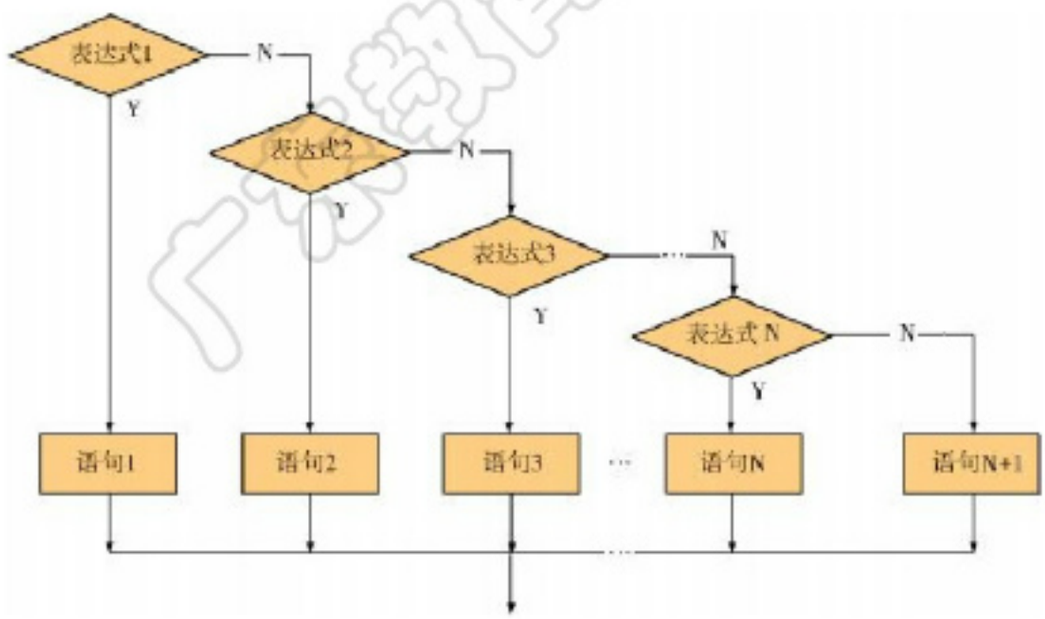


图4-10　if...elif...else语句执行过程

微信公众号：电子课本大全



项目小组成员经过比较，打算在某购物网站购买三种笔记本作为奖品和纪念品，它们的单价分别为6元、5元和4元。若项目小组共有经费x元，他们希望买尽可能多的笔记本作为奖品和纪念品，并能把经费用完。那么，这三种笔记本各可以买多少本？



根据上述问题给出的条件，思考如何编程找出符合条件的采购方案。



解决上述问题的方法有多种，其中一种方法步骤如下：

（1）分析问题。

因为要买尽量多的笔记本，所以选择4元的笔记本，可以买到的本数最多为$\frac{x}{4}$本，但不一定能刚好把经费用完。因此，项目小组可以按如下方案购买笔记本：

如果买完$\frac{x}{4}$本单价为4元的笔记本，还剩1元，则单价为4元的笔记本少买1本，换成买一本单价为5元的笔记本即可；如果买完$\frac{x}{4}$本单价为4元的笔记本，还剩2元，则单价为4元的笔记本少买1本，换成买一本单价为6元的笔记本即可；如果买完$\frac{x}{4}$本单价为4元的笔记本，还剩3元，则单价为4元的笔记本少买2本，换成买一本单价为5元的和一本单价为6元的笔记本即可。

（2）设计算法。

根据分析，可先设购买笔记本的经费总费用共为x元，购买单价为6元、5元、4元笔记本的数量分别为m，n，k，具体算法描述如下：

①输入x的值；

②$k=x//4$；

③$y=x\%4$；

④如果y的值为0，则$m=0$，$n=0$；

⑤如果y的值为1，则$m=0$，$n=1$，$k=k-1$；

⑥如果y的值为2，则$m=1$，$n=0$，$k=k-1$；

⑦如果y的值为3，则$m=1$，$n=1$，$k=k-2$；

⑧输出m，n，k的值；

⑨结束。

（3）编写程序。

根据设计的算法，编写出解决该问题的程序（见配套学习资源包“第四章\课本素材\程序4-4”）如下：

微信公众号：电子课本大全

程序4-4

```
x=int(input("输入经费的值："))
k=x//4
y=x%4
if y==0:
    m=0
    n=0
elif y==1:
    m=0
    n=1
    k=k-1
elif y==2:
    m=1
    n=0
    k=k-1
elif y==3:
    m=1
    n=1
    k=k-2
print("6元的笔记本为：%d 本"% m)
print("5元的笔记本为：%d 本"% n)
print("4元的笔记本为：%d 本"% k)
```

（4）调试程序。

将程序4-4输入计算机进行调试运行，如输入x的值为398，结果如下：

```
输入经费的值：398
6元的笔记本为：1本
5元的笔记本为：0本
4元的笔记本为：98本
```



各小组讨论以下问题：

（1）if语句和if...elif...else语句在格式、功能、应用等方面的异同。

（2）选择结构的特点。

微信公众号：电子课本大全



某学习小组的三名同学为判断某年是否为闰年的问题，各编写程序如下：

程序1：

```
x=int(input("请输入年份："))
year=x
if x%100==0:
    if x%400==0:
        print("%d年是闰年"%d x)
    else:
        print("%d年不是闰年"%d x)
elif x%4==0:
    print("%d年是闰年"%d x)
else:
    print("%d年不是闰年"%d x)
```

程序2：

```
x=int(input("请输入年份："))
year=x
if x%100==0:
    if x%400==0:
    x=x/100
if x%4>0:
    print("%d年不是闰年"%d year)
else:
    print("%d年是闰年"%d year)
```

程序3：

```
x=int(input("请输入年份："))
year=x
if x%100==0 or (x%100>0 and x%4==0 ):
    print("%d年是闰年"%d x)
else:
    print("%d年不是闰年"%d x)
```

各小组完成以下实践内容：

（1）上机调试程序，分别验证三个程序的正确性。

（2）在这三个程序中，分别采用了不同的表达式作为判断闰年的条件。分析、讨论哪一种更好。

微信公众号：电子课本大全

拓展

Python语言程序调试

Python 是解释型语言，翻译一行就执行一行，因此在运行程序的过程中，只要任何一行出现错误，程序就会立即显示出错误信息并停止执行。此种方式虽然在调试上比较低效（因为每次只会发现一个错误），但是却很容易知道错误发生的原因以及位置。

%debug命令用于调试代码，它有两种用法：一种是在执行代码之前设置断点进行调试；第二种则是在代码抛出异常之后，执行%debug命令查看调用堆栈。如图4-11是代码执行后出现的异常信息。

```
import math

def sinc(x):
    return math.sin(x) / x

[sinc(x) for x in range(5)]
```

```
ZeroDivisionError                         Traceback (most recent call last)
<ipython-input-28-9b09eaad97fe> in <module>()
      4     return math.sin(x) / x
      5
----> 6 [sinc(x) for x in range(5)]

<ipython-input-28-9b09eaad97fe> in sinc(x)
      2
      3 def sinc(x):
----> 4     return math.sin(x) / x
      5
      6 [sinc(x) for x in range(5)]

ZeroDivisionError: float division by zero
```

图4-11　异常提示

上面的程序中出现了ZeroDivision Error异常，下面用%debug查看调用堆栈。在调试模式下可以使用ipdb模块提供的调试命令，例如用p x显示变量x的值，如图4-12所示。

```
%debug
> <ipython-input-28-9b09eaad97fe>(4)sinc()
      3 def sinc(x):
----> 4     return math.sin(x) / x
      5

ipdb> p x
0
ipdb> q
```

图4-12　%debug调试

微信公众号：电子课本大全

调试程序时，我们还可以先设置断点，然后运行程序。但是%debug的断点需要指定文件名和行号，使用起来并不方便。用%%func_debug单元命令，可以中断指定运行的函数。在如图4-13所示的例子中，程序是在np.unique( )的第一行中断运行，然后通过输入n命令单步运行程序，最后输入c继续运行。

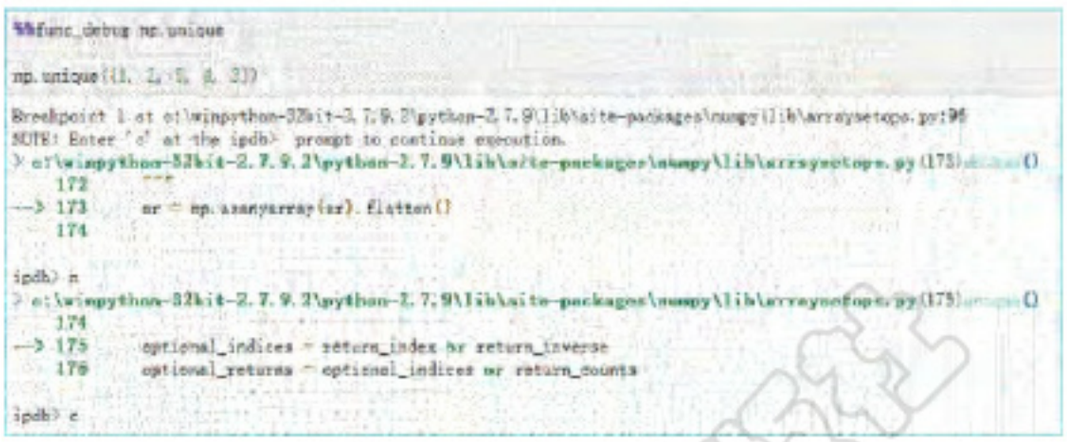


图4-13 %%func_debug调试

4.4 运用循环结构描述问题求解过程

程序的循环结构是程序设计中常用到的基本结构之一。循环是让计算机自动完成重复工作的常见方式。for循环、while循环、循环嵌套、循环控制等都是程序循环结构的主要内容。

4.4.1 for循环的应用

for循环语句的一般格式如下：

```
for 循环变量 in 列表:
    语句或语句组
```

for循环依次把列表中的每个元素逐个迭代出来。冒号“：”代表下一行是循环的第一行，但是循环体在书写时需要缩进。在for循环中，从for语句末尾的冒号开始至其下面没有缩进的代码之间都是其包含的需要重复执行的部分。因此，for循环所包含的语句是以缩进为标志的。具体如表4-9所示。

微信公众号：电子课本大全

表4-9　缩进循环体示例

是否缩进	循环体示例代码	示例结果
有缩进	for value in [1,2,3,4,5]: 　　print(value) 　　print('good')	1 good 2 good 3 good 4 good 5 good
无缩进	for value in [1,2,3,4,5]: 　　print(value) print('good')	1 2 3 4 5 good

需要注意的是，列表可以有几种表达方式：

（1）直接罗列的方式。

如表4-9中代码“for value in [1,2,3,4,5]:”就是直接罗列的方式。

（2）先定义后使用的方式，如表4-10所示。

表4-10　先定义后使用的列表示例

循环体先定义后使用示例代码	示例结果
values=[1,2,3,4,5] for value in values: 　　print(value) print('good')	1 2 3 4 5 good

（3）使用range( )函数。

range( )函数是让Python从指定的第一个值开始数，并在到达所指的第二个值后停止，因此不包含第二个值。如range(1,6)其实是生成一个1~6的整数（不包含6），如表4-11所示。

微信公众号：电子课本大全

表4-11 range( )函数示例

range ( ) 函数示例代码	示例结果
for value in range(1,6): print(value) print('good')	1 2 3 4 5 good

使用range( )函数时，还可指定步长，如表4-12所示。

表4-12 range( )函数指定步长示例

range ( ) 函数指定步长示例代码	示例结果
for value in range(2,11,2): print(value)	2 4 6 8 10

当然，使用range( )函数时，数值列表也可以是降序的，如range(10,1,-2)就是生成10至2的偶数（降序），具体如表4-13所示。

表4-13 range( )函数降序功能示例

range ( ) 函数降序示例代码	示例结果
for value in range(10,1,-2): print(value)	10 8 6 4 2



假如项目小组经过讨论，决定先网购单价为1元至10元的十种笔记本各1本，且每一种只买1本，那么该如何编程计算一共要花多少元？



如何用for循环语句编程解决上述“网购物品累计费用计算”问题？

微信公众号：电子课本大全

观察

解决上述问题的方法有很多，其中一种方法步骤如下：

（1）分析问题。

这个问题相当于要计算从1到10的累加值。1~10的累加就是计算1+2+…+10，可以用for循环结构编程来解决这个问题。

（2）设计算法。

根据上述分析，可设计如下算法：需要有一个循环变量从1变化到10，将该变量命名为*i*；还需要另外一个临时变量不断和该变量进行加法运算，并记录运算结果，将临时变量命名为*sum*；变量*i*每增加1时，就和变量*sum*进行一次加法运算，变量*sum*记录的是累加的结果。该算法流程图如图4-14所示。

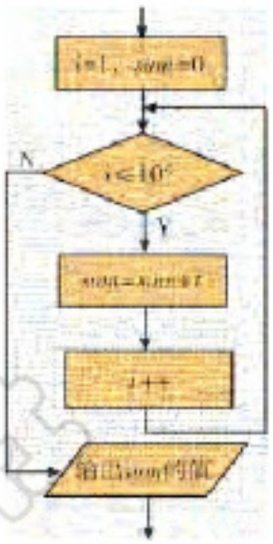


图4-14　累加算法流程图

（3）编写程序。

根据上面设计的算法，编写出解决该问题的程序（见配套学习资源包“第四章\课本素材\程序4-5”）如下：

程序4-5

```
sum=0
for i in range(1,11):
    sum=sum+i
print("the result:",sum)
```

（4）调试程序。

将程序4-5输入计算机进行调试运行，所得结果为“the result:55”。

4.4.2　while循环的应用

while循环是当型循环，只要条件满足，就不断循环；否则，退出循环。while循环语句的格式：

```
while (表达式):
    语句或语句组
```

（1）表达式一般是一个关系表达式或一个逻辑表达式，表达式的值应该是一个逻辑值真（True）或假（False）。

（2）当表达式的值为真时，执行循环体语句；当表达式的值为假时，退出循环，执行循环体外的下一条语句（即while后面没有缩进的第一条语句）。

（3）每次循环都是执行完循环体语句后回到表达式处重新开始判断，重新计算表达

微信公众号：电子课本大全

式的值。一旦表达式的值为假就退出循环，为真时就继续执行循环体语句。

（4）循环体中必须有改变条件表达式值的语句，否则将成为死循环。

（5）while循环可以用流程图来演示执行过程，如图4-15所示。



图4-15 while循环执行过程流程图

问题

项目小组为了筹集购买奖品和纪念品的费用，准备开展一次募捐活动，计划向第一个赞助人募集20元，向第二个赞助人募集25元……后一个比前一个多募捐5元，依次下去，第几个人募捐后募集到的总费用累计超过500元？

思考

如何用while循环语句编程解决上述的募集费用计算问题？

观察

用while循环语句编程解决上述问题的方法有多种，编程时要注意条件表达式的构成，现给出其中一种方法的步骤如下：

（1）分析问题。

设所拉赞助人的数量为*num*（*num*=1，2，3，…），每次募集的费用为*money*，每次募集之后总费用为*sum*，则*sum*可以用如下的式子表示：

sum=20+25+30+…

于是这个问题就可以转化为：当*sum*≥500时，求*num*的值。

（2）设计算法。

根据上述的分析，设计算法如下：

①令*num*=1；

②令*money*=20；

③令*sum*=20；

微信公众号：电子课本大全

④*num*=*num*+1；

⑤*money*=*money*+5；

⑥*sum*=*sum*+*money*；

⑦如果*sum*<500，则返回步骤④；

⑧输出*num*的值；

⑨结束。

（3）编写程序。

根据上面设计的算法，编写出解决该问题的程序（见配套学习资源包“第四章\课本素材\程序4-6”）如下：

程序4-6

```
num=1
money=20
sum=20
while sum<500:
    num=num+1
    money=money+5
    sum=sum+money
print(num)
```

（4）调试程序。

将程序4-6输入计算机进行调试运行，得出结果为12。

交流

各小组讨论、交流如下问题：

（1）如何根据上述解法的自然语言描述画出其算法流程图？

（2）对于上述问题的解法，如果改变几个变量的初始值，假设令“*num*=0，*money*=20，*sum*=0”，那么，while循环体中的程序段应该如何编写？它们之间的语序能改变吗？为什么？

4.4.3　循环嵌套的应用

单独使用for循环或while循环单循环语句，往往无法解决复杂的问题。如需要解决类似“工作要做很多遍，而每一遍工作都是需要重复做一些事情”的问题，就要用到多重循环，即“循环嵌套”。如果某一种循环语句中包含着自身或其他循环语句，就称为循环的“嵌套”。for循环或while循环之间可以相互嵌套。

微信公众号：电子课本大全



项目小组经过讨论，最后选择购买单价分别为1.8元的笔记本、1.9元的笔、2.1元的小饰品作为活动的奖品和纪念品，每一种物品至少买100件，并且尽可能地用完1000元经费。如何设计方案，才能实现购买物品数量最多呢？如数量相同的情况下余额最小的方案为最佳，则最佳方案中每一种物品的数量是多少？余额是多少？



如何运用循环嵌套的方法编程解决上述的最优采购方案设计问题？



要解决这个问题，需要用到循环嵌套的知识，下面提供了其中一种解法：

（1）分析问题。

设购买单价为1.8元、1.9元、2.1元的物品数量分别为x，y，z，当前余额为r元，所买物品总数量为s，根据题意有下面关系式：$1.8x+1.9y+2.1z\leqslant 1000$（$x$，$y$，$z\in \mathbf{N}^*$，$\mathbf{N}^*$为自然数），我们分别将$x$，$y$，$z$从100至555，526，476逐个穷举，如果$x$，$y$，$z$的值满足式子$1.8x+1.9y+2.1z\leqslant 1000$（$x$，$y$，$z\in \mathbf{N}^*$，$\mathbf{N}^*$表示自然数），如$x+y+z>s$，则$s=x+y+z$，$r=1000-(1.8x+1.9y+2.1z)$，并记录此时$x$，$y$，$z$的值；如$x+y+z=s$，且$1000-(1.8x+1.9y+2.1z)<r$，则$r=1000-(1.8x+1.9y+2.1z)$，并记录此时$x$，$y$，$z$的值。

（2）设计算法。

根据上述的分析，可设计出如下的算法：

①令$x1=100$，令$y1=100$，令$z1=100$，令$s=300$，令$r=1000-(100\times 1.8+100\times 1.9+100\times 2.1)$；

②x从100~555穷举；

③y从100~526穷举；

④z从100~476穷举；

⑤如果$1.8x+1.9y+2.1z\leqslant 1000$，若$x+y+z>s$，则$s=x+y+z$，$r=1000-(1.8x+1.9y+2.1z)$，$x1=x$，$y1=y$，$z1=z$；若$x+y+z=s$，且$1000-(1.8x+1.9y+2.1z)<r$，则$r=1000-(1.8x+1.9y+2.1z)$，$x1=x$，$y1=y$，$z1=z$；

⑥转步骤④；

⑦转步骤③；

⑧转步骤②；

⑨输出$x1$，$y1$，$z1$，s和r的值。

（3）编写程序。

根据上面设计的算法，编写出解决该问题的程序（见配套学习资源包“第四章\课本素材\程序4-7”）如下：

微信公众号：电子课本大全

程序4-7

```
x1=100
y1=100
z1=100
s=300
r=1000-(100*1.8+100*1.9+100*2.1)
for x in range(100,556):
    for y in range(100,527):
        for z in range(100,477):
            if 1.8*x+1.9*y+2.1*z<=1000:
                if x+y+z>s:
                    s=x+y+z
                    r=1000-(1.8*x+1.9*y+2.1*z)
                    x1=x
                    y1=y
                    z1=z
                if x+y+z==s and r>=1000-(1.8*x+1.9*y+2.1*z):
                    s=x+y+z
                    r=1000-(1.8*x+1.9*y+2.1*z)
                    x1=x
                    y1=y
                    z1=z
            else:
                break #当1.8*x+1.9*y+2.1*z>1000,则可以退出循环,
                      避免循环空转而浪费时间
print("符合条件的最优方案是：")
print("单价1.8元的物品的数量是：",x1)
print("单价1.9元的物品的数量是：",y1)
print("单价2.1元的物品的数量是：",z1)
print("共买到的礼物总数量为：",s)
print("余款还剩：",r,"元")
```

（4）调试运行。

将程序4-7输入计算机进行调试运行，结果如下：

```
符合条件的最优方案是：
单价1.8元的物品的数量是：327
单价1.9元的物品的数量是：106
单价2.1元的物品的数量是：100
共买到的礼物总数量为：533
余款还剩：0.0元
```


微信公众号：电子课本大全

交流

各小组讨论、交流for循环、while循环语句在格式、功能、应用等方面的异同及其各自的特点。

实践

1. 我们用for循环计算了从1到10的累加值，并且使用的循环控制变量为递增方式。借助此问题的学习，尝试设计用递减方式的循环控制编写代码。

2. 用while循环编程实现计算从1到10的累加值，并对for循环、while循环的使用进行对比。

4.4.4 循环控制的应用

循环控制包含控制循环变量的变化方式和控制循环的跳转。无论是for循环还是while循环，都需要一个控制循环的变量来控制循环的执行。但是，有些循环仅靠循环变量自身的变化难以达到想要的控制效果，因此需要用到控制循环的跳转语句。实现控制循环的跳转需要用到break和continue两个关键字。其中，break是中断循环，continue是跳出本次循环体的执行。

1. break语句

在循环结构中，可以用break语句跳出当前循环体，从而中断当前循环。在实际运用中，往往是在循环体中使用某一个条件来选择是否中断循环，即跳出当前循环，继而执行当前循环外的下一条语句。也就是说，当某个条件成立的时候，需要立即退出循环，不再运行循环中余下的代码，也不管循环控制变量的条件测试的结果如何。使用break语句，即可让程序能按照要求执行要执行的代码。

2. continue语句

continue语句是对break语句的补充。continue不是立即跳出循环体，而是当条件符合需要跳转的时候，跳过该次循环结束前的语句，回到循环开头的条件测试部分，重新开始执行循环。同理，在实际运用中，往往是在循环体中使用某一个条件来选择是否跳转循环。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，比较程序设计几种基本结构的特点，利用几种程序结构完成相应的程序编写，进一步完善该项目方案中的各项学习活动，并参照项目范例的样式，撰写相应的项目成果报告。

微信公众号：电子课本大全

成果交流

各小组运用数字化学习工具，将所完成的项目成果，在小组或班级上进行展示与交流，共享创造、分享快乐。

活动评价

各小组根据项目选题、拟订的项目方案、实施情况以及所形成的项目成果，利用教科书附录2的“项目活动评价表”，开展项目学习活动评价。

拓展

函数的定义与调用

在面对较大项目的时候，随着堆砌的代码越来越多，编程人员发现有很多功能重复的模块被反复地键入和执行。因此需要考虑是否可以将这些功能具体且经常使用的代码段，从程序中抽离出来单独封装。于是，函数的概念出现在编程语言里。在对函数的设计方面，我们可以向函数提供必要的参数输入，同时可以调用函数获取所需的返回值。Python采用def作为关键词来定义一个函数，具体例子如下：

定义和调用函数getValue(*b*,*r*,*n*)，根据本金*b*、年利率*r*和年数*n*，计算最终收益*v*。提示：$v=b(1+r)^n$。

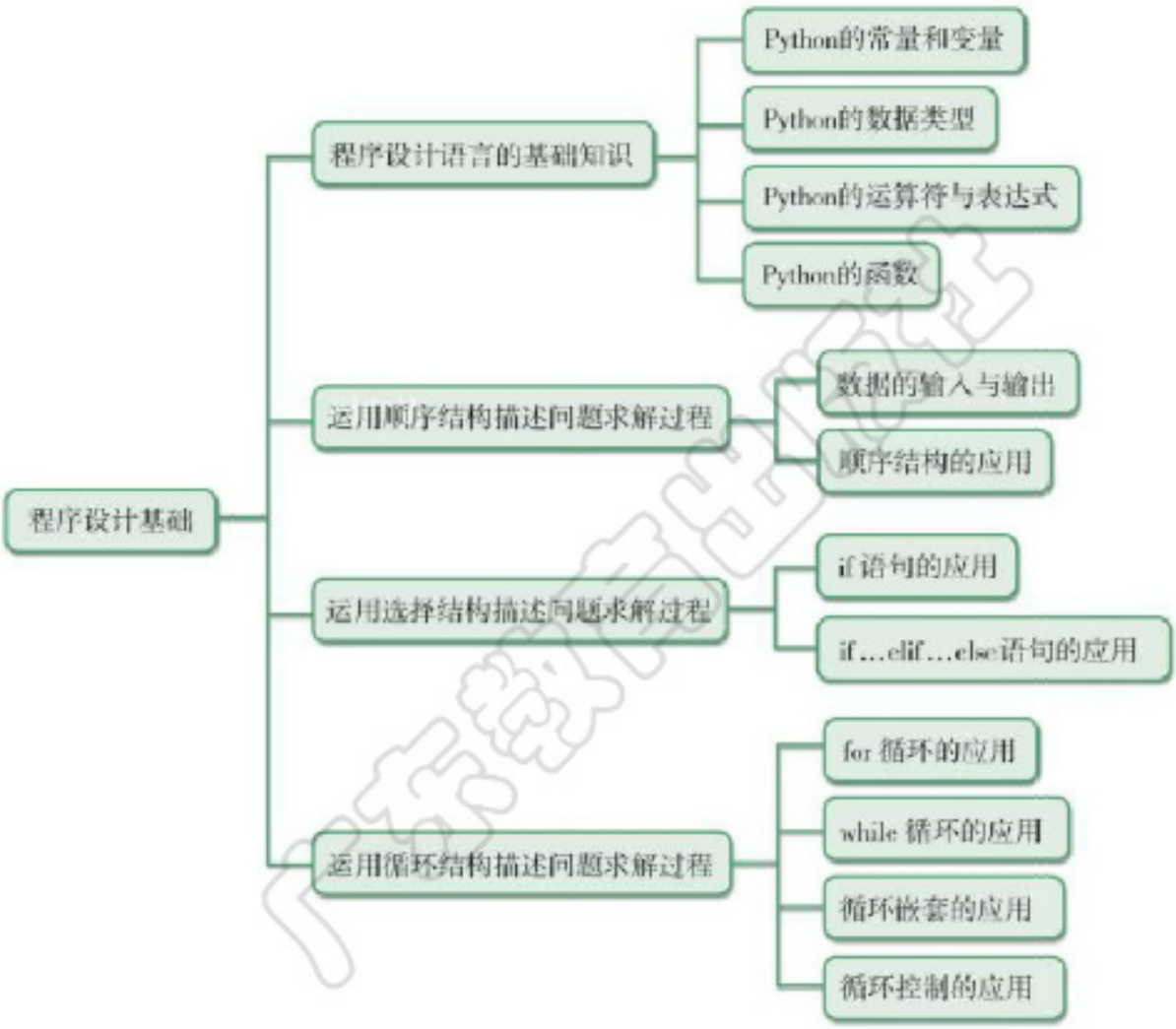
```
def getValue(b,r,n):          #创建函数对象getValue
    v=b*((1+r)**n)            #计算最终收益v
    return v                  #使用return返回值
total = getValue(1000,0.05,5)    #调用函数getValue
print(total)                  #打印结果
```

程序运行结果为：1276.2815625000003。

微信公众号：电子课本大全

本章扼要回顾

同学们通过本章学习，根据“程序设计基础”知识结构图，扼要回顾、总结、归纳学过的内容，建立自己的知识结构体系。



回顾与总结

微信公众号：电子课本大全

本章学业评价

同学们完成下列测试题（更多的测试题可以在教科书的配套学习资源包中查看），并通过“本章扼要回顾”以及本章的项目活动评价，综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。

1. 单选题

（1）在Python中，元素是无序不重复的是（　　）。

A. 列表　　B. 元组　　C. 集合　　D. 字符串

（2）在以下运算符中，优先级最低的运算符是（　　）。

A. *　　B. !=　　C. +　　D. =

（3）有以下程序段：

```
sum=1
for i in range(1,5):
  sum=sum*i
print(sum,i)
```

程序的运行结果为（　　）。

A. 24　5　　B. 120　5　　C. 24　4　　D. 120　4

2. 思考题

编写函数，接收一个正偶数为参数，输出两个素数，并且这两个素数之和等于原来的正偶数。如果存在多组符合条件的素数，则全部输出。

3. 情境题

依法纳税是每一个公民应尽的义务。同学们上网学习2018年8月31日第七次修正的《中华人民共和国个人所得税法》，了解居民个人应纳税所得额的含义和个人所得税税率表（综合所得适用），试设计一个程序计算以下人员应缴个人所得税额。

人员	A	B	C	D	E
全年应纳税所得额/元	763 000	550 000	360 000	125 000	35 000
应缴个人所得税额/元					

微信公众号：电子课本大全

第五章
数据处理和可视化表达

随着移动互联网和物联网的飞速发展，人类社会产生的数据以惊人的速度增长。海量的数据几乎包含一切形式的结构化、半结构化以及非结构化的数据，如网络日志、音频、视频、图片、地理位置信息等。如何高效地对这些数据进行采集、存储、处理，并从中发掘到有价值的信息，是数据分析处理需要解决的问题。

本章将通过“网络平台的客户行为数据分析和可视化表达”项目，进行自主、协作、探究学习，让同学们认识大数据及其特征，认识大数据对人们日常生活的影响，了解数据采集、分析和可视化表达的基本方法；学会选用恰当的软件工具或平台处理数据，完成分析报告；理解对数据进行保护的意义，从而将知识建构、技能培养与思维发展融入运用数字化工具解决问题和完成任务的过程中，促进信息技术学科核心素养达成，完成项目学习目标。

- 认识大数据
- 数据的采集
- 数据的分析
- 数据的可视化表达

微信公众号：电子课本大全

项目范例　网络购物平台客户行为数据分析和可视化表达

情境

网络购物中，消费者购物的任何行为细节都会被服务器所记录。通过大数据分析，商家可以了解消费者的偏好甚至预测其购买行为，如图5-1所示。大数据分析使得商家可以追踪客户的行为并确定最有效的方式以提升客户对购物平台的忠诚度，根据客户的个性化需求提供相应的产品或服务以获得更大的市场占有率。

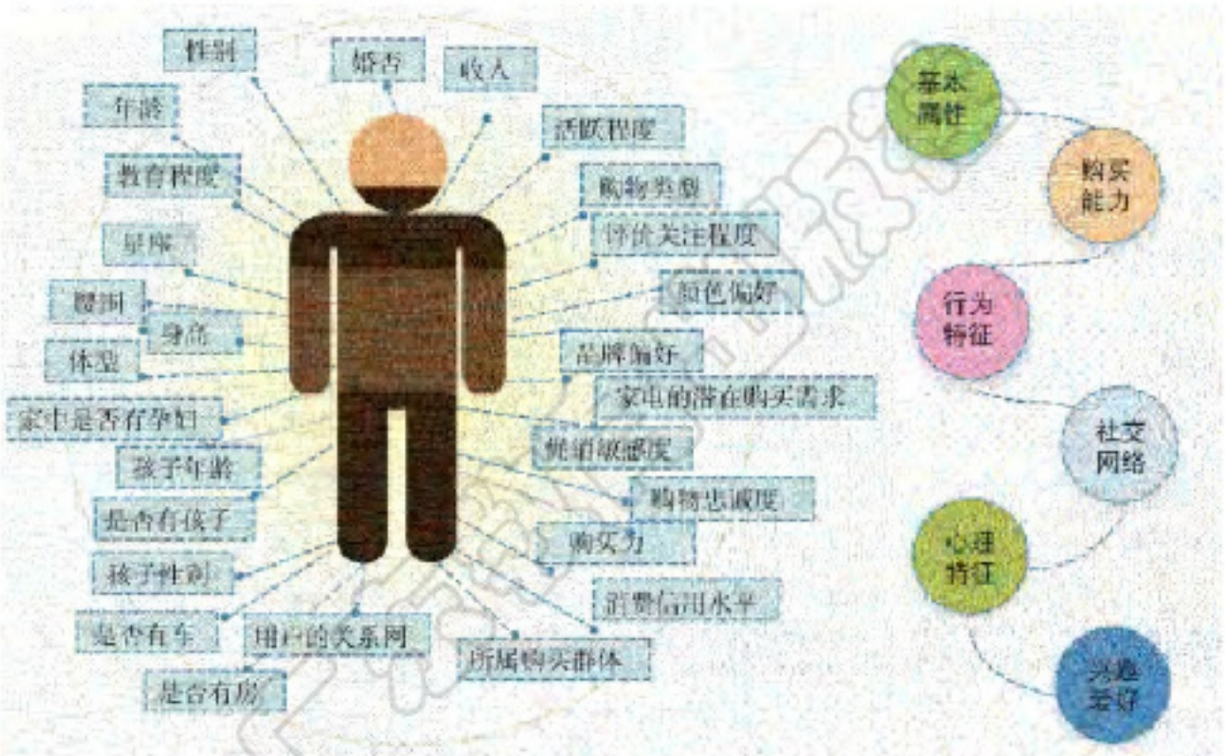


图5-1　客户网购行为分析

主题

网络购物平台客户行为数据分析和可视化表达

规划

根据项目范例的主题，在小组中组织讨论，利用思维导图工具，制订项目范例的学习规划，如图5-2所示。

微信公众号：电子课本大全

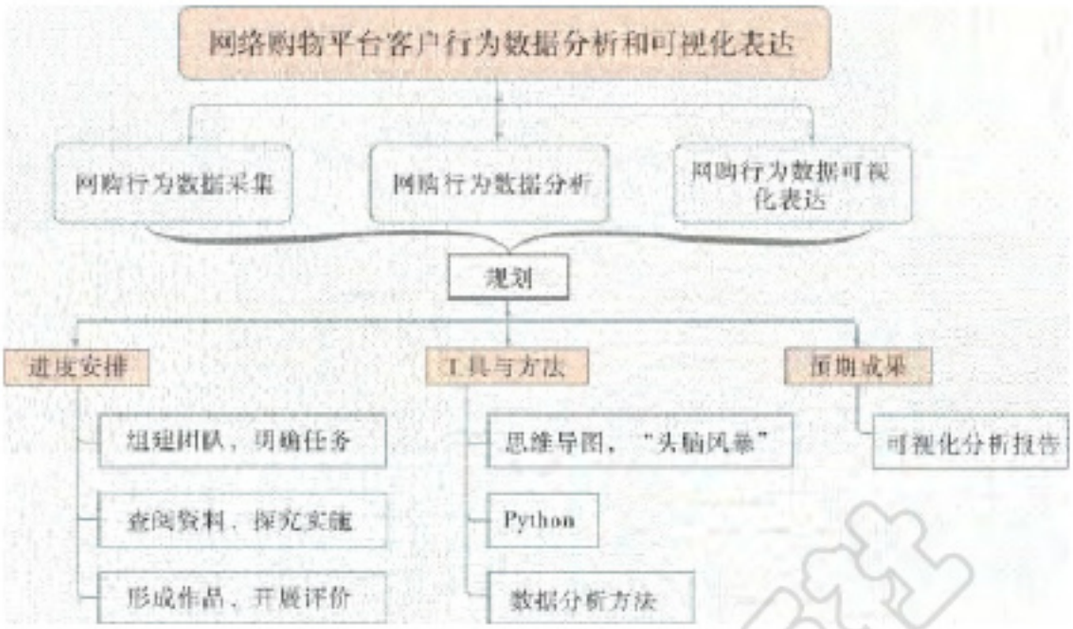


图5-2 “网络购物平台客户行为数据分析和可视化表达”项目学习规划

探究

根据项目学习规划的安排，通过调查、案例分析、文献阅读和网上资料搜索，开展“网络购物平台客户行为数据分析和可视化表达”项目学习探究活动，如表5-1所示。

表5-1 “网络购物平台客户行为数据分析和可视化表达”项目学习探究活动

探究活动	学习内容	知识技能
网购行为数据采集	大数据及其特征。 数据的来源。 数据对日常生活的影响。 数据保护的意义。	认识大数据及其特征。 了解数据采集的基本方法。 理解数据对日常生活的影响。 理解对数据进行保护的意义。
网购行为数据分析	数据分析方法。 选择数据分析处理工具。	了解数据分析的基本方法。 学会选用恰当工具处理数据。
网购行为数据可视化表达	可视化的方法和方式。 选择数据可视化表达处理工具。	了解数据可视化表达的基本方法。 学会选用恰当工具可视化表达数据。

实施

实施项目学习各项探究活动，进一步认识消费者网络购物行为特点。

微信公众号：电子课本大全

成果

在小组开展项目范例学习过程中，利用思维导图工具梳理小组成员在“头脑风暴”活动中的观点，建立观点结构图，运用多媒体创作工具（如演示文稿、在线编辑工具等），综合加工和表达，形成项目范例可视化学习成果，并通过各种分享平台发布，共享创造、分享快乐。例如，运用在线编辑工具制作的“网络购物平台客户行为数据分析和可视化表达”可视化报告，可以在教科书配套学习资源包中查看，其目录截图如图5-3所示。



图5-3　“网络购物平台客户行为数据分析和可视化表达”可视化报告目录截图

评价

根据教科书附录2的“项目活动评价表”，对项目范例的学习过程和学习成果在小组或班级上进行交流，开展项目学习活动评价。

项目选题

同学们以3~6人组成一个小组，选择下面一个参考主题，或者自拟一个感兴趣的主题，开展项目学习。

1. 网约车平台客户行为的数据分析和可视化表达
2. 社交平台用户行为的数据分析和可视化表达
3. 网络学习平台学生行为的数据分析和可视化表达

项目规划

各小组根据项目选题，参照项目范例的样式，利用思维导图工具，制订相应的项目方案。

微信公众号：电子课本大全

方案交流

各小组将完成的方案在全班进行展示交流，师生共同探讨、完善相应的项目方案。

5.1 认识大数据

在日常工作、生活和学习等活动中，人们的一举一动基本上都可以数字化。例如，从家中出门到达工作地点选择的交通工具、路线和所用时间；每刷一次微博、拨打一次电话、发送一条短信、网上银行转账或者浏览相关网站，甚至所在的地理位置信息等都产生了大量数据。

在第一章，我们已经介绍了数据的基本概念与特征，下面我们将学习大数据的基本概念与特征。

5.1.1 大数据

大数据是指无法在可承受的时间范围内用常规软件工具进行高效捕捉、管理和处理的数据集合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

大数据的产生是与人类日益普及的网络行为所伴生的：物联网、云计算、移动互联网、车联网、手机、电脑以及遍布地球各个角落的各种各样的传感器，无一不是数据的来源或是承载的方式。互联网生成的数据量，不仅远超此前一切人类所生成的数据量的总和，而且在以大爆发性的速度不断增长。

在天气预报、大气监测、地球物理探矿和天体运动观测等科学实验和科学观察等活动中，各种各样的传感器每时每刻都在产生大量的数据。

大数据已成为人们提取信息、做出决策的重要依据，是推动信息社会发展的重要资源。

5.1.2 大数据的特征

我们正处于一个大数据时代，大数据使人们的生活、工作与思维方式等都产生了巨大的变革。那么，大数据究竟具有哪些特征？

微信公众号：电子课本大全

探究活动

分析

仔细阅读以下案例，分析大数据具有哪些特征。

案例1：某市交通智能化分析平台的数据来源于道路交通、电信、地理信息系统等各行各业。例如，交通卡刷卡每天产生1900万条记录，手机定位数据每天产生1800万条，出租车运营数据每天产生100万条，电子停车收费系统数据每天产生50万条，等等。这些数据在体量和速度上都达到了大数据的规模。

案例2：搜索引擎公司通过跟踪网民对“感冒症状”以及“治疗”等关键词的搜索，发现某个时段在某个区域内搜索数量急剧增长，从而成功预测了甲型H1N1流感的暴发时间、地域。

案例3：目前，某基于大数据的网约车平台已覆盖全国400多个城市，涵盖出租车、快车、顺风车、代驾、专车、试驾以及租车等多项业务，为人们的出行带来极大的便利。我们只需在网约车APP上输入或者说出目的地，强大的智能系统就立刻分配订单，即时通知附近司机；借助定时定位系统，我们可以看到司机的大致位置以及预计到达时间。

（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（Volume）、多样（Variety）、低价值密度（Value）、高速（Velocity）。

第一，数据体量巨大。从TB级别跃升到PB级别。

第二，数据类型繁多。如网络日志、视频、图片、地理位置信息等。

第三，价值密度低。以视频为例，在连续不间断的监控过程中，有用的数据可能仅仅一两秒。

第四，变化速度快。数据来自世界各地的网络终端，且以秒为单位快速变化。

（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。

第一，在大数据时代强调数据要全量而不是抽样，即强调数据规模全量，而不是强调数量巨大。大数据时代有了更好的数据采集手段，让获取全量数据成为可能。

第二，大数据时代研究的数据如此之多，追求的不是精确性，而是模糊性。在大数据时代，只要掌握了大体的发展方向即可，适当忽略微观层面上的精确度，会让我们在宏观层面拥有更好的洞察力。

第三，大数据时代不是因果关系，而是相关关系。在大数据时代，无须再紧盯事物之间的因果关系，而应该寻找事物之间的相关关系；相关关系也许不能准确地告诉我们某件事情为何会发生，但是会提醒我们这件事情正在发生。

（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。

第一，大数据存储在互联网不同的服务器与各客户终端。

第二，应用分布式并行计算处理互联网的大数据。

微信公众号：电子课本大全

讨论

以小组为单位上网搜索资料，讨论传统数据与大数据的区别，并填入表5-2中。

表5-2 传统数据与大数据的区别

传统数据	大数据

5.1.3 大数据对日常生活的影响

1. 大数据使人们日常生活更为便捷

（1）方便支付。中国的移动支付发展得特别快。在中国，每三个手机用户，就有两个在使用移动支付。中国是全球最大的移动支付市场。医院、餐厅、菜市场、加油站，甚至路边摊，都在使用移动支付。中国人今天的生活，已经越来越有科技含量。

（2）方便出行。应用交通系统的大数据，网络约车出行，智能导航行车避免堵车，无人驾驶，智能地图方便寻路，等等。

（3）方便购物与产品推介。网络购物不但节省人们出行购物的时间，而且帮助企业有效判断用户的信息需求和消费需求，对客户进行产品推介，方便人们选购产品。

（4）方便看病与诊病。应用网络预约挂号，减轻与节省患者排队挂号看病的辛劳与时间；同时，又方便医生提前分析患者的病史数据，以便更科学诊病。

2. 大数据对人们日常生活产生的负面影响

（1）个人信息泄露。在大数据时代，我们使用的手机、计算机、网络、信用卡等信息科技，都会产生数据。这些数据时刻存在泄露的风险。

（2）信息伤害与诈骗。在大数据时代，我们的网络信息随时都可能被不法分子窃取，并对我们及身边的亲人造成伤害。

微信公众号：电子课本大全

5.2 数据的采集

数据的来源广泛（主要是互联网和物联网）、类型丰富、规模巨大。采集数据首先要明确数据应用项目的需求，围绕选定的项目主题，制订数据采集的需求清单和内容大纲，再采用适当的方法和工具进行采集。

5.2.1 数据采集的方法和工具

数据采集的基本方法包括系统日志采集法、网络数据采集法和其他数据采集法。

1. 系统日志采集法

在信息系统中，系统日志是记录系统中硬件、软件和系统问题的信息文件。系统日志包括操作系统日志、应用程序日志和安全日志。系统日志采集数据的方法通常是在目标主机上安装一个小程序，将目标主机的文本、应用程序、数据库等日志信息有选择地定向推送到日志服务器进行存储、监控和管理。

通过日志服务器可以监视系统中发生的事件，可以检查错误发生的原因，或者寻找受到攻击时攻击者留下的痕迹。例如，安全管理信息系统就是以系统日志服务器采集原始日志数据，以日志记录文本文件实现日志数据的监控和保存，以数据库操作进行日志有效信息的管理工作。

2. 网络数据采集法

网络数据采集是指通过网络爬虫或网站公开API（Application Programming Interface，应用程序接口）等方式从网站上获取数据信息。网络爬虫从一个或若干初始网页的URL（Uniform Resource Locator，统一资源定位符）开始，获得初始网页上的URL，在抓取网页的过程中，不断从当前页面上抽取新的URL放入队列，直到满足系统的一定停止条件（图5-4）。该方法可以将非结构化数据从网页中抽取出来，将其存储为统一的本地数据文件，并以结构化的方式存储。它支持图片、音频、视频等文件或附件的采集，附件与正文可以自动关联。

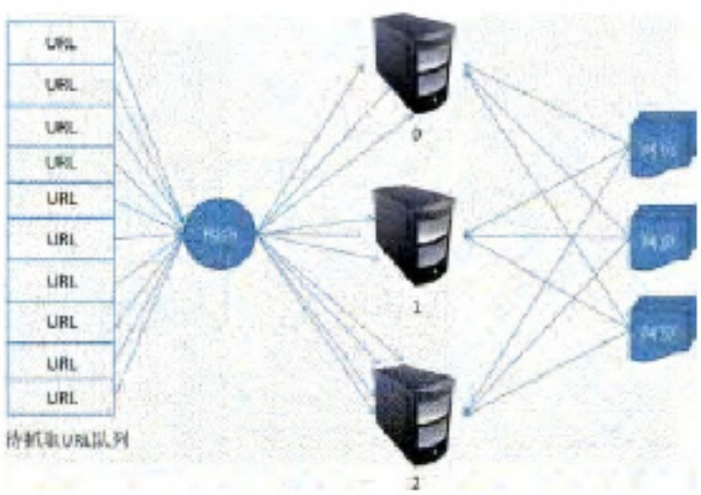


图5-4　网络爬虫

微信公众号：电子课本大全

拓展

Python网络数据采集程序使用的扩展库

在众多的数据采集工具中，Python以其简洁、开源和包容的特性在数据采集和分析领域独树一帜。由于Python可以安装第三方扩展库模块来扩展功能，因此使用Python进行网络数据采集和分析显得简单易用。以下是使用Python进行网络数据采集和分析所需要的一些第三方扩展库。

（1）NumPy。

NumPy（Numerical Python）是构建科学计算最基础的软件库，为Python中的n维数组和矩阵的操作提供了大量有用的功能。该库还提供了NumPy数组类型的数学运算向量化，可以提升性能，加快执行速度。

（2）SciPy。

SciPy是一个工程和科学软件库，包含线性代数、优化、集成和统计的模块。SciPy库的主要功能建立在 NumPy 的基础之上，因此它的数组大量使用了NumPy。它通过其特定的子模块提供高效的数值例程操作，如数值积分等。SciPy的所有子模块中的函数都有详细的介绍文档。

（3）Pandas。

Pandas是一个Python包，旨在通过标记（labeled）和关系（relational）数据进行工作，简单直观。Pandas主要用于快速简单的数据操作、聚合和可视化呈现。库中有两个主要的数据结构，一维数组（Series）和二维数组（DataFrame）结构。

（4）Matplotlib。

Matplotlib是Python的一个2D绘图库，以各种硬拷贝格式和跨平台的交互式环境生成出版质量级别的图形。在NumPy、SciPy和Pandas的帮助下，通过 Matplotlib，开发者仅需输入几行代码，便可以生成绘图、直方图、功率谱、条形图、散点图等。

在Python模块库中有大量模块可供使用，要想使用这些文件，就需要用import语句把指定模块导入当前程序中。使用import语句导入模块的语法如下：

```
import module
关键字 模块名
```

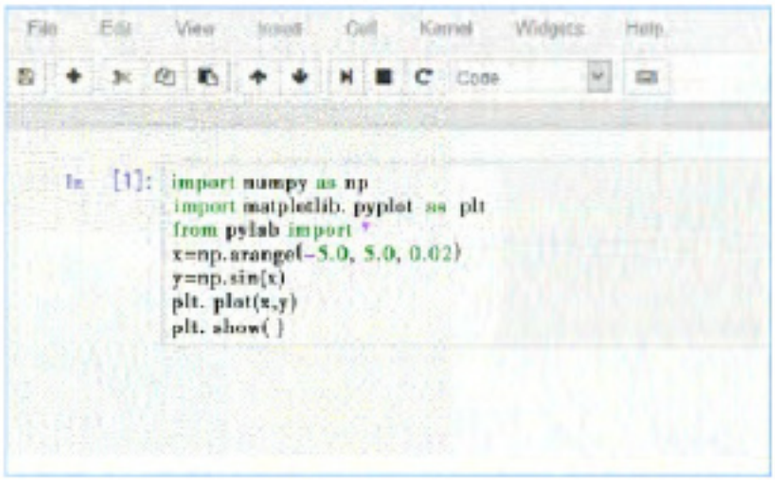
from import语句也是导入模块的一种方法，是导入指定模块内的指定函数方法。使用from import语句导入模块内指定方法的语法如下：

```
from module import name
关键字 模块名 关键字 方法名
```

体验

打开并运行配套学习资源包“第五章\课本素材\程序5-1”文件，其截图如图5-5所示，指出程序中import语句的作用。

微信公众号：电子课本大全



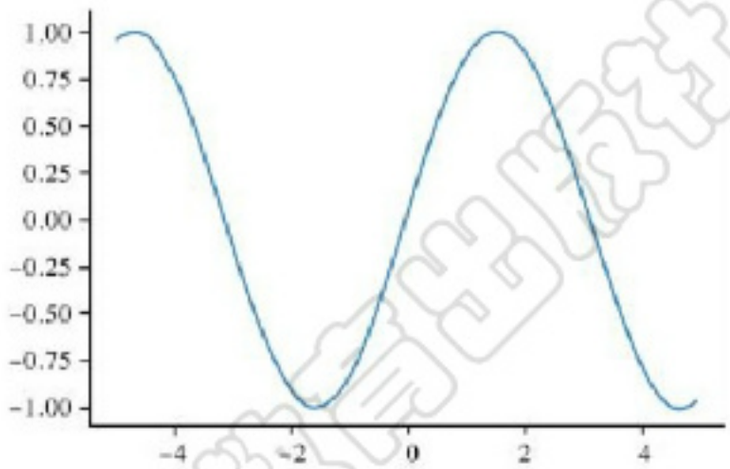


图5-5　程序5-1和绘图结果



利用配套学习资源包提供的装虚拟机文件，搭建“网络商城”网站，具体操作详见配套学习资源包的说明文件。

打开并运行配套学习资源包“第五章\课本素材\程序5-2”。其中，语句import bs4和import requests用于调用爬虫模块，import re用于调用正则表达式模块，import xlwt用于调用Excel模块，import datetime用于调用日期和时间模块。程序5-2的关键过程及语句如下：

（1）爬取页面设置。

```
for i in range(0, 5):       # 循环5次，就是5个页面的商品数据
    payload['page'] = i+ 1   # 此处为页码，根据网页参数具体设置
    resp = requests.get(url, params=payload)
    soup = bs4.BeautifulSoup(resp.text, "html.parser")
    print(resp.url)          # 打印访问的网址
    resp.encoding = 'utf-8'  # 设置编码
```


微信公众号：电子课本大全

（2）爬取商品标题。

```
all_title = soup.find_all('a', class_=re.compile("history-item product"))
    for j in all_title:
        soup_title = bs4.BeautifulSoup(str(j), "html.parser")
        title.append(soup_title.a['title'])
```

（3）爬取商品价格。

```
all_price = soup.find_all('span', itemprop="price")
    for k in all_price:
        soup_price = bs4.BeautifulSoup(str(k), "html.parser")
        price.append(soup_price.span.string)
```

（4）爬取商品订单量。

```
all_order = soup.find_all('a', class_=re.compile("order-num-a"))
    for l in all_order:
        soup_order = bs4.BeautifulSoup(str(l), "html.parser")
        order.append(soup_order.em.string)
```

（5）爬取出售商品的店铺名称。

```
all_store = soup.find_all('div', class_="store-name util-clearfix")
    for m in all_store:
        soup_store = bs4.BeautifulSoup(str(m), "html.parser")
        store.append(soup_store.a.string)
```

3. 其他数据采集法

对于企业生产经营或科学研究等保密性要求较高的数据，可通过与企业或研究机构合作，使用特定系统接口等相关方式收集数据。例如，科学研究的数据是通过科学实验的各种传感器采集，并传输到数据库管理系统中的。

5.2.2 数据的存储和保护

1. 数据的存储

存储数据主要有两种方式，一种是把数据存在本地内部，另一种是把数据放在第三方

微信公众号：电子课本大全

公共或私有的“云端”存储，如图5-6所示。

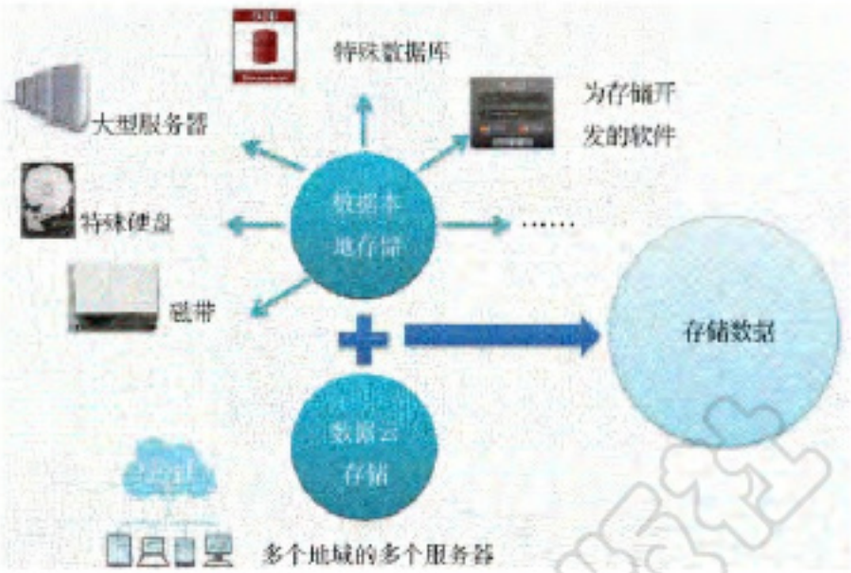


图5-6　数据的存储

云存储已经成为存储发展的一种趋势，其技术也日益成熟。云存储是把各类数据存储在虚拟的逻辑模型里，其物理空间存储在跨越多个地域放置的众多服务器中，为用户提供统一、灵活、安全的“云存储服务”。云存储供应商拥有并管理这些服务器，负责管理数据的使用和访问权限，以及云存储环境的日常运营和维护。对于用户而言，无须关注云存储系统的具体运行，仅需获取存储空间，把自己的数据存储进去。

数据的存储采用分布式文件存储或NoSQL数据库存储。分布式文件存储的特点之一是为了解决复杂问题而将大任务分解为多项小任务，通过让多个处理器或多个计算机节点并行计算来提高解决问题的效率。分布式文件存储系统能够支持多台主机通过网络同时访问共享文件和存储目录，大部分采用了关系数据模型并且支持SQL语句查询。

2. 数据的保护

如今，无论是政府部门、企业还是个人，对数据的依赖性已越来越强。然而，数据安全的隐患无处不在，一旦数据泄密或丢失，造成的损失和影响将是巨大的。因此，对数据安全的保护非常重要。研究表明，如果在发生数据灾难后的两个星期内无法恢复公司的业务系统，75%的公司业务将会完全停顿，43%的公司将再也无法开业。在信息化社会，对数据的保护刻不容缓。

（1）数据安全保护技术。数据安全保护指数据不被破坏、更改、泄露或丢失。安装杀毒软件和防火墙只能防备数据安全隐患，而采用拷贝、备份、复制、镜像、持续备份等技术进行数据保护才是更为彻底、有效的方法。一般的数据安全保护技术的使用特点如表5-3所示。

微信公众号：电子课本大全

表5-3 数据安全保护技术的使用特点一览表

数据安全保护技术	适用场合	备份介质	备份距离	管理
拷贝/FTP	简单小数据量备份，个人不定期的文件保护等。	磁盘。	近。	手动执行，占用人力资源。
备份	有归档需求的用户等。	磁带机、磁带库、磁盘。	近，以本地备份为主。	备份软件对使用者要求较高，需要掌握数据库、文件系统等综合知识。
复制技术	企业等。	磁盘。	远近皆可。	设定策略后无须人工干预，复制与恢复的过程都很简单。
镜像技术	企业等。	磁盘。	近，带宽和距离影响延迟时间和性能，因此多以本地为主。	简单。
持续备份	企业等。	磁盘。	远近皆可。	连续备份，可以实现过去任意一个时间点的数据恢复。

为了防止他人对机密的数据、数据库进行非法访问、删除、修改、拷贝等操作，可以采用对数据进行加密等方法，保护数据在存储和传递过程中不被修改或泄露。选择何种加密算法、需要多高的安全级别、各算法之间如何协作等，都是进行数据加密要考虑的因素。加密技术通常分为对称式加密和非对称式加密两大类。对称式加密指加密和解密用的是同一个密钥。非对称式加密指加密和解密用的是两个不同的密钥，必须配对使用，否则不能打开加密数据。

（2）数据的隐私保护。任何事物都有两面性，数据正在变成生活的第三只眼，敏锐地洞察却也正监控着我们的生活。我们上网浏览、出行、购物等数据，统统都被记录了，人人都成了数据的产生者和贡献者。数据带来的整体性变革，也使得数据的隐私保护的形势显得越发严峻。

隐私泄露的问题不是大数据时代特有的，在没有大数据的时候，我们就已经有很多隐私泄露的问题。可是到了大数据时代，数据发布多了，信息范围扩大了，信息传播和共享速度加快了，若不加以控制，其所含的商业信息或私密信息就可能泄露（图5-7）。解决办法有三个：一是技术手段，常用的隐私保护有：①数据收集时进行数据精度处理；②数据共享时进行访问控制；③数据发布时进行人工加扰；④数据分析时进行数据匿名处理等。二是提高自身的保护意识。三是要对数据使用者进行道德和法律上的约束。

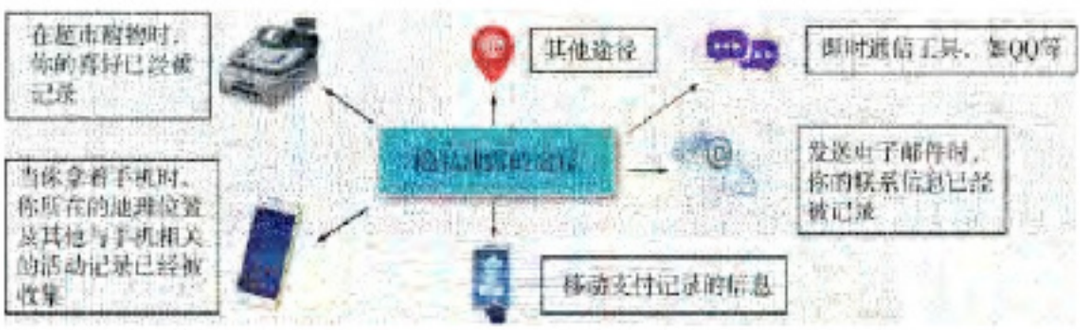


图5-7 隐私泄露的途径

微信公众号：电子课本大全

数据作为一种资产，用在什么地方、掌握在谁手里都是次要的，关键是看怎样利用这个工具。在安全的前提下，实现数据共享，才能真正创造数据价值，发挥数据真正的作用。



某用户在电脑客户端打开了某电商的网站，随意搜索了一下“行李箱”的购物信息，浏览了几个店铺，便关闭了该网站。然而，当他打开其他网站时，却“惊喜”地发现几乎所有的网页上都闪烁着行李箱的图片和价格，即使网页上没有出现广告，右下角也会弹出一个广告框，里面的内容依然是行李箱。

用户主动访问某些网站时，便产生了网络缓存（Cookies）。Cookies是指用户浏览网页时，网络服务器以文本格式存储在用户电脑硬盘上的少量数据。Cookies的主要目的是帮助网站记忆用户之前可能进行的操作。网站通过获取并分析属于用户的Cookies，为用户提供设置或推送服务。这样使得网站更人性化，操作更方便快捷。



1. 选择网页浏览器，右键选择“属性”。
2. 在弹出的“Internet选项”窗口选择“设置”按钮，如图5-8所示。
3. 在弹出的新窗口“网站数据设置”中，当前位置就是Cookies所在的位置，可以选择“查看文件”按钮找到文件所在处，如图5-9所示。

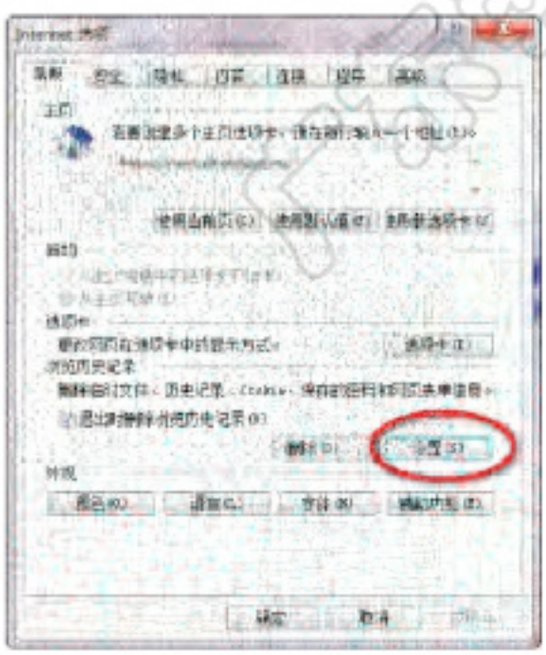


图5-8　“Internet选项”窗口

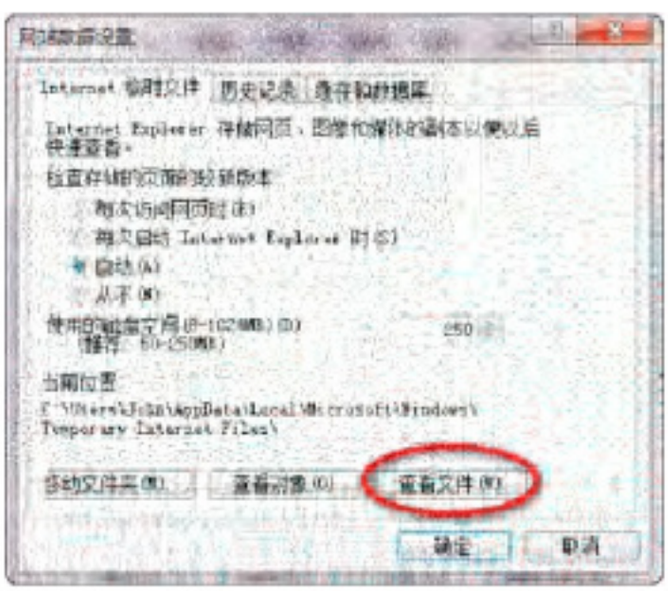


图5-9　“网站数据设置”窗口

微信公众号：电子课本大全

交流

很多网站利用Cookies跟踪用户行为。有人认为网站可能存在侵犯用户隐私的问题，也有人认为网站利用Cookies给用户带来了很多便利，免去了用户重复登录网站的麻烦。在小组中讨论交流，谈谈如何有效阻止Cookies跟踪或者跨站跟踪的行为。

拓展

《中华人民共和国网络安全法》（节录）

（2016年11月7日第十二届全国人民代表大会常务委员会第二十四次会议通过）

第十八条　国家鼓励开发网络数据安全保护和利用技术，促进公共数据资源开放，推动技术创新和经济社会发展。国家支持创新网络安全管理方式，运用网络新技术，提升网络安全保护水平。

第二十七条　任何个人和组织不得从事非法侵入他人网络、干扰他人网络正常功能、窃取网络数据等危害网络安全的活动；不得提供专门用于从事侵入网络、干扰网络正常功能及防护措施、窃取网络数据等危害网络安全活动的程序、工具；明知他人从事危害网络安全的活动的，不得为其提供技术支持、广告推广、支付结算等帮助。

第三十一条　国家对公共通信和信息服务、能源、交通、水利、金融、公共服务、电子政务等重要行业和领域，以及其他一旦遭到破坏、丧失功能或者数据泄露，可能严重危害国家安全、国计民生、公共利益的关键信息基础设施，在网络安全等级保护制度的基础上，实行重点保护。关键信息基础设施的具体范围和安全保护办法由国务院制定。国家鼓励关键信息基础设施以外的网络运营者自愿参与关键信息基础设施保护体系。

项目实施

各小组根据项目选题及拟订的项目方案，并结合本节所学知识，开展以下活动。

1. 完成相应的数据采集。
2. 总结归纳数据采集的方法和步骤。

5.3 数据的分析

数据分析就是在一大批杂乱无章的数据中，运用数字化工具和技术，探索数据内在的结构和规律，构建数学模型，并进行可视化表达，通过验证将模型转化为知识，为诊断过去、预测未来发挥作用。数据分析一般包括特征探索、关联分析、聚类与分类、建立模型和模型评价等。

微信公众号：电子课本大全

5.3.1 特征探索

数据特征探索的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，求最大值、最小值、极差等描述性统计量。

探究活动

实践

打开并运行配套学习资源包“第五章\课本素材\程序5-3数据预处理”，观察数据预处理结果。程序5-3数据预处理的关键过程及部分程序段如下：

（1）数据清洗，发现缺失值。

```
x=0
data["price"][(data["price"]==0)]=None
for i in data.columns:
    for j in range(len(data)):
        if(data[i].isnull())[j]:
            data[i][j]="64"
            x+=1
print(x)
```

（2）异常值处理中，利用画散点图发现异常值部分。

```
data2=data.T
price=data2.values[2]
comt=data2.values[3]
pyl.plot(price,comt,"o")
pyl.show()
```

（3）求最大值和最小值。

```
pricemax=da2[2].max()
pricemin=da2[2].min()
commentmax=da2[3].max()
commentmin=da2[3].min()
```


微信公众号：电子课本大全

（4）求极差。

```
pricerg=pricemax-pricemin
commentrg=commentmax-commentmin
```

（5）求组距。

```
pricedst=pricerg/13
commentdst=commentrg/13
```

（6）绘制价格直方图。

```
pricesty=npy.arange(pricemin,pricemax,pricedst)
pyl.hist(da2[2],pricesty)
pyl.show()
```

（7）绘制评论数直方图。

```
commentsty=npy.arange(commentmin,commentmax,commentdst)
pyl.hist(da2[2],commentsty)
pyl.show()
```

5.3.2 关联分析

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。关联分析的基本算法如下：

（1）扫描历史数据，并对每项数据进行频率次数统计。

（2）构建候选项集$C1$，并计算其支持度，即数据出现频率次数与总数的比。

（3）对候选项集的支持度进行筛选，筛选的数据项支持度应当不小于最小支持度，从而形成频繁项集$L1$。

（4）对频繁项集$L2$进行连接生成候选项集$C2$，重复上述步骤，最终形成频繁K项集或者最大频繁项集。



打开并运行教科书配套学习资源包“第五章\课本素材\程序5-4关联分析”，观察程序的运行结果。程序5-4关联分析的关键过程及程序段如下：

（1）自定义连接函数，用于实现L_{k-1}到C_k的连接。

微信公众号：电子课本大全

```
def connect_string(x,ms):
    x=list(map(lambda i:sorted(i.split(ms)),x))
    l=len(x[0])
    r=[]
    for i in range(len(x)):
        for j in range(i,len(x)):
            if x[i][:l-1]==x[j][:l-1] and x[i][l-1]!=x[j][l-1]:
                r.append(x[i][:l-1]+sorted([x[j][l-1],x[i][l-1]]))
    return r
```

（2）寻找关联规则。

```
result=pd.DataFrame(index=['support','confidence']) #定义输出结果
    support_series=1.0*d.sum()/len(d) # 支持度序列
    column=list(support_series[support_series>support].index) #初步根据支持度筛选
    k=0
    while len(column)>1:
        k=k+1
        print(u'\n正在进行第%s次搜索...'%k)
        column=connect_string(column,ms)
        print(u'数目：%s...'%len(column))
        sf=lambda i:d[i].prod(axis=1,numeric_only=True) #新一批支持度的计算函数
```

（3）创建连接数据。

```
d_2=pd.DataFrame(list(map(sf,column)),index=[ms.join(i)for i in column]).T
support_series_2=1.0*d_2[[ms.join(i) for i in column]].sum()/len(d)
column=list(support_series_2[support_series_2>support].index)
support_series=support_series.append(support_series_2)
 column2=[]
```

（4）遍历可能的推理。

```
for i in column:
    i=i.split(ms)
    for j in range(len(i)):
    column2.append(i[:j]+i[j+1:]+i[j:j+1])
```


微信公众号：电子课本大全

（5）计算置信度序列。

```
for i in column2:
    cofidence_series[ms.join(i)]=support_series[ms.join(sorted(i))]/support_series[
ms.join(i[:len(i) - 1])]
```

（6）置信度筛选。

```
for i in cofidence_series[cofidence_series>confidence].index:
        result[i]=0.0
        result[i]['confidence']=cofidence_series[i]
        result[i]['support']=support_series[ms.join(sorted(i.split(ms)))]
```

5.3.3 聚类分析

聚类分析是一种探索性的分析，在分类的过程中，人们不必事先给出一个分类的标准，聚类分析能够从样本数据出发，自动进行分类。聚类分析的算法有很多，其中K-平均（K-Means）算法是一种经典的自下而上的聚类分析方法。K-平均算法的基本思想就是在空间N个点中，初始选择K个点作为中心聚类点，然后将N个点分别与K个点计算距离，选择自己最近的点作为自己的中心点，再不断更新中心聚集点，以达到“物以类聚，人以群分”的效果，如图5-10所示。

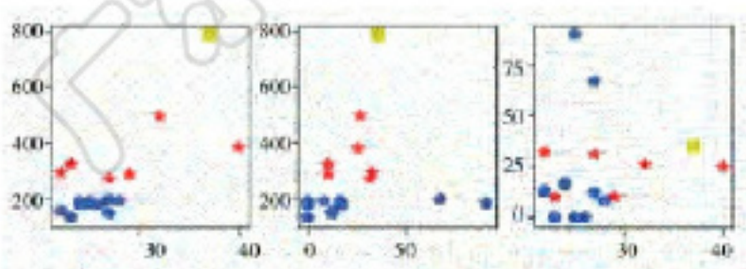


图5-10 聚类分析K-平均算法

聚类分析的基本算法如下：

（1）从数据点集合中随机选择K个点作为初始的聚集中心，每个中心点代表着每个聚集中心的平均值。

（2）对其余的每个数据点，依次判断其与K个中心点的距离，距离最近的表明它属于这项聚类。

（3）重新计算新的聚簇集合的平均值即中心点。整个过程不断迭代计算，直到达到预先设定的迭代次数或中心点不再频繁波动。

微信公众号：电子课本大全



打开并运行教科书配套学习资源包“第五章\课本素材\程序5-5聚类分析”文件，观察程序的运行结果。程序5-5直接调用了Python语言的sklearn机器学习模块，其对数据进行聚类分析的关键程序段如下：

```
from sklearn.cluster import KMeans
kms=KMeans(n_clusters=3)
y=kms.fit_predict(x)
print(y)
```

5.3.4　数据分类

数据分类是数据分析处理中最基本的方法。数据分类通常的做法是，基于样本数据先训练构建分类函数或者分类模型（也称为分类器），该分类器具有将待分类数据项映射到某一特点类别的功能。数据分类和回归分析都可用于预测，预测是指从基于样本数据记录，根据分类准则自动给出对未知数据的推广描述，从而实现对未知数据进行预测。

贝叶斯分类技术在众多分类技术中占有重要地位，也属于统计学分类的范畴，是一种非规则的分类方法。贝叶斯分类技术通过对已分类的样本子集进行训练，学习归纳出分类函数（对离散变量的预测称作分类，对连续变量的分类称为回归），利用训练得到的分类器实现对未分类数据的分类。



如表5-4所示是某网络商城客户购物行为特征的一组统计资料。已知某客户购物行为特征A为数值182.8，特征B为数值58.9，特征C为数值26，请问这人是重要客户还是普通客户？

表5-4　客户购物行为特征表

客户	特征A	特征B	特征C
重要客户	182.8	81.6	30
重要客户	180.4	86.1	29
重要客户	170.0	77.1	30
重要客户	180.4	74.8	28
普通客户	152.4	45.3	24
普通客户	167.6	68.0	26

微信公众号：电子课本大全

（续表）

客户	特征A	特征B	特征C
普通客户	165.2	58.9	25
普通客户	175.2	68.0	27

根据表5-4的资料，得到一个已分类的样本子集：

X=[[182.8,81.6,30],[180.4,86.1,29],[170.0,77.1,30],[180.4,74.8,28],[152.4,45.3,24],[167.6,68.0,26],[165.2,58.9,25],[175.2,68.0,27]]和Y=[1,1,1,1,0,0,0,0]（1代表重要客户，0代表普通客户）。

打开教科书配套学习资源包“第五章\课本素材\程序5-6”文件，把已分类的样本子集（X,Y）和测试样本[182.8,58.9,26]输入到程序5-6中，具体（高斯朴素贝叶斯）如下：

程序5-6

```
import numpy as np
X=np.array([[182.8,81.6,30],[180.4,86.1,29],[170.0,77.1,30],[180.4,74.8,28],
[152.4,45.3,24],[167.6,68.0,26],[165.2,58.9,25],[175.2,68.0,27]])
Y=np.array([1,1,1,1,0,0,0,0])
from sklearn.naive_bayes import GaussianNB
clf=GaussianNB().fit(X,Y)
print(clf.predict([[182.8,58.9,26]]))
```

运行程序5-6后，得到结果为：[0]。由此可以判断这人为普通客户。

程序5-6使用语句from sklearn.naive_bayes import GaussianNB调用了sklearn.naive_bayes模块，使得程序简单明了。

拓展

贝叶斯与概率分类

贝叶斯（Thomas Bayes, 1701—1761），英国数学家，发明了概率统计学原理，将归纳推理法用于概率论基础理论，创立了贝叶斯统计理论，对统计决策函数、统计推断、统计的估算等做出了贡献。

在概率统计理论中，条件概率是指事件A在另外一个事件B已经发生条件下的发生概率，表示为$P(A|B)$，读作“在B的条件下A的概率”。贝叶斯发现在事件B出现的前提下事件A出现的概率，等于事件A出现的前提下事件B出现的概率乘以事件A出现的概率再除以事件B出现的概率。这就是著名的贝叶斯定理。具体计算公式为：

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

微信公众号：电子课本大全

运用贝叶斯定理对事物进行分类，是一种非常有效的思维方法，是贝叶斯决策理论方法的基本思想。例如，假设有一个数据集，由两类组成，且已知每个样本的分类，数据分布如图5-11所示。用$p_1(x,y)$表示数据点(x,y)属于红色一类的概率，用$p_2(x,y)$表示数据点(x,y)属于蓝色一类的概率。

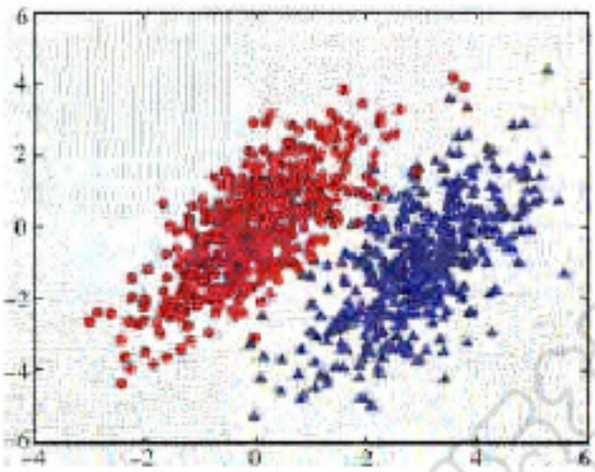


图5-11　数据分布

那么如何判断对于一个新的点$C(x,y)$属于红色还是蓝色类别呢？通常人们会按以下步骤解答这个问题：

（1）求新的点$C(x,y)$属于红色一类的概率$p_1(x,y)$。

（2）求新的点$C(x,y)$属于蓝色一类的概率$p_2(x,y)$。

（3）选择概率高的一类作为新点$C(x,y)$的分类。即如果$p_1(x,y)>p_2(x,y)$，则$C(x,y)$为红色一类；如果$p_1(x,y)<p_2(x,y)$，则$C(x,y)$为蓝色一类。

同样，若有多个分类，而要求新点属于哪一类，只需求出新点在所有类别中概率最大的一类。这就是贝叶斯决策理论的核心思想，即选择具有最高概率的决策。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，完成相应的数据分析。

1. 采用适当的方法完成相应项目选题的数据分析工作。
2. 总结和归纳数据分析的方法和步骤。

5.4　数据的可视化表达

在大数据时代，复杂繁多且不同类型的数据大量涌来，往往超出了人们的处理能力，人类的大脑也难以从堆积如山的数据中快速发现核心问题，而数据可视化可以把枯燥乏味

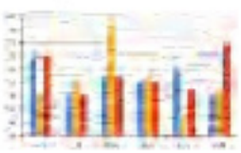
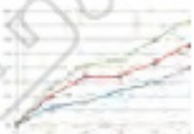
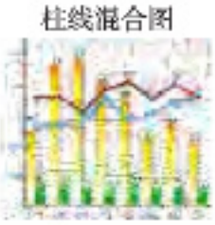
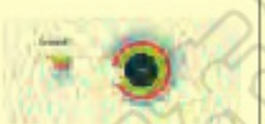
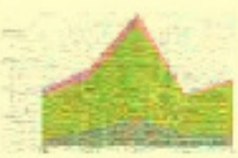
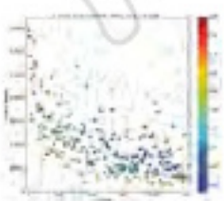
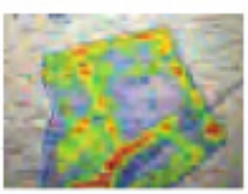
微信公众号：电子课本大全

的海量数据以丰富的视觉效果呈现数据所反映的本质问题，有效提升数据分析的效率。数据可视化是指以图形、图像、地图、动画等生动、易于理解的方式展示数据和诠释数据之间的关系、趋势与规律等，以便更好地理解数据。

5.4.1 数据可视化表达的方式

数据可视化可以让人们快速抓住要点信息，让关键的数据点从人们的眼睛快速通往心灵深处。从常用和实用的维度，数据可视化的呈现类型主要分为探索和解释两种不同的类型。探索类型可以帮助人们发现数据背后的价值，而解释类型则把数据简单明了地解释给人们。按照处理数据类型的不同，可视化呈现的主要方式如表5-5所示。

表5-5 数据分析类型及其对应的可视化呈现

<table>
<tr><th colspan="2">数据分析类型</th><th colspan="4">可视化呈现类型</th></tr>
<tr><td colspan="2">有关趋势的分析</td><td>分组柱形图
</td><td>堆叠柱形图
</td><td>折线图
</td><td>柱线混合图
</td></tr>
<tr><td colspan="2">有关比例的分析</td><td>百分比圆环图
</td><td>饼图
</td><td>圆图
</td><td>堆叠面积图
</td></tr>
<tr><td rowspan="2">有关关系的分析</td><td>逻辑关系</td><td>散点图
</td><td>雷达图
</td><td>网络关系图
</td><td>词云
</td></tr>
<tr><td>空间关系</td><td colspan="2">数据地图
</td><td>动态热力图
</td><td>3D动态显示图
</td></tr>
</table>

微信公众号：电子课本大全

探究活动

体验

随着技术的发展，分享和传达数据信息的方式不仅具有视觉冲击力，更具备丰富的交互功能，如数据地图、词云、热力图、树图、网络图等。如在网络上经常看到的词云图（图5-12）就可以使得过去难以实现的非结构化数据文本挖掘成为轻而易举的事情。运用在线词云图生成网站，尝试复制日记、作文、演讲稿、学期计划、新闻、社交网络的聊天记录等文字信息，个性化地生成词云图，将文本挖掘可视化。



图5-12　词云图

5.4.2　数据可视化表达的工具

数据可视化呈现的形式千变万化，实现的工具众多。其中，Python语言因其开源和包容的特性，嵌入了大量数据可视化的工具，深受大家的喜爱。在Python中，除前面我们经常用到的绘图工具模块Matplotlib外，专业的数据可视化工具模块还包括Seaborn和Bokeh等。

1. Seaborn

Seaborn主要关注统计模型的可视化。例如，直方图既可以总结数据，也可以描绘总体分布。Seaborn基于且高度依赖于Matplotlib。

实践

打开并运行教科书配套学习资源包“第五章\课本素材\直方图程序5-7”文件，结果如图5-13所示。

微信公众号：电子课本大全

```
In [1]: import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        sns.set(palette='muted', color_codes=True)
        rs = np.random.RandomState(10)
        d = rs.normal(size=100)
        f, axes = plt.subplots(2, 2, figsize=(7, 7), sharex=True)
        sns.distplot(d, kde=False, color="b", ax=axes[0, 0])
        sns.distplot(d, hist=False, rug=True, color="r", ax=axes[0, 1])
        sns.distplot(d, hist=False, color="g", kde_kws={'shade': True}, ax=axes[1, 0])
        sns.distplot(d, color="m", ax=axes[1, 1])
        plt.show()
```

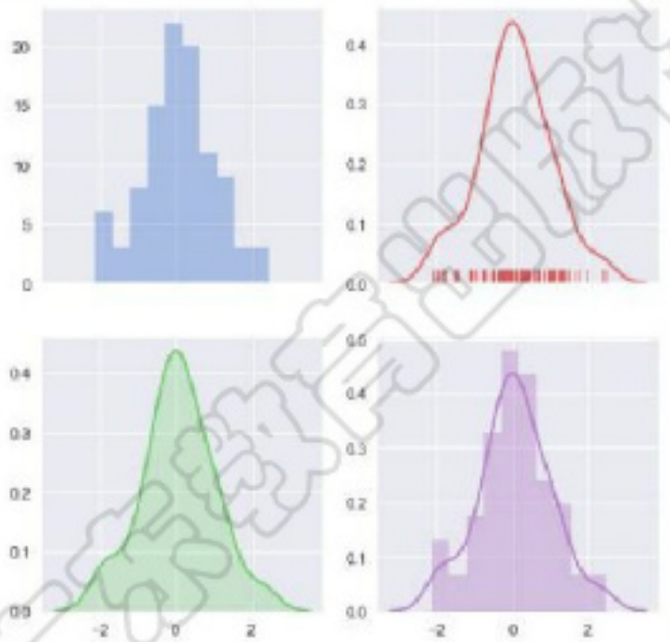


图5-13 直方图程序5-7

2. Bokeh

Bokeh也是一个很好的可视化库，可实现交互式可视化。与其他库相反，Bokeh是独立于Matplotlib的。Bokeh的重点在其交互性，且是通过浏览器以数据驱动文档的风格呈现。

观实

打开教科书配套学习资源包“第五章\课本素材\Bokeh示例程序5-8”文件，结果如图5-14所示。

微信公众号：电子课本大全

```
In [2]: import numpy as np

        from bokeh.layouts import gridplot
        from bokeh.plotting import figure, output_file, show

        N = 100
        x = np.linspace(0, 4*np.pi, N)
        y0 = np.sin(x)

        output_file("sinewave.html")

        sine = figure(width=600, plot_height=600, title='Sine')
        sine.circle(x, y0, size=10, color="navy", alpha=0.5)

        p = gridplot([[sine]], toolbar_location=None)

        show(p)
```

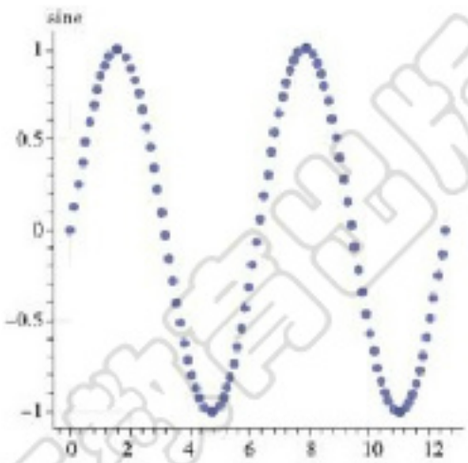


图5-14　Bokeh示例程序5-8

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，了解数据可视化表达的方式与工具，进一步完善该项目方案中的各项学习活动，并参照项目范例的样式，撰写相应的项目成果报告。

成果交流

各小组运用数字化学习工具，将所完成的项目成果，在小组或班级上进行展示与交流，共享创造、分享快乐。

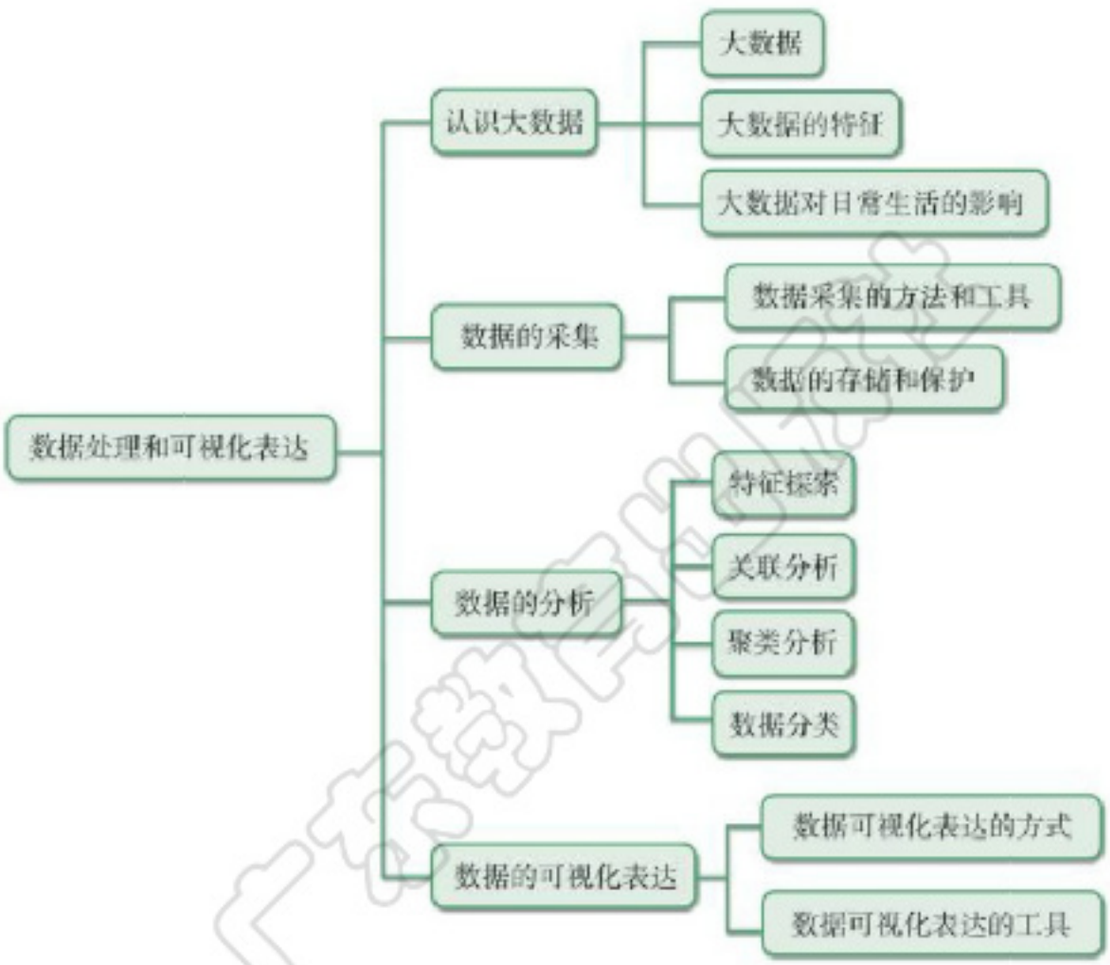
活动评价

各小组根据项目选题、拟订的项目方案、实施情况以及所形成的项目成果，利用教科书附录2的“项目活动评价表”，开展项目学习活动评价。

微信公众号：电子课本大全

本章扼要回顾

同学们通过本章学习，根据“数据处理和可视化表达”知识结构图，扼要回顾、总结、归纳学过的内容，建立自己的知识结构体系。



回顾与总结

微信公众号：电子课本大全

本章学业评价

同学们完成下列测试题（更多的测试题可以在教科书的配套学习资源包中查看），并通过“本章扼要回顾”以及本章的项目活动评价，综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。

1. 单选题

（1）从互联网产生大数据的角度来看，大数据具有的特征是（　　）。

A. “4V”特征：大量（Volume）、多样（Variety）、低价值密度（Value）、高速（Velocity）

B. 样本渐趋于总体，精确让位于模糊，相关性重于因果

C. 分布式存储，分布式并行计算

D. 没有特征

（2）下列可以用于分析数据趋势的是（　　）。

A. 饼图　　B. 折线图　　C. 动态热力图　　D. 词云图

（3）数据分析的方法不包括（　　）。

A. 线性分析　　B. 关联分析　　C. 聚类分析　　D. 数据分类

2. 思考题

若要求你对本班同学喜欢看的书籍进行分类统计，并对其进行数据分析，你会如何做？谈谈你的想法。

3. 情境题

智能手环作为一种智能可穿戴设备，可以记录用户日常生活中的锻炼、睡眠、身体状况等实时数据，并利用数据给用户提供健康建议。简要分析智能手环采集数据的过程，并分析智能手环是如何利用采集的数据给予用户建议的。



智能手环

微信公众号：电子课本大全

第六章
人工智能及其应用

自计算机诞生以来，人类就提出并努力实现最具雄心的目标：制造真正像人一样会思考的机器。随着时光的飞逝和科学技术的高速发展，人类离实现这个目标越来越近，未来的一抹曙光也正在天际逐渐明朗。当前，人类已经可以利用人工智能技术实现汽车无人驾驶、人脸识别、新药研发等。人工智能已经走出了科幻小说与电影，进入人们的现实生活。

本章将通过“剖析企业智能客服机器人”项目，进行自主、协作、探究学习，让同学们在具体感知智能信息系统的基础上，了解智能信息处理的巨大进步和应用潜力，认识人工智能在信息社会中的重要作用，从而将知识建构、技能培养与思维发展融入运用人工智能技术和工具解决问题和完成任务的过程中，促进信息技术学科核心素养达成，完成项目学习目标。

- 认识人工智能
- 人工智能的应用

微信公众号：电子课本大全

项目范例　剖析空调企业智能客服机器人

情境

人工智能技术的高速发展，将对各行各业产生重大的影响。此外，人工智能技术的广泛应用也将深刻地影响人们日常生活。例如，当人们进入某电商平台向客服咨询商品情况时，智能客服机器人（图6–1）将能准确、快速解答咨询的问题，为人们提供高质量的服务体验。



图6–1 智能客服机器人概念图

主题

剖析空调企业智能客服机器人

规划

根据项目范例的主题，在小组中组织讨论，利用思维导图工具，制订项目范例的学习规划，如图6–2所示。

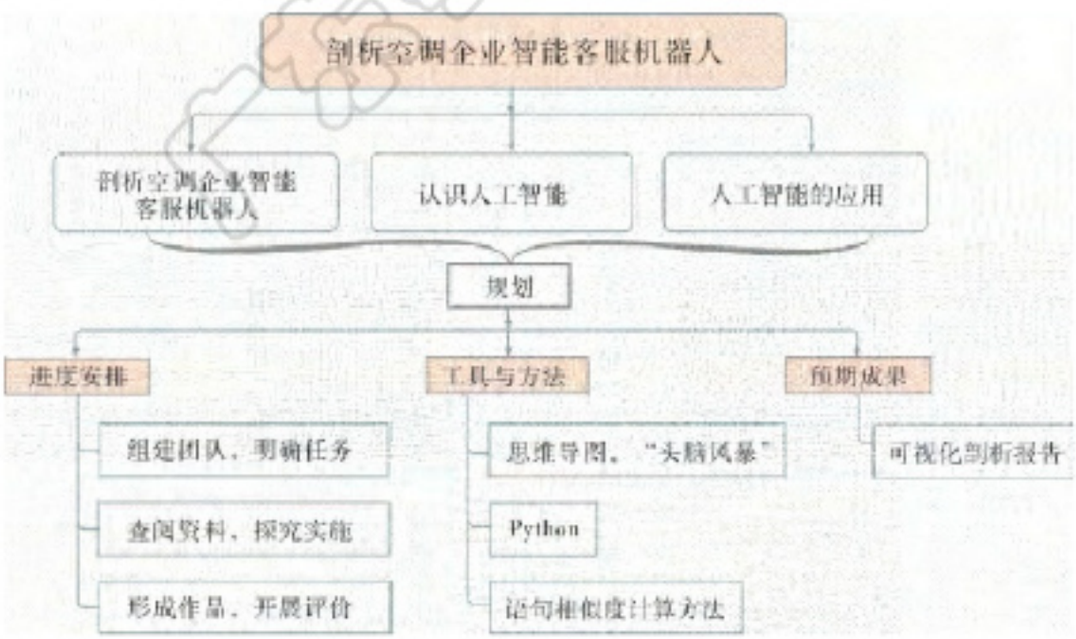


图6–2 “剖析空调企业智能客服机器人”项目学习规划

微信公众号：电子课本大全

探究

根据项目学习规划的安排，通过调查、案例分析、文献阅读和网上资料搜索，开展“剖析空调企业智能客服机器人”项目学习探究活动，如表6-1所示。

表6-1 “剖析空调企业智能客服机器人”项目学习探究活动

探究活动	学习内容	知识技能	
剖析空调企业智能客服机器人	领域知识智能问答系统。	领域知识智能问答系统的概念。	体验人工智能典型应用过程，剖析人工智能典型案例。
	智能信息处理的过程。	常见问题解答模块（FAQ）。	
		文档库（知识库）模块。	
		问题理解模块。	
		信息检索模块。	
		答案抽取模块。	
认识人工智能	人工智能的诞生与发展。	人工智能的概念。	了解智能信息处理的巨大进步和应用潜力。
		人工智能的诞生。	
		人工智能的发展。	
人工智能的应用	人工智能的作用。	智能制造。	认识人工智能在信息社会中的重要作用。
		智能家居。	
		智能教育。	
		智能交通。	
		智能安防。	
		智能医疗。	
		智能物流。	

实施

实施项目学习各项探究活动，进一步剖析空调企业智能客服机器人。

成果

在小组开展项目范例学习过程中，利用思维导图工具梳理小组成员在“头脑风暴”活动中的观点，建立观点结构图，运用多媒体创作工具（如演示文稿、在线编辑工具等），综合加工和表达，形成项目范例可视化学习成果，并通过各种分享平台发布，共享创造、分享快乐。例如，运用在线编辑工具制作的“剖析空调企业智能客服机器人”可视化报告，可以在教科书的配套学习资源包中查看，其目录截图如图6-3所示。

微信公众号：电子课本大全



图6-3 “剖析空调企业智能客服机器人”可视化报告目录截图

评价

根据教科书附录2的“项目活动评价表”，对项目范例的学习过程和学习成果在小组或班级上进行交流，开展项目学习活动评价。

项目选题

同学们以3～6人组成一个小组，选择下面一个参考主题，或者自拟一个感兴趣的主题，开展项目学习。

1. 剖析洗衣机企业智能客服机器人
2. 剖析微波炉企业智能客服机器人
3. 剖析电冰箱企业智能客服机器人

项目规划

各小组根据项目选题，参照项目范例的样式，利用思维导图工具，制订相应的项目方案。

方案交流

各小组将完成的方案在全班进行展示交流，师生共同探讨、完善相应的项目方案。

微信公众号：电子课本大全

6.1 认识人工智能

在移动互联网、大数据、超级计算、传感网、脑科学等新理论新技术以及经济社会发展强烈需求的共同驱动下，人工智能加速发展。

6.1.1 人工智能

人工智能是计算机科学的一个分支，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（如学习、推理、思考、规划等）的学科。人工智能是引领未来的战略性技术，将深刻改变人类生产生活方式。人类要保持对人工智能的控制能力，防范人工智能失控的风险和对人类社会未来发展的潜在威胁。

在互联网时代，人们“遇上新问题，就上网搜索，寻找出答案”已成为一种习惯。然而，一般的信息搜索需要人们在搜索引擎上输入一些关键字，然后在返回的无数个相关网页中，一个一个地打开查看网页内容是否包含自己所需要的信息。这种搜索方式非常耗时，严重影响查找信息的效率。如何快速准确地获取所需要的信息，成为人们亟待解决的问题。

探究活动

体验

利用教科书配套学习资源包提供的虚拟仿真系统，体验某空调生产企业的智能客服机器人所提供的服务，如图6-4所示。

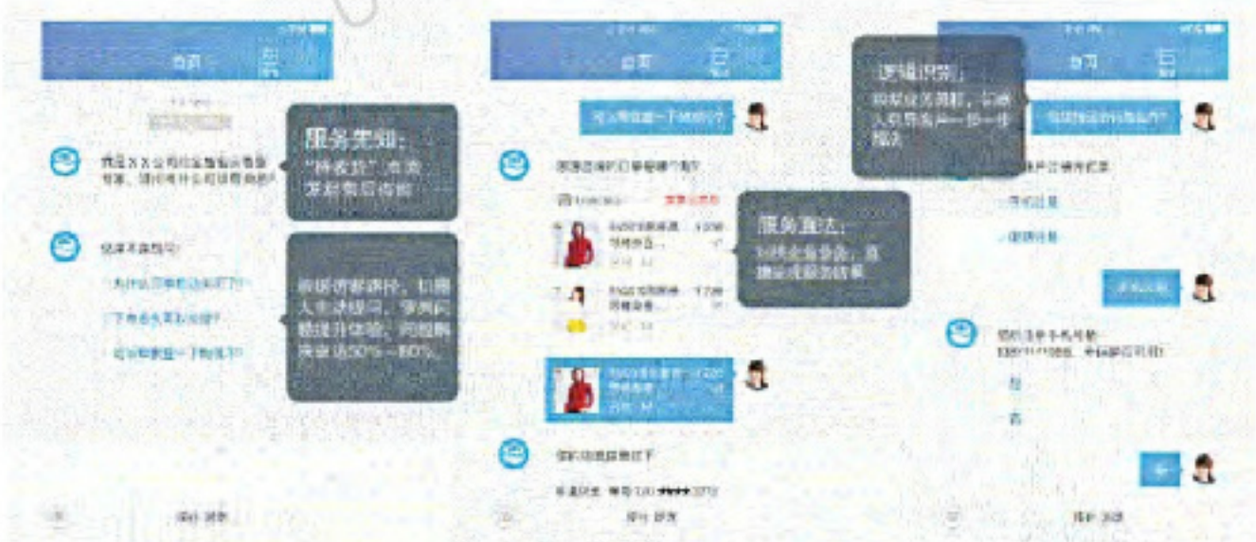


图6-4 智能客服机器人虚拟仿真对话示例截图

微信公众号：电子课本大全

上述的空调企业智能客服机器人是一款领域知识智能问答机器人。它是在某一领域知识范围内实现人机自动问答的智能信息系统，是一种新颖的信息检索系统。它可以准确理解人们的信息检索意图，并通过语料库检索出相关文档和答案，最后从相关的候选答案中提取最接近问题的结果。使用智能问答机器人时，人们可以不采用输入关键字的方式提问，而是直接使用自然语言组成的问句进行提问。因为智能问答机器人采用的自然语言处理技术，可以根据相关领域知识，对人们的提问做必要处理后返回结果，且返回的结果不同于一般搜索引擎返回成千上万的网页，而是通过对答案的抽取，直接从成千上万的网页中快速、准确地检索出所需结果。

6.1.2　剖析领域知识智能问答机器人

领域知识智能问答机器人是在某一领域知识范围内实现人机自动问答的智能信息系统。在智能问答系统中，用户能够直接向机器人提出自己关心的问题；然后机器人根据用户的提问把系统中已有的相似问题的正确答案按问句相似度进行排序，再把排序后的结果作为答案反馈给用户；最后将认为最满意的答案标记为最佳答案。典型的智能问答系统主要包括常见问题解答（FAQ）、问题理解、信息检索、文档库、答案抽取五大模块，结构如图6–5所示。

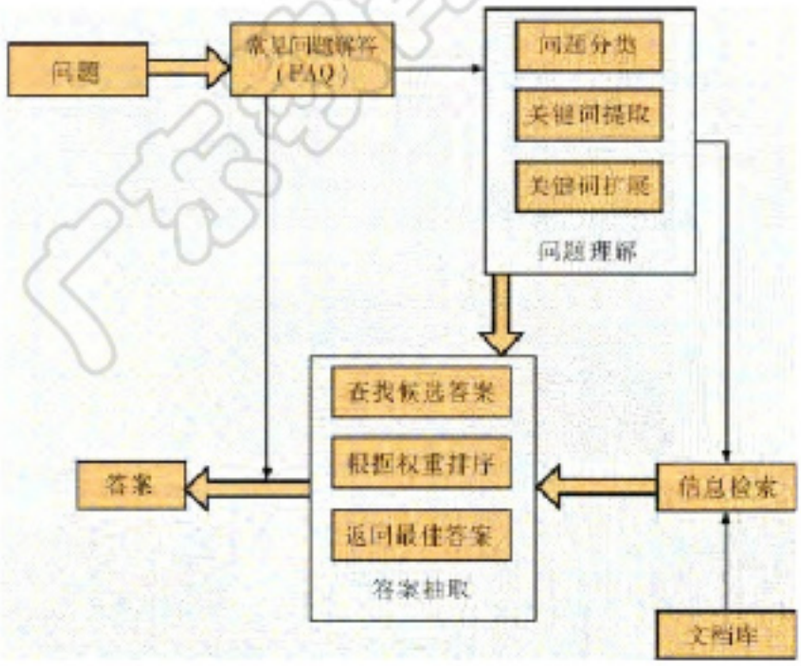


图6–5　典型的智能问答系统结构示意图

1. 常见问题解答模块

用户发起提问后，智能问答系统首先从FAQ库中寻找是否包括用户的问题（问句），如果包含问句，则直接返回问句对应的答案给用户，从而省去后面一系列步骤；否则，就

微信公众号：电子课本大全

进入问题理解模块。FAQ库主要用于提高智能问答系统的效率，其主要技术就是问句相似度计算和候选问句的选择。

问句相似度计算通常采用Jaccard相似度系数算法。Jaccard相似度系数表示两集合的交集元素个数与两集合的并集元素个数之比，系数越高，两集合的相似度越高。例如，计算问句A“空调机有什么功能？”与问句B“空调机的功能有哪些？”的Jaccard相似度系数流程如下：

（1）运用分词系统处理问句。处理后得到集合A={“空调机”，“有”，“什么”，“功能”，“？”}和集合B={“空调机”，“的”，“功能”，“有”，“哪些”，“？”}。

（2）计算集合A和集合B的交集与并集。集合A和集合B的交集={“空调机”，“有”，“功能”，“？”}，集合A和集合B的并集={“空调机”，“有”，“什么”，“功能”，“的”，“哪些”，“？”}。

（3）计算交集元素的数量和并集元素的数量的比值。交集元素的数量为4，并集元素的数量为7，交集元素的数量和并集元素的数量的比值为4÷7≈0.57。因此，将0.57作为问句A和问句B的相似度系数。



基于上述的流程，可以用 Python 语言编程实现 Jaccard 相似度系数算法，程序如下：

```
# -*- coding:utf-8 -*-
# 计算jaccard系数
def jaccard(p,q):
    c=[val for val in p if val in q] #求p和q的交集
    return float(len(c)/(len(p)+len(q)-len(c)))
# 注意：在使用之前必须对两个数据集去重
a=["空调机","有","什么","功能","?"]
b=["空调机","的","功能","有","哪些","?"]
print(jaccard(a,b))
```

程序结果：0.5714285714285714。



计算下列两个语句的Jaccard相似度系数，并利用Python语言编程求出Jaccard相似度系数。

A：我爱祖国的壮丽山河；我爱祖国悠久的历史和璀璨的文化。

B：我爱祖国人民的勤劳、智慧与勇敢；我更爱祖国那生生不息、辉煌雄壮的民族魂！

微信公众号：电子课本大全

2. 问题理解模块

该模块主要实现计算机理解用户的问题，确定问题的关键词和问题的类型，为后面的信息检索和答案提供服务。问题理解模块的实现过程一般包括问题预处理、问题分类、关键词提取和关键词扩展等。其中，问题分类主要确定问题的类别，以方便信息检索和答案抽取。问题理解模块主要运用的技术有分词、同义词词典、分类方法等。

3. 信息检索模块

该模块主要从互联网或者知识库中找到与问题相关的文档作为答案提取的原材料。信息检索的方法一般有两种，一种是直接利用搜索引擎检索信息；另一种是建立特定的知识库，然后根据知识库建立索引模块，从而可以方便、快速地找到相关文档，并根据特点的排序算法对文档进行排序。信息检索模块运用的技术主要包括查询扩展、语料库的构建技术、词汇索引、文档排序等。

4. 文档库模块

文档库用于存放专家提供的知识，其内部含有大量某个领域的常识性知识和专家水平的知识与经验总结，且能够利用专家的知识和解决问题的方法来处理该领域问题。知识的表示形式有产生式、框架、语义网络等，运用得较为普遍的知识是产生式规则。产生式规则以“if...then...”的形式出现，即如果前提条件得到满足，就产生相应的动作或结论。

5. 答案抽取模块

该模块主要利用问题的类型构建相应的答案抽取策略，从信息检索后的文档中对排序靠前的文档进行答案的定位和输出，所用技术主要有答案抽取模板的制定、模式匹配、聚类等。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，开展以下活动。

1. 体验相应的智能客服机器人服务过程，完成对智能客服机器人的剖析工作。
2. 总结归纳智能客服机器人的技术原理以及采用的方法和步骤。

6.1.3　人工智能的诞生与发展

人工智能是通过智能机器延伸、增强人类改造自然和治理社会能力的新兴技术。怎样才能判断一台机器是否具备了思维能力呢？阿兰·图灵提出：在测试人与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。经过多次测试后，如果被测试者超过70%的答复不能使测试人确认出哪个是人、哪个是机器的回答，那么这台机器就通过了测试，并被认为具有人类智能。测试示意图如图6-6所示。



图6-6　图灵测试示意图

微信公众号：电子课本大全

探究活动

讨论

查阅资料，与同学讨论图灵测试的操作过程，并分析图灵测试的意义。

在20世纪50年代，有几位科学家提出想做一台能够像人那样认知、思考和学习的机器，也就是用计算机来模拟人的智能。他们列举了几个准备去攻克的难点，包括系统计算机、用程序对计算机进行编程、神经网络、计算的复杂性、自我学习与提高、抽象等。从此，人工智能便迈上了它的征途，人工智能发展历程大致分为三个阶段，如图6–7所示。

第一阶段（20世纪50年代—80年代）。这一阶段人工智能刚诞生，基于抽象数学推理的可编程数字计算机已经出现，符号主义（Symbolism）快速发展，但由于很多事物不能形式化表达，建立的模型存在一定的局限性。此外，随着计算任务的复杂性不断加大，计算能力远远不能满足需求，人工智能发展一度遇到瓶颈。

第二阶段（20世纪80年代—90年代末）。在这一阶段，专家系统得到快速发展，数学模型有重大突破，但由于专家系统在知识获取、推理能力等方面的不足，以及开发成本高等原因，人工智能的发展又一次进入低谷期。

第三阶段（21世纪初至今）。大数据的积聚、理论算法的革新、计算能力的提升，为人工智能发展提供了丰富的数据资源，协助训练出更加智能化的算法模型。人工智能的发展模式也从过去追求“用计算机模拟人工智能”，逐步转向以机器与人结合而成的增强型混合智能系统，用机器、人、网络结合成新的群智系统，以及用机器、人、网络和物结合成的更加复杂的智能系统。人工智能在很多应用领域取得了突破性进展，迎来了又一个繁荣时期。

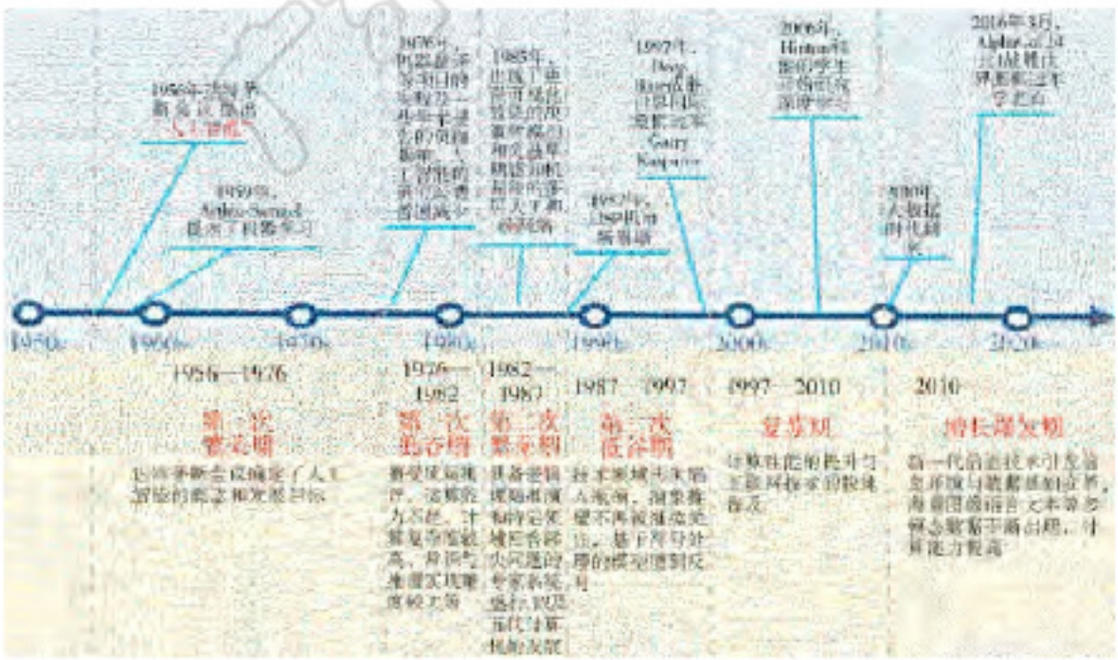


图6–7　人工智能发展历程

微信公众号：电子课本大全

调查

在人工智能的发展历史中，发生了许多具有划时代意义的事件，影响了人工智能发展的方向，如1956年达特茅斯会议确定了人工智能的概念和发展目标等。现以小组为单位，调查人工智能发展历史上发生的重大事件，并阐明其意义。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，开展以下活动。

1. 探索相应领域智能客服机器人的发展过程。
2. 总结归纳领域智能客服机器人发展过程所采用的技术方法。

6.2 人工智能的应用

人工智能的飞速发展，为制造、家居、教育、交通、安防、医疗、物流等各行各业的发展和社会服务带来前所未有的变化，深刻改变人类的社会生活、改变世界，让人们的学习更个性，工作更便捷，生活更美好。

6.2.1 智能制造

智能制造是基于新一代信息通信技术与先进制造技术深度融合，贯穿于设计、生产、管理、服务等制造活动的各个环节，具有自感知、自学习、自决策、自执行、自适应等功能的新型生产方式。智能制造对人工智能的需求主要表现在以下三个方面：一是智能装备，包括自动识别设备、人机交互系统、工业机器人以及数控机床等具体设备，涉及跨媒体分析推理、自然语言处理、虚拟现实智能建模及自主无人系统等关键技术。二是智能工厂，包括智能设计、智能生产、智能管理以及集成优化等具体内容，涉及跨媒体分析推理、大数据智能、机器学习等关键技术。三是智能服务，包括大规模个性化定制、远程维护以及预测性维护等具体服务模式，涉及跨媒体分析推理、自然语言处理、大数据智能、高级机器学习等关键技术。如图6-8所示是智能汽车工厂作业图。



图6-8　智能汽车工厂作业图

微信公众号：电子课本大全

探究活动

体验

通过“剖析空调企业智能客服机器人”，我们知道了智能客服机器人的关键技术是对“知识库模块”的构建，主要是将现有涉及空调企业客服问题的纸质化文件，通过自然语言处理形成数字化资料，然后对形成的数字化资料进行结构化数据处理，形成深度学习所需的训练数据，从而构建空调企业客服问题分析的神经网络和知识库，为下一步进行问题诊断、优化服务质量提供决策依据。打开教科书配套学习资源包“第六章智能客服机器人”仿真系统，运用学习资源包提供的训练数据，修改并运行配套学习资源包“第六章\课本素材\训练模型程序6-1”文件。训练模型程序的逻辑结构如图6-9所示。

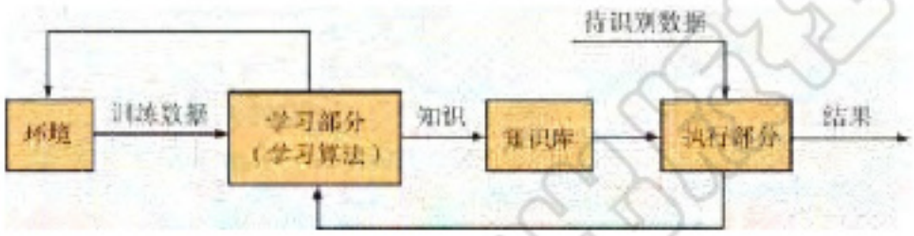


图6-9 训练模型程序的逻辑结构

在程序6-1中，问题库用列表变量q表示，答案库用列表变量a表示。q和a中的元素一一对应，基本句子如下：

```
q=["你好，在吗？","可以货到付款吗?","包邮吗？",……"什么时候发货？","几天能到？",]
a=["在的，欢迎您光临“××官方旗舰店”，我是售前客服小智，请问有什么可以帮您？",……"您好，我们是根据店铺订单量的多少来确定的。正常情况16：00之前付款的，当天都可以发货。订单量过多或者其他特殊情况则是次日发货。"]
```

程序6-1的关键程序段如下：

```
qcut=[]
for i in q:
    data1=""
    thisdata=jieba.cut(i)
    for item in thisdata:
        data1+=item+" "
    qcut.append(data1)
docs=qcut
```


微信公众号：电子课本大全

```
tall=[[w1 for w1 in doc.split()]
            for doc in docs]
dictionary=corpora.Dictionary(tall)
corpus=[dictionary.doc2bow(text) for text in tall]
tfidf=models.TfidfModel(corpus)
num=len(dictionary.token2id.keys())
index = similarities.SparseMatrixSimilarity(tfidf[corpus], num_features=num)
fh=open("dictionary.pk","wb")
pickle.dump(dictionary,fh)
fh.close()
fh=open("tfidf.pk","wb")
pickle.dump(tfidf,fh)
fh.close()
fh=open("index.pk","wb")
pickle.dump(index,fh)
fh.close()
fh=open("a.pk","wb")
pickle.dump(a,fh)
fh.close()
```

6.2.2　智能家居

智能家居以住宅为平台，基于物联网技术，是由硬件（智能家电、智能硬件、安防控制设备、家具等）、软件系统、云计算平台构成的家居生态圈，实现人远程控制设备、设备间互联互通、设备自我学习等功能，并通过收集、分析用户行为数据为用户提供个性化生活服务，使家居生活安全、节能、便捷等。例如，借助智能语音技术，用户应用自然语言实现对家居系统各设备的操控，如开关窗帘（窗户）、操控家用电器和照明系统、打扫卫生等操作；借助机器学习技术，智能电视可以从用户看电视的历史数据中分析其兴趣和爱好，并将相关的节目推荐给用户；通过应用声纹识别、脸部识别、指纹识别等技术进行开锁等；通过大数据技术可以使智能家电实现对自身状态及环境的自我感知，具有故障诊断能力；通过收集产品运行数据，发现产品异常，主动提供服务，降低故障率；还可以通过大数据分析、远程监控和诊断，快速发现问题、解决问题及提高效率。如图6-10所示是指纹识别锁。



图6-10　指纹识别锁

微信公众号：电子课本大全

6.2.3 智能教育

随着人工智能技术的发展，越来越多的人工智能工具被应用于教育领域，成为教师教学和学生学习的得力助手。智能导师是人工智能在教育领域的一个重要应用，它主要通过自然语言处理和语音识别技术，由计算机模拟教师教学的经验和方法，对学生实施一对一的教学，并向具有不同需求和特征的学习者传递知识。它能够在学习者学习的过程中实时跟踪、记录和分析学习者的学习过程与结果，以了解其个性化的学习特点，并根据这一特点为每一名学习者选择合适的学习资源，制订个性化的学习方案。如图6-11所示是智能导师系统结构模型。

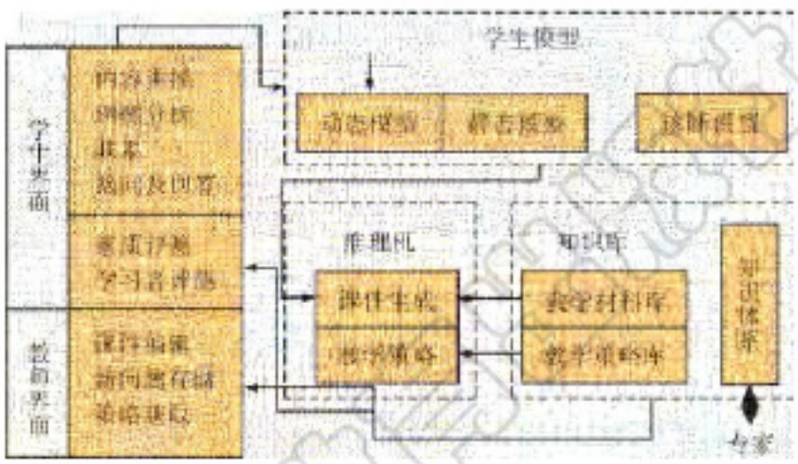


图6-11 智能导师系统结构模型

6.2.4 智能交通

智能交通系统（Intelligent Traffic System，ITS）是通信、信息和控制技术在交通系统中集成应用的产物。ITS 借助现代科技手段和设备，将各核心交通元素联通，实现信息互通与共享以及各交通元素的彼此协调、优化配置和高效使用，形成人、车和交通的一个高效协同环境，建立安全、高效、便捷和低碳的交通运输管理系统。例如，通过交通信息采集系统采集道路中的车辆流量、行车速度等信息，信息分析处理系统处理后形成实时路况，决策系统据此调整道路红绿灯时长，调整可变车道或潮汐车道的通行方向等，信息发布系统将路况推送到导航软件和广播中，让人们合理规划行驶路线。通过不停车电子收费系统（ETC），实现对通过 ETC 入口站的车辆身份及信息自动采集、处理、收费和放行，有效提高通行能力、简化收费管理、降低环境污染。如图6-12所示是智能交通示例图。



图6-12 智能交通示例图

微信公众号：电子课本大全

6.2.5　智能安防

智能安防技术是一种利用人工智能对视频、图像进行存储和分析，从中识别安全隐患并对其进行处理的技术。智能安防与传统安防的最大区别在于智能化，传统安防对人的依赖性比较强，非常耗费人力，而智能安防能够通过机器实现智能判断，从而尽可能实现实时的安全防范和处理。从技术方面来讲，目前智能安防分析技术主要集中在两大类：一类是采用画面分割前景提取等方法对视频画面中的目标进行提取检测，通过不同的规则来区分不同的事件，从而实现不同的判断并产生相应的报警联动等，如区域入侵分析、打架检测、人员聚集分析、交通事件检测等；另一类是利用模式识别技术，对画面中特定的物体进行建模，并通过大量样本进行训练，从而达到对视频画面中的特定物体进行识别，如车辆检测、人脸检测、人头检测（人流统计）等应用。智能安防目前涵盖众多的领域，如街道社区、道路、楼宇建筑、机动车辆的监控，移动物体监测等。今后智能安防还要解决海量视频数据分析、存储控制及传输问题，将智能视频分析技术、云计算及云存储技术结合起来，构建智慧城市下的安防体系。如图6–13所示是安防系统中的摄像头。



图6–13　摄像头

6.2.6　智能医疗

人工智能的快速发展，为医疗健康领域向更高的智能化方向发展提供了非常有利的技术条件。近年，智能医疗在辅助诊疗、疾病预测、医疗影像辅助诊断（图6–14）、药物开发等方面发挥了重要作用。在辅助诊疗方面，通过人工智能技术可以有效提高医护人员的工作效率，提升一线全科医生的诊断治疗水平。例如，利用智能语音技术可以实现电子病历的智能语音录入；利用智能影像识别技术，可以实现医学图像自动读片；利用智能技术和大数据平台，构建辅助诊疗系统。在疾病预测方面，人工智能借助大数据技术可以进行疫情监测，及时有效地预测并防止疫情的进一步扩散和发展。以流感为例，很多国家都有规定，当医生发现新型流感病例时需告知疾病控制与预防中心，但由于人们可能患病不及时就医，同时信息传达回疾控



图6–14　医疗影像辅助诊断

微信公众号：电子课本大全

中心也需要时间，因此，通告新流感病例的时间往往会有一定的延迟，而人工智能通过疫情监测能够有效缩短响应时间。在医疗影像辅助诊断方面，影像判读系统的发展是人工智能技术的产物。早期的影像判读系统主要靠人手工编写判定规则，存在耗时长、临床应用难度大等问题，从而未能得到广泛推广。影像组学是通过医学影像对特征进行提取和分析，为患者预前和预后的诊断与治疗提供评估方法和精准诊疗决策。这在很大程度上简化了人工智能技术的应用流程，节约了人力成本。

拓展

生物特征识别技术

生物特征识别技术是指通过个体生理特征或行为特征对个体身份进行识别认证的技术。从应用流程看，生物特征识别通常分为注册和识别两个阶段。注册阶段通过传感器对人体的生物表征信息进行采集，如利用图像传感器对指纹和人脸等光学信息、麦克风对说话声音等声学信息进行采集，利用数据预处理以及特征提取技术对采集的数据进行处理，得到相应的特征进行存储。识别过程采用与注册过程一致的信息采集方式，对待识别人进行信息采集、数据预处理和特征提取，然后将提取的特征与存储的特征进行比对分析，完成识别。

目前生物特征识别作为重要的智能化身份认证技术，在金融、公共安全、教育、交通等领域得到广泛的应用。

（1）指纹识别。

指纹识别过程通常包括数据采集、数据处理、分析判别三个过程。数据采集通过光、电、力、热等物理传感器获取指纹图像，数据处理包括预处理、畸变校正、特征提取三个过程，分析判别是对提取的特征进行分析判别的过程。

（2）人脸识别。

人脸识别是典型的计算机视觉应用，从应用过程来看，可将人脸识别技术划分为检测定位、面部特征提取以及人脸确认三个过程。人脸识别技术的应用主要受到光照、拍摄角度、图像遮挡、年龄等多个因素的影响，在约束条件下人脸识别技术相对成熟，在自由条件下人脸识别技术还在不断改进。

（3）指静脉识别。

指静脉识别是利用了人体静脉血管中的脱氧血红蛋白对特定波长范围内的近红外线有很好的吸收作用这一特性，采用近红外光对指静脉进行成像与识别的技术。由于指静脉血管分布随机性很强，其网络特征具有很好的唯一性，且属于人体内部特征，不受到外界影响，因此模态特性十分稳定。指静脉识别技术应用面临的主要难题来自成像单元。

（4）声纹识别。

声纹识别是指根据待识别语音的声纹特征识别说话人的技术。声纹识别技术通常可以分为前端处理和建模分析两个阶段。声纹识别的过程是将某段来自某个人的语音

微信公众号：电子课本大全

经过特征提取后与多复合声纹模型库中的声纹模型进行匹配，常用的识别方法可以分为模板匹配法、概率模型法等。

6.2.7　智能物流

智能物流就是利用条形码、射频识别技术、传感器、全球定位系统等方面优化改善运输、仓储、配送装卸等物流业基本活动，同时也在尝试使用智能搜索、推理规划、计算机视觉以及智能机器人等技术，实现货物运输过程的自动化运作和高效率优化管理，提高物流效率。例如，在仓储环节，利用大数据智能分析大量历史库存数据，建立相关预测模型，实现物流库存商品的动态调整。大数据智能也可以支撑商品配送规划，进而实现物流供给与需求匹配、物流资源优化与配置等。在货物搬运环节，加载了计算机视觉、动态路径规划等技术的智能搬运机器人（如搬运机器人、货架穿梭车、分拣机器人等）得到广泛应用，大大减少了订单出库时间，使物流仓库的存储密度、搬运的速度、拣选的精度均有大幅度提高。如图6–15所示是分拣机器人概念图。



图6–15　分拣机器人概念图

中国正处在一个新时代，人工智能的迅速发展，正在让老百姓生活更为便捷、工作更为轻松、学习更为智能，让企业发展更具活力，让国家实力更加强大！



结合实际，讨论在生活中具体应用人工智能技术的例子，感受人工智能对社会发展的影响，并上网了解更多应用人工智能的其他领域，与同学一起讨论、分享。



1997年，深蓝（Deep Blue）战胜了世界国际象棋冠军卡斯帕罗夫（Garry Kasparov），让人类见识了人工智能的强大。2016年3月，AlphaGo以4比1总比分战胜世界围棋冠军李世石，更引发了人类对人工智能的激烈讨论。有人认为人工智能可以给社会各领域带来巨大

微信公众号：电子课本大全

变化，但同时也会引发社会问题。据此，与同学们一起交流，如何正确认识人工智能对人类社会的影响，思考人工智能广泛使用可能会引发的社会问题及应对策略。

项目实施

各小组根据项目选题及拟订的项目方案，结合本节所学知识，了解人工智能的广泛应用，进一步完善该项目方案中的各项学习活动，并参照项目范例的样式，撰写相应的项目成果报告。

成果交流

各小组运用数字化学习工具，将所完成的项目成果，在小组或班级上进行展示与交流，共享创造、分享快乐。

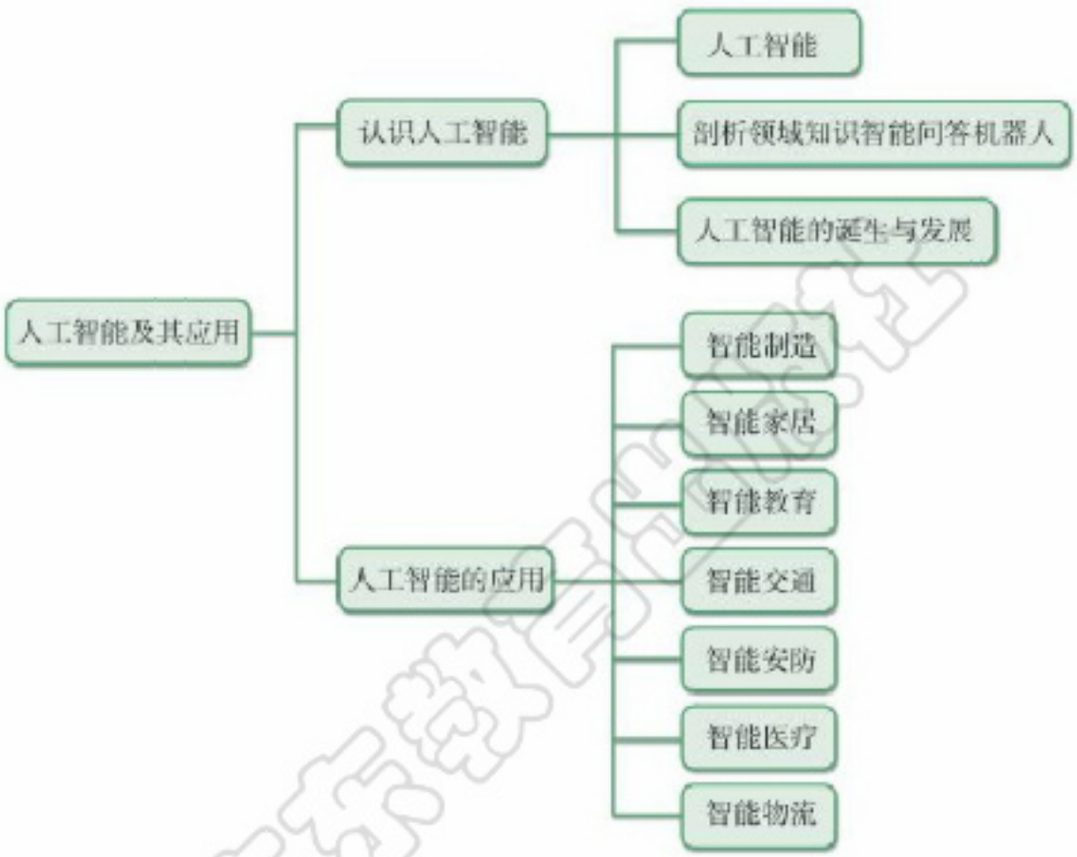
活动评价

各小组根据项目选题、拟订的项目方案、实施情况以及所形成的项目成果，利用教科书附录2的“项目活动评价表”，开展项目学习活动评价。

微信公众号：电子课本大全

本章扼要回顾

同学们通过本章学习，根据“人工智能及其应用”知识结构图，扼要回顾、总结、归纳学过的内容，建立自己的知识结构体系。



回顾与总结

微信公众号：电子课本大全

本章学业评价

同学们完成下列测试题（更多的测试题可以在教科书的配套学习资源包中查看），并通过“本章拓要回顾”以及本章的项目活动评价，综合评价自己在信息技术知识与技能、解决实际问题的过程与方法，以及相关情感态度与价值观的形成等方面，是否达到了本章的学习目标。

1. 单选题

（1）典型的智能问答系统中，主要负责知识的存储的模块是（　　）。

A. 文档库模块

B. 信息检索模块

C. 答案抽取模块

D. 常见问题解答模块

（2）到目前为止，人工智能的发展历史大致可分为（　　）。

A. 两个阶段

B. 三个阶段

C. 四个阶段

D. 五个阶段

（3）智能医疗可以在（　　）等方面发挥重要作用。

A. 辅助诊疗、疾病预测、医疗影像辅助诊断、药物开发

B. 货物搬运、仓储管理、货物配送

C. 人脸识别、指纹解锁、产品检测

D. 远程维护、企业管理、个性化定制

2. 思考题

如下图所示为智能手机上语音输入后即时识别成文字输入的截图，准确率高，极大地提升了人们汉字输入的速度。思考语音输入法主要运用了什么技术，并说说此技术的特点。



微信公众号：电子课本大全

3. 情境题

随着人工智能技术在各个领域的广泛运用，越来越多的个人信息被记录和分析，不可避免地涉及个人隐私的合理使用问题。例如，有一些工具利用人工智能技术在采集个人数据后，可以推导出与其有关的其他人员（如朋友、亲人、同事）信息（如在线行为、人际关系等），这难免引起我们担忧的隐私泄露问题（如下图）。如何解决人工智能技术发展所带来的隐私泄露问题，也正成为社会必须面对的挑战。请谈谈你对如何正确看待人工智能时代个人信息数据的合理使用以及保护个人隐私问题的想法。



微信公众号：电子课本大全

附录1　部分术语、缩略语中英文对照表

ASCII（American Standard Code for Information Interchange）	美国信息交换标准代码（1）
API（Application Programming Interface）	应用程序接口（5）
BMP（Bitmap）	标准图像文件格式（1）
Cookies	网络缓存（5）
E-Learning（Electronic Learning）	数字化学习（2）
ETC（Electronic Toll Collection）	不停车电子收费系统（6）
FAQ（Frequently Asked Questions）	常见问题解答（6）
FTP（File Transfer Protocol）	文件传输协议（5）
GBK（Chinese Internal Code Specification）	汉字内码扩展规范（1）
ITS（Intelligent Traffic System）	智能交通系统（6）
MOOC（Massive Open Online Courses）	大型开放式网络课程（2）
Python	程序设计语言（2）
Unicode	统一码（1）
URL（Uniform Resource Locator）	统一资源定位符（5）

微信公众号：电子课本大全

附录2　项目活动评价表

以培养信息素养为目标，以知识体系为载体，以项目学习活动过程与评价为途径，促进同学们的信息技术学科核心素养达成。

项目学习主题：_____

项目学习过程	学科核心素养达成	一级指标	二级指标	评价结果	支撑材料
选定项目	从现实世界中选择明确的项目主题，形成对信息的敏感度和信息价值的判断力。 分析项目目标与可行性。	项目选题	从现实世界选择项目主题的能力。 化抽象概念为现实问题的能力。 对信息的敏感度和价值的判断力。	□优秀 □良好 □中等 □仍需努力	
		项目分析	分析项目目标的能力。 分析项目可行性的能力。 从现实世界发现项目素材的能力。	□优秀 □良好 □中等 □仍需努力	
规划设计	组建团队与明确项目任务，体现正确的信息社会责任意识。 规划项目与交流方案。	项目规划	组建团队与明确项目任务的能力。 规划项目学习工具与方法的能力。 预期项目成果的能力。	□优秀 □良好 □中等 □仍需努力	
		方案交流	交流项目方案的能力。 完善项目方案的能力。 体现正确的信息社会责任意识。	□优秀 □良好 □中等 □仍需努力	
活动探究	通过团队合作，围绕项目进行自主、协作学习。 开展探究活动，提升信息获取、处理与应用、创新能力。	团队合作	自主学习能力。 分工与协作能力。 交流与沟通能力。	□优秀 □良好 □中等 □仍需努力	
		探究活动	信息获取与处理能力。 探究与联想能力。 实践与创新能力。	□优秀 □良好 □中等 □仍需努力	

微信公众号：电子课本大全

（续表）

项目学习过程	学科核心素养达成	一级指标	二级指标	评价结果	支撑材料
项目实施	针对给定的任务进行分解，明确需要解决的关键问题，并采用计算机科学领域的思想方法，在形成问题解决方案的过程中产生一系列思维活动。 完成方案中预设的目标。	工具方法	采用计算机领域的思想方法能力。 使用数字化工具与资源能力。 数字化学习能力。	□优秀 □良好 □中等 □仍需努力	
		实施方案	针对给定的任务进行分解。 明确需要解决的关键问题。 完成方案中预设的目标。	□优秀 □良好 □中等 □仍需努力	
项目成果交流与评价	与团队成员共享创造与分享快乐，提升批判性思维能力与信息社会责任感。 评价项目目标与成果质量效果。	成果交流	清晰表达项目主题与过程。 与团队成员共享创造与分享快乐。 提升批判性思维能力与信息社会责任感。	□优秀 □良好 □中等 □仍需努力	
		项目评价	运用新知识与技能实现项目目标。 项目成果的可视化表达质量。 项目成果解决现实问题效果。	□优秀 □良好 □中等 □仍需努力	
综合评价	□优秀　□良好　□中等　□仍需努力				

注：1. 评价得分90－100分为优秀（A）；75－89分为良好（B）；60－74分为中等（C）；60分以下为仍需努力（D）。

2. 综合得分=互评×30%+自评×30%+教师评×40%。

微信公众号：电子课本大全



广东教育出版社

微信公众号：电子课本大全