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

人民教育出版社 中国地图出版社

普通高中教科书

信息技术

必修 1

数据与计算

人民教育出版社课程教材研究所信息技术课程教材研究开发中心
中国地图出版社教材出版分社
编著

总主编 祝智庭 樊 磊

人民教育出版社 中国地图出版社
·北京·

总 主 编：祝智庭　樊　磊
副总主编：高淑印　郭　芳　李　锋
本册主编：李　锋　高淑印

编写人员：程建娜　刘姝弘　夏燕萍　王　岚　史弘文

责任编辑：兰大鹏　慈黎利
美术编辑：李　媛　徐海燕

普通高中教科书　信息技术　必修 1　数据与计算
人民教育出版社课程教材研究所信息技术课程教材研究开发中心
中国地图出版社教材出版分社　编著

出　　版　人民教育出版社
（北京市海淀区中关村南大街 17 号院 1 号楼　邮编：100081）
中国地图出版社
（北京市西城区白纸坊西街 3 号　邮编：100054）
网　　址　http://www.pep.com.cn
http://www.ditu.cn
电子邮箱　sinomaps@yeah.net
发　　行
印　　刷
版　　次　2019 年 6 月第 1 版
印　　次　2019 年 6 月第 1 次印刷
开　　本　890 毫米 ×1240 毫米　1/16
印　　张　10
字　　数　200 千字
印　　数
书　　号　ISBN 978-7-107-33845-8
定　　价
审 图 号　GS（2019）3861 号

版权所有 · 未经许可不得采用任何方式擅自复制或使用本产品任何部分 · 违者必究
如发现内容质量问题，请登录中小学教材意见反馈平台：jcyjfk.pep.com.cn
如发现印、装质量问题，影响阅读，请与出版社联系调换。电话：010-83543867

前言

同学们，欢迎探索信息技术这个神奇而充满魅力的世界。

在以往的学习、生活中，你们已经积累了许多信息技术方面的知识与技能，例如：在网上查阅资料，用手机与亲朋好友保持联系，使用移动终端、自动柜员机等设备……你们知道这些应用中都包含哪些关键技术，涉及哪些领域吗？怎样有效地利用这些技术帮助我们培养信息意识，提升计算思维，进而通过数字化学习与创新，承担起信息社会责任呢？即将开始的这门课程，会帮助你们对信息技术有更多的认识和思考，获得更丰富的体验和感受。

为了很好地掌握信息技术，希望同学们按以下三个要求去努力。

1. 认真阅读教科书，理解基本概念和原理。信息技术发展非常迅猛、各类信息系统不断涌现，但信息系统的基础和运行体系相对稳定，离不开算法的设计及对数据的利用。只有夯实基础，才能学好本领，跟上时代发展的步伐。

2. 敢于动手，勤于实践。信息技术是一门实践性较强的课程。实践能帮助同学们熟练操作技能，进一步掌握知识。因此，要认真阅读理解每章的主题学习项目，并逐步完成“实践活动”“思考活动”“技术支持”“阅读拓展”等栏目的学习内容，在实践中获取知识和经验。

3. 要有积极探究、锲而不舍的精神。掌握信息技术的知识与技能需要一个过程，不可能一蹴而就。信息技术学科内容非常丰富，各知识点之间联系密切，但名词术语多，有可能令人感到繁杂，甚至产生畏难情绪。学习新知识，首先要知其然，接着通过不断学习，积极动手操作，大胆请教，加深对知识的理解，然后才能知其所以然，在不断的探索过程中取得进步。

本书中涉及的配套资源，可在教科书配套教学资源平台的信息技术栏目中获得。让我们开始一段信息技术新旅程，成长为信息社会中合格的中国公民！

目录

第1章　认识数据与大数据　1

主题学习项目：体质数据促健康　2

1.1　数据、信息与知识　3

1.1.1　感知数据　4

1.1.2　认识信息　6

1.1.3　理解知识　9

1.2　数字化与编码　13

1.2.1　数字化及其作用　14

1.2.2　二进制与数制转换　17

1.2.3　数据编码　18

1.2.4　数据压缩　22

1.3　数据科学与大数据　25

1.3.1　数据科学的兴起　26

1.3.2　大数据及其应用　27

总结评价　36

中国地图出版社
人民教育出版社

第2章　算法与程序实现　37

主题学习项目：编程控灯利出行　38

2.1　解决问题的一般过程和用计算机解决问题　39

2.1.1　解决问题的一般过程　40

2.1.2　用计算机解决问题的过程　41

2.2 算法的概念及描述 47
2.2.1 认识算法 48
2.2.2 描述算法 49
2.3 程序设计基本知识 55
2.3.1 数据类型 56
2.3.2 常量、变量及表达式 57
2.3.3 语句与程序结构 59
2.4 常见算法的程序实现 70
2.4.1 基于解析算法的问题解决 71
2.4.2 基于枚举算法的问题解决 73
2.4.3 算法与程序实现的综合应用 76
总结评价 82

第3章 数据处理与应用 83

主题学习项目：用水分析助决策 84
3.1 数据处理的一般过程 85
3.1.1 数据处理 86
3.1.2 数据处理的过程 88
3.2 数据采集与整理 91
3.2.1 数据采集 92
3.2.2 数据整理 97
3.2.3 数据安全 100

中国地图出版社
人民教育出版社

3.3 数据分析与可视化 105
3.3.1 数据分析 106
3.3.2 数据可视化 113
3.4 数据分析报告与应用 125
3.4.1 数据分析报告 126
3.4.2 数据处理与问题解决 127
总结评价 130

第4章 走进智能时代 131

主题学习项目：智能交互益拓展 132
4.1 认识人工智能 133
4.1.1 人工智能的产生与发展 134
4.1.2 感受人工智能的魅力 136
4.2 利用智能工具解决问题 140
4.2.1 人工智能平台中的智能工具 141
4.2.2 编程调用平台中的智能工具 142
4.3 人工智能的应用与影响 146
4.3.1 人工智能的应用 147
4.3.2 人工智能的影响 149
总结评价 151

项目评价 152

中国地图出版社
人民教育出版社

第 1 章
认识数据与大数据

中医是我国传统医学的瑰宝。有经验的医家，通过“望闻问切”四诊法获取人体信息，经过辨证论治，对人体健康状况做出诊断，提出健身、防病的建议，是为中医的“治未病”理念。

目前，国内一些信息技术企业，让“治未病”理念插上信息化的翅膀，研制出可穿戴的智能设备，实时采集使用者的心率、血压等数据，通过数据比对，实现对身体健康状况的监测。或许不久的将来，这些高度智能化的设备能将望、闻、问、切的信息与科学测量的数据综合起来，研判监测指标变化对设备使用者的健康状况的意义，提示患病风险并提出针对性解决方案，“治未病”将成为人们普遍享有的福利。

当今社会，人们生活在数据世界里，生活中时常应用着数据，享受着数据带来的便利。今天就让我们进入奇妙的数据世界，领略数据的魅力。

在本章的学习中，我们将借助数字化学习工具，认识数据世界，理解数据、信息与知识三者之间的相互关系，以“体质数据促健康”为主题，开展项目活动，体验数据的作用和价值。

中国地图出版社
人民教育出版社

1 主题学习项目：体质数据促健康

项目目标

本章围绕“体质健康”开展项目学习，建议从“个人健康评估、运动习惯养成、体质改善措施”等方面讨论与思考，研究制订增强体质、促进健康的方案与措施，并以多媒体作品的方式进行全班交流。

1. 理解体质数据反映的信息，领会数据、信息与知识三者的关系，使用思维导图设计增强体质的方案。

2. 理解数字化的作用和数据编码的方法，以多媒体作品的方式呈现项目成果。

3. 借助数字化工具协同学习，提高团队合作意识和数字化学习的能力。

项目准备

为完成项目，需做如下准备。

- 全班分成若干小组，建议每组3～5人，明确目标和分工。
- 查阅《国家学生体质健康标准》及其他文献资料，讨论增强体质的方案。
- 收集多媒体素材，选用一款多媒体集成工具，为制作多媒体作品做准备。

项目过程

在学习本章内容的同时开展项目活动。为了保证本项目的顺利完成，要在以下各阶段检查项目的进度。

1 设计方案

确定项目主题和目标，收集体质数据，分析其中的信息，讨论增强体质、促进健康的方法和途径，以思维导图的形式呈现讨论结果。

P12

2 制作作品

根据完成的思维导图，获取相应多媒体素材，完成多媒体作品的初稿。

P24

3 交流作品

结合大数据特征，提出增强体质、促进健康的新方法，完善多媒体作品并交流。

P35

项目总结

完成本章项目后，各小组提交项目学习成果（包括思维导图、多媒体作品等）。开展作品交流与评价，体验小组合作、项目学习和知识分享的过程，认识数据和大数据在信息社会中的作用与价值。

中国科技出版社
中国教育出版社

1.1
数据、信息与知识

学习目标

- 理解数据、信息的概念，能够举例说明两者的区别和联系。
- 领会数据、信息与知识之间的关系，能有意识地开展数字化学习。

体验探索

发现身边的数据

数据应用十分广泛，随着数据处理方式与工具的变革，数据对我们生活和学习的影响越来越大。

为了更好地提高身体素质，小明同学坚持每天跑步锻炼。自从使用了智能运动手环，他可以实时获取自己的运动数据，灵活调整运动方式（如图1.1.1所示）。例如，跑步前可预先设置心率提示上限。当心率过高时，手环可以通过振动、语音等方式进行提醒。此时就要适当放慢步频，调整心率。

图1.1.1　数据改变运动方式

思考：

1. 为了更科学地制订运动计划，智能运动手环还需要采集哪些方面的运动和生理数据？试加以说明。

2. 在日常生活中，你曾经借助哪些数据来支持个人的生活与学习呢？这些数据又是通过怎样的手段或工具获取的呢？

1.1.1 感知数据

日常生活和学习中，我们经常会接触到数据。例如，同学们的体质数据、图书馆中的图书数据、超市中的商品数据、道路上行驶的车辆数据、企业内产品的生产和销售数据等。随着信息社会的发展，人们对数据的认识越来越深刻，对数据的使用也越来越广泛，数据体现出了前所未有的价值。

数据

什么是数据呢？其实数据就在我们身边。例如：日常交流中，我们可能会听到“这位同学的体重是60 kg”“今天的气温是25 ℃”“这件衬衫的价格是49元”等表述，这其中的60，25，49作为数据描述了不同事物的特征。

随着计算机技术的发展，数据有了更丰富的含义。在计算机中，我们能够看到不同格式的文件，例如，“doc”“xls”“ppt”等文件，这些文件所包含的内容也是由数据组成的，通过计算机中的相应软件编辑文稿、制作表格、编排演示文稿，都是对数据进行处理。

步入信息社会后，互联网技术的发展加速了数据的产生与传输，使社会数据量迅猛增长。例如：在物联网中，通过各种传感器可以采集到大量的数据，人们在社交网站的交流过程中也产生着海量数据……这些数据被广泛应用在各行各业中，深刻影响着社会的经济结构与生产方式，改变着人们的工作、学习和生活习惯。

思考活动

用数据描述体质状况

体质指人体在遗传性和获得性的基础上表现出来的功能和形态上相对稳定的固有特性。为了增强体质，保持健康体魄，就有必要了解反映体质的指标数据，有针对性地锻炼身体。

1.从身体形态、身体机能和身体素质等方面，分组讨论，反映体质状况的评价指标有哪些？

2.查阅《国家学生体质健康标准》，分析个人指标数据，思考数据是如何描述体质状况的。

中国地图出版社
人民教育出版社

数据可用于描述事物。例如，通过频率和振幅等数据来描述声音；通过体重指数（BMI）、肺活量等数据来描述体质健康状况；通过学号、姓名和性别等数据来描述学生基本信息。用来描述事物的一系列数据形成了数据集。

数据是信息的载体。例如，居民身份证号码能够反映出生日期和性别等信息；邮政编码代表着所投递邮件的邮局专用代号信息；商品的条形码能够传递出商品的产地、厂家、名称、生产日期等信息；目前在移动设备中广泛使用的二维码则通过链接的电话号码、网址、图片等数据来描述事物，可以用来表示更多种类的信息，应用于多种场景，如

图1.1.2所示。

数据是可加工，可处理的。对于一些未经加工的原始数据，其含义不易理解。例如，数据“147175270170360172”单独看很难说出它的实际意义。当该数据以表1.1.1的方式呈现时，就能判读出它表示的是三名学生的体质信息。可见，对于所获得的原始数据，还需要在具体情境下，按照一定的规则，选用恰当的工具进行处理，才能获得有意义的信息。如今，计算机已成为人们处理数据的一种重要工具。

图1.1.2　二维码应用场景

综上所述，数据是描述事物的符号记录，是信息的载体。在计算机科学中，数据是计算机识别、存储和加工的对象，例如字符、图像和音频等。

表1.1.1　经过处理的数据

序号	体重/kg	身高/cm
1	47	175
2	70	170
3	60	172

数据的应用

如今，数据已广泛应用于社会的方方面面。上网收发电子邮件，使用搜索引擎查找网络信息，借助电子地图进行导航等，都是在与数据打交道。数据正在改变着人们的生活学习和生产方式。

■ 体质数据有助于人们了解身体状况。通过可穿戴设备能及时收集人体数据，从每日步数、人体热量消耗等基础运动数据，到血氧饱和度、血压、心率、皮肤温度和睡眠状况等生理数据，便于个人及时了解身体状况，接受健康方面的建议，调整运动方式乃至生活方式（如图1.1.3所示）。可穿戴设备还可融合云计算、物联网和人工智能等技术，帮助人们更好地获取并处理自身与外部世界的数据，提供个性化的专属服务。

中国地图出版社
人民教育出版社

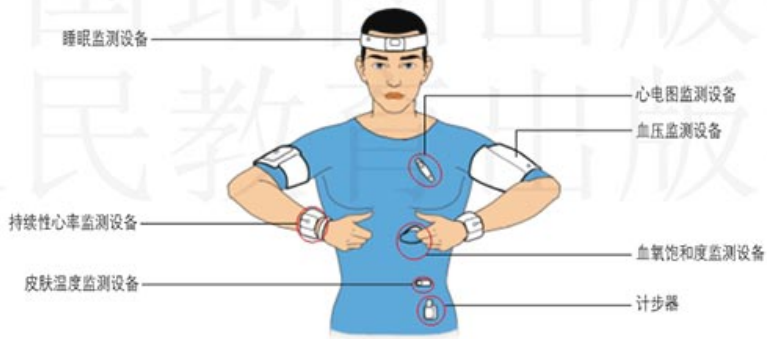


图1.1.3　可穿戴设备

■ 在线学习数据可帮助教师了解学情，调整教学进度，支持学习者的持续发展。通过在线学习平台可以收集学习者的在线行为（如浏览课件、提交作业、练习考试和师生交互

等）数据，对这些数据进行统计分析，可及时判断学习者的学习状况，并可提供个性化的学习改善建议，以便更合理地安排学习进度，提高学习效率。

■ 营销数据有助于企业做出决策、组织生产以及为客户提供个性化服务等。图1.1.4中的数字腕带可以用来采集游客的个人偏好数据，如记录游客来公园的次数，去过哪些地方，参与了什么娱乐项目，在什么时候用餐，喜欢什么食物等。管理部门能够利用这些数据所反映的信息，决定营销策略，合理安排人力和物力等资源，为游客提供个性化的服务，如推送商品信息、自助点餐等，从而提升游客的满意度。

图1.1.4　数字腕带

随着移动互联网的快速普及，信息技术与人们的生产生活逐渐交汇融合，全球数据呈现爆发式增长、海量集聚的特点。数据对改善人民生活、促进经济发展、推动社会进步等方面，起着越来越重要的作用。懂得数据、善于利用数据来解决问题是信息社会公民的一项重要能力。

阅读拓展

从“数字”“数值”到“数据”

计算机问世前，“数据”多指用于统计的“数字”或“数值”。计算机问世初期，其主要用于“数值计算”，如科学研究、军事等方面的科学计算。随着计算机技术的发展，其逐步应用于“数据处理”，如工资、库存管理等。

随着计算机和互联网的普及与发展，现代的“数据”已不再限于“数字”或“数值”，而是有了更多的内涵和更广的外延。“数据”可以是字符、图像、音频和视频等，可以对各行业相关信息进行记录。例如，“交通数据”包括城市道路实况、违章记录和车辆流量等；“教育数据”包括适龄学生人数、教师人数和学习成绩等；“医疗数据”包括病人信息、病历记录和诊断结果等。

现代社会中，数据用途广泛，与人们的生活息息相关，有人把“数据”比作石油和矿石，合理加工后能够生成多种产品和价值。因此，更好地利用数据为我们服务，具有重要的意义。

中国地图出版社
人民教育出版社

1.1.2　认识信息

信息

数据是信息的载体，信息则是数据所表示的意义。信息可以用来消除接收者某种认识上的不确定性。也就是说，当接收者获得信息后，这种“不确定性”就可以减少或消除。信息量的大小可用消除不确定性的多少来表示。例如，播报足球比赛新闻时，收听者预先不知道比赛情况，因此足球新闻报道对急需了解比赛情况的听众来说就是很重要的信息，报道越详细，则信息量越大。假如广播时有外界干扰，收听受到影响，听者所听到的信息也会受到损失。

思考活动

理解信息的涵义

例1：隋唐时期发明的雕版印刷术，在宋代得到广泛推广。但雕刻书版费工费时，保存书版也非常不便。北宋时期的匠人毕昇发明了活字印刷术（图1.1.5）。这种印刷方法方便、省时、经济，大大提高了印刷效率。印刷术发明后，逐渐传到朝鲜和日本，甚至传到中亚、西亚和北非。印刷术的发明，对世界文化的传播、交流和发展起到了巨大的推动作用。

图1.1.5　泥活字板（模型）

例2：每逢大型体育赛事，电视台会现场直播比赛（图1.1.6）。电视台利用现场解说、技术统计分析、比赛慢动作回放及现场比分播报等全方位、多视角的直播手段，使观众能够通过电视机观看比赛实况，身临其境地感受赛场的气氛。

图1.1.6　体育赛事

例3：地图导航软件是出行的好助手（图1.1.7）。开启定位功能，就能使用导航软件确定使用者的位置，规划出行线路，还能随时搜索周边的商店、银行、医院和公交站台等信息。

图1.1.7　地图导航

阅读上述事例，参考表1.1.2中的例1，完成例2和例3的事例分析。

表1.1.2　信息分析表

分析项	活字印刷	电视赛事直播	地图导航
载体类型	☑文字　☐图像 ☐音频　☐视频 ☐其他：_____	☐文字　☐图像 ☐音频　☐视频 ☐其他：_____	☐文字　☐图像 ☐音频　☐视频 ☐其他：_____
传播方式	☐现场交流　☑印刷品 ☐电话　☐广播电视 ☐网络　☐其他：____	☐现场交流　☐印刷品 ☐电话　☐广播电视 ☐网络　☐其他：____	☐现场交流　☐印刷品 ☐电话　☐广播电视 ☐网络　☐其他：____
感知方式	☑视觉　☐听觉 ☐触觉　☐其他：_____	☐视觉　☐听觉 ☐触觉　☐其他：_____	☐视觉　☐听觉 ☐触觉　☐其他：_____
达到效果	☑认识事物　☑形成判断 ☑做出决策　☑指导行动 ☐其他：_____	☐认识事物　☐形成判断 ☐做出决策　☐指导行动 ☐其他：_____	☐认识事物　☐形成判断 ☐做出决策　☐指导行动 ☐其他：_____

中国地图出版社
人民教育出版社

通过上面的案例可以看出，依附字符、图像和音频等载体，通过各种途径进行传播，能够帮助获取者了解情况、形成判断或做出决策的内容可以称之为信息。当人们获取各种形式的信息后，经过分析、判断与理解，把它们变成知识或引发相应的行为，此时信息便发生了作用。例如，听到上课铃声，学生进入教室、保持安静、准备上课；根据交通信号灯的变化，驾驶员和行人决定“行”和“停”的行为；根据个人体质数据，人们采取合理的保健措施。由此可见，在信息社会，信息的有效获取和合理利用已成为一种重要素养。

信息的特征

信息在人际交流、生产管理、知识传播和科学研究等方面都发挥着巨大的作用。了解信息的特征，有助于我们加深对信息的认识和理解。

■ 信息的表示、传播和存储需要依附于某种载体，用来反映其内容和含义。例如，许多古籍通过文字记录了我国古代的历史与文化；体育比赛实况是以音视频形式，通过电视、网络等途径传播的；导航信息则是以地图和语音的形式，通过移动通信网络、卫星通信系统等途径传输的。

■ 信息的传递和共享体现了信息的意义。信息可进行复制、传播或分享，不会因使用人数和次数的累积而耗损其内容。例如，电视台直播的比赛实况可以将信息传播给众多的观众，提高了信息的利用率，还节省了社会成本。信息的无损共享使信息资源能够发挥最大的效用，促进人们的交流与沟通，从而推动社会的进步和发展，如图 1.1.8 所示。

图 1.1.8　信息的传递和共享

■ 信息的产生和利用具有时效性。信息往往反映了事物在某一时间段的状态，而事物状态会随着时间的推移发生变化，信息就可能失去原来的作用。例如，气象信息、交通信息、市场信息等常常瞬息万变、时过境迁，只有善于把握时机，才能发挥信息的效用。

■ 信息的价值和效用因人而异。信息蕴含着价值，信息价值的大小是相对的，它取决于接收信息者的需求及其对信息的理解、认识和应用能力。例如，天气预报中关于台风的信息，对于台风影响到地区的居民和没有影响到地区的居民，其价值是不一样的。

人类很早就将信息应用于生产和生活实践，只是受当时社会环境的制约，信息利用的规模和范围十分有限。如今进入了信息社会，信息技术的变革正强有力地推动着社会的进步。人们借助信息对自然界中的资源进行有效地获取、分配和利用，促进了人类社会的可持续发展。

阅读拓展

香农与信息论

1948年，信息论创始人香农（Claude Elwood Shannon）发表了一篇题为《通信的数学理论》的论文，建立了比较系统的信息理论基础，即信息论。香农为信息科学的发展做出了很大的贡献。

第一，他提出了信息的形式化。香农认为，在通信中要排除信息的语义，只考虑信息的形式因素。在通信中，如果接收端把发送端发出的信息从形式上复制出来，那么也就复制了信息的语义内容。信息形式化后，就有可能用数学进行表示和处理。

第二，他提出了一个通信系统模型，如图1.1.9所示。信息源给出要传输的信息；编码器把信息转换为信号，使之能在信道中传输；解码器把信号恢复成信息传给信息的接收方，即信宿。

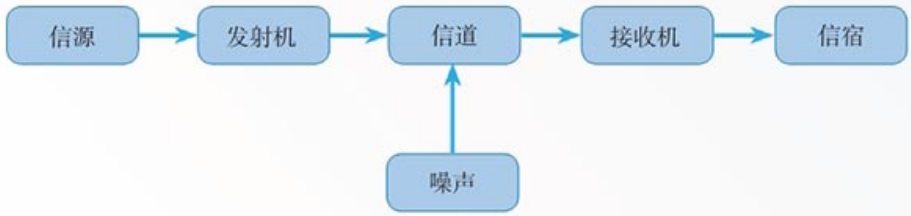


图1.1.9　通信系统模型示意图

第三，他进一步完善了“奈奎斯特－香农采样定理”，阐述了采样频率与信号频谱之间的关系。这个定理最初是由奈奎斯特（Harry Nyquist）在研究电报信号中的失真问题时提出的，香农对奈奎斯特的理论进行了明确说明并正式作为定理引用。采样定理是实现数字化的理论基础，在数字通信、信息处理、采样控制和遥测系统等领域都有应用。

1.1.3　理解知识

知识

知识是人们在社会实践中所获得的认识和经验的总和，是人类在实践中认识世界的成果。“知识就是力量”这句名言一直流传至今。在人类历史文明的长河中，人们可以通

过学习各种各样的资料继承前人的知识，也可以通过实践探究发现知识。经过人们一代一代的探索与传承，人类社会的知识得以不断丰富。

思考活动

医生靠什么诊断病情

某患者发热、咳嗽，去医院检查。医生让他先去做血液化验，其化验单如图1.1.10所示。医生通过化验单确诊病症是由细菌引起的呼吸道感染，并制订了相应的治疗方案。

检验项目	结果	参考区间	提示	单位
1 白细胞计数	10.25	3.50～9.50	↑	10^9/L
2 中性粒细胞数	8.33	1.80～6.30	↑	10^9/L
3 淋巴细胞数	1.36	1.10～3.20		10^9/L
4 单核细胞数	0.49	0.01～0.60		10^9/L
5 嗜酸性粒细胞数	0.02	0.02～0.52		10^9/L
6 嗜碱性粒细胞数	0.05	0.00～0.06		10^9/L
7 中性粒细胞百分比	81.2	40.0～75.0	↑	%
8 淋巴细胞百分比	13.3	20.0～50.0	↓	%
9 单核细胞百分比	4.8	3.0～10.0		%
10 平均红细胞容积	105.6	82.0～100.0	↑	fL
11 嗜酸性粒细胞百分比	0.2	0.4～8.0	↓	%
12 嗜碱性粒细胞百分比	0.5	0.0～1.0		%
13 红细胞计数	3.39	4.3～5.8	↓	10^{12}/L
14 血红蛋白测定	105	115～150	↓	g/L
15 红细胞比积测定	35.8	35.0～45.0		%
16 平均红细胞血红蛋白浓度	293	316～354	↓	g/L
17 平均红细胞血红蛋白含量	31.0	27.0～34.0		pg
18 红细胞分布宽度标准差	62.4	37.0～50.0	↑	fL

图1.1.10　血液化验结果

思考：

患者可以从化验单上看到各种指标数据和提示信息，但很难准确地判断自己的病情。医生同患者相比，在病情认识上的差异体现在哪里？

血液化验单列举的是患者的血液指标数据，其中蕴含着有用的信息。例如，化验指标行中的提示箭头表明相应的指标异常。指标异常的信息很多，如何抓住其中的关键信息来诊断病情呢？在上述案例中，医生通过观察患者血液化验结果的“白细胞计数”和“中性粒细胞百分比”两项指标的异常来对病情进行初步诊断。医生之所以能够迅速从众多化验数据中获取到有价值的信息，主要得益于医疗行业的专业基础知识和实践经验，而这些知识和经验是医务工作者在学习和工作过程中逐步积累起来的。

数据、信息与知识的关系

数据作为描述事物的符号记录，经过处理后，就有可能转换为相应的信息。人们对信息进行提炼和归纳后，获得实践中解决问题的观点、经验和技能，信息才会内化为知识。人们综合运用知识和技能解决问题、发明创造或预测未来时，知识就升华为智慧。表1.1.3以案例的方式说明了数据、信息与知识的关系。

表1.1.3　数据、信息与知识的关系

<table>
<tr><th>概念</th><th>案例</th><th>说明</th></tr>
<tr><td>数据</td><td>147175270170360172</td><td>对事物的符号记录</td></tr>
<tr><td>信息</td><td>列表并计算三名高一男生的体重指数（BMI），计算公式为：体重指数（BMI）=体重/身高^2
<table>
<tr><th>序号</th><th>体重/kg</th><th>身高/m</th><th>BMI/(kg·m^{-2})</th></tr>
<tr><td>1</td><td>47</td><td>1.75</td><td>15.35</td></tr>
<tr><td>2</td><td>70</td><td>1.70</td><td>24.22</td></tr>
<tr><td>3</td><td>60</td><td>1.72</td><td>20.28</td></tr>
</table></td><td>经过处理的数据</td></tr>
<tr><td>知识</td><td>国家有关部门根据学生体质健康数据，进行统计分析，全面了解学生健康状况及变化趋势，制定了《国家学生体质健康标准》，其中高一男生的正常体重指数为16.5 ~ 23.2</td><td>通过组织化、结构化的信息，形成解决问题的方法和途径</td></tr>
</table>

表1.1.3中提到的体重指数对于不同年龄和性别的学生有着不同的标准。例如，对于高一男生来说，体重指数为15.35属低体重，但对于小学一年级男生来说就是正常体重。因此，《国家学生体质健康标准》是针对我国青少年学生制定的，是国家有关部门在长期研究的基础上知识积累的结果，属于知识的范畴。

数据、信息和知识的转换还依赖于它们使用的环境和应用者的知识结构。例如，有专业知识背景的医生可以通过血液化验数据做出有针对性的诊断。因此，在进行数据、信息和知识的关系分析时，就需要结合特定的环境与应用者的知识结构。

在知识快速更新的时代，人们创造了许多知识管理工具，用来认识事物、表达思想，使学习和工作更有效率。思维导图就是一种高效的知识管理工具，它能够帮助学习者结构化地思考问题，直观化地展现思维脉络，使知识的交流与分享更清晰有效。

技术支持　思维导图的绘制方法

思维导图是一个简单、有效的知识管理工具，可以帮助我们进行发散性、结构化的思考，通过文字、图标、线条和颜色等要素，图文并茂地展示思考内容，帮助分析与记忆。思维导图的绘制方法如图1.1.11。

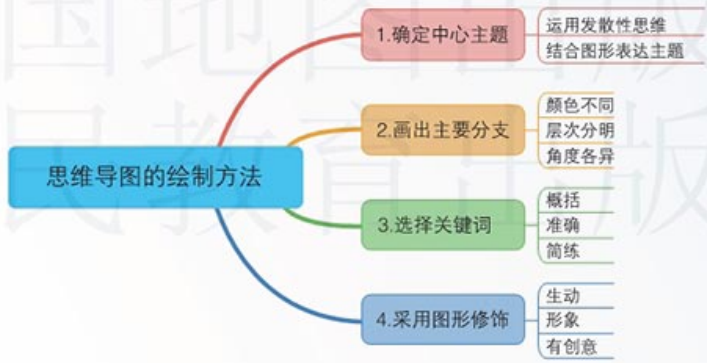


图1.1.11　思维导图的绘制方法

互联网上有许多在线或离线的思维导图绘制工具，教学资源平台中也提供了相关资源或微视频，可以帮助我们获得更多的学习支持。

中国地图出版社
人民教育出版社

项目实施

制订项目方案，获取体质数据

一、项目活动

1. 依据项目目标从项目内容、项目分工、项目过程和项目检查等方面制订项目活动方案。

2. 按照表1.1.4收集小组同学的体质数据，参考《国家学生体质健康标准》，查阅相应的得分和等级，填写到表格中。

表1.1.4　学生体质数据表

学号		性别		年龄	
年级		身高 / m		体重 / kg	
评价指标			成绩	得分	等级
体重指数 /（kg · m^{-2}）					
肺活量 / mL					
50 米跑 / s					
坐位体前屈 / cm					
立定跳远 / cm					
引体向上（男）/ 次					
1 分钟仰卧起坐（女）/ 次					
1 000 米跑（男）/ min					
800 米跑（女）/ min					

3. 参考教学资源平台提供的《体质数据汇总表》，将小组同学的体质数据整理到电子表格中，从“身体形态、身体机能和身体素质”等方面提炼相关信息，从“个人健康评估、运动习惯养成、体质改善措施”等方面开展讨论，制订增强体质、促进健康的方案与措施，用思维导图呈现讨论结果。

二、项目检查

1. 各小组从体育锻炼、日常饮食、生活作息等方面绘制促进体质健康的思维导图，并在全班交流。

2. 依据思维导图，讨论多媒体作品的设计思路，明确分工和任务。

练习提升

1. 参考表1.1.3数据、信息与知识的案例，举例说明三者的关系。

2. 智能运动手环可以记录运动数据和生理数据，这些同时也是个人的隐私数据。其实，智能手机中还包含了更多的用户隐私数据，应该如何保护这些数据呢？

3. 试围绕某一学科的知识要点，绘制思维导图，梳理知识结构体系。

1.2
数字化与编码

学习目标

- 了解数字化的意义与作用，感受数字技术对生活的影响。
- 理解数据编码的基本方式，合理选用数字化工具解决问题，感受数字化工具的优势。

体验探索

体验数字化生活

信息技术的发展与普及为我们创造了一个全新的数字化生活环境。例如，共享单车、在线购物、网络通信和远程医疗等，它们在给我们带来生活便利的同时，也在逐渐地改变着我们的生活方式（如图1.2.1所示）。

图1.2.1　信息技术改变我们的生活

思考：

1. 从使用方式、应用工具、时间效率和操作流程等方面对实体超市购物和在线购物、个人单车和共享单车、书信交流和网络交流等进行比较，分别说出它们的异同。

2. 举例说明在日常生活和学习中自己经常使用哪些数字化工具，分析它们如何改变自己的生活与学习方式。

中国科学技术出版社
人民教育出版社

1.2.1 数字化及其作用

信息技术现已应用到了社会的方方面面。例如，数字图书馆让我们可以足不出户就获取所需的学习资料；公共场所的数字摄像设备可及时捕捉安保信息，有力地保障社会的安全；数字医疗设施的广泛使用，提高了病例筛选的精准度。数字化正悄然改变着我们的学习、生活与工作方式，推动着社会的变革与发展。

模拟信号和数字信号

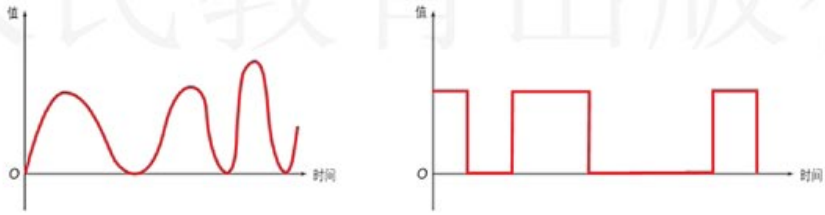
在自然环境中，人们可以感知各种各样的信息。例如，水流速度快慢、环境温度变化、空气污染状况等，这些信息可以通过连续的状态表示出来，也可以通过离散的状态表示出来。例如，水银体温计是通过水银连续移动的状态来表示体温信息的；数字体温计以离散的状态给出体温信息。图 1.2.2 是两种体温计分别用连续和离散的状态来表示温度。

（a）水银体温计　　（b）数字体温计

图 1.2.2　两种常见的体温计

一般来说，通信系统中有两种主要的信息传输形式：一种是模拟信号，其信号波形可以表示为时间的连续函数，如图 1.2.3（a）所示；另一种是数字信号，它在取值上是离散的、不连续的，例如以“0”和“1”来表示二进制数字信号，如图 1.2.3（b）所示。模拟信号和数字信号虽然是两种不同形式的信号，但它们在传输过程中是可以相互转换的。例如，用录音设备采集到的语音最初为模拟信号，通过采样、量化、编码等步骤可以将模拟信号转换为数字信号，并以二进制方式存储在计算机中。同样，也可以将数字信号转换为模拟信号。例如，计算机中的音频文件通过转换设备可将数字信号转换为模拟信号，并通过播放设备播放出来。

人民教育出版社



（a）模拟信号　　（b）数字信号

图 1.2.3　模拟信号和数字信号

思考活动

录像带中视频的数字化

小明家中有一些老式录像带，它们记录着家庭中美好的回忆。他在整理这些录像带时，发现播放这些录像带的设备现在已很难找到了。为了更好地保存这些美好回忆，他希望能把录像带中的视频转换为计算机中可以存储和处理的视频文件。

思考：

查找资料，指出将录像带中的视频转换为计算机中存储和处理的视频文件的工具与方法有哪些。

生活在数字世界

数字化是将复杂多样的事物属性或特征转化为计算机可处理对象的过程。这些信息被数字化后可以更方便地在数字设备中存储、处理和传输。例如，在数字化实现之前，我们获取知识多是通过阅读纸质书籍，异地交流比较多的是通过写信或打电话，听音乐常常是通过录音机，看电影主要使用胶片放映机……这些信息数字化后，可以通过网络线路来传输，阅读、交流、听音乐和看电影都能通过数字设备来完成。数字化让人们的生活、学习和工作变得更加便捷，如图1.2.4所示。

图1.2.4　信息的数字化

当今，社会中复杂多样的信息也可以用数字化形式表达。例如，网络商城的卖家需要通过数字化方式来呈现商品信息：商品的说明书、商品的外观图片、商品的操作视频等都可以用数字化的方式来表示。此外，借助数字化设备我们可以方便地获取身边事物的相关数据，例如，借助温度传感装置可以方便地采集空气温度的数据。通过网络还可以实现物物相连，并能把所获取的数据进行及时的处理和应用。可以说，信息技术的广泛应用已把整个世界变成了一个数字世界，如图1.2.5所示。

中国地图出版社
人民教育出版社

图 1.2.5　数字化改变世界

信息技术的发展给我们的学习和生活带来了深远的影响。例如，数字图书馆方便我们检索信息；数字博物馆方便我们欣赏展品和资料；数字化校园为教师和学生提供了全方位的网络服务，可以提高工作和学习效率。如今，数字化学习资源与工具层出不穷，支持着个性化学习和终身学习。有效地利用数字化学习工具可以帮助我们建构知识，解决问题，实现创新。

阅读拓展

数字图书馆

数字图书馆是虚拟的、没有“围墙”的图书馆，是基于网络环境共建共享的、可扩展的知识网络系统，是超大规模、便于使用、没有时空限制、智能检索的知识中心。

数字图书馆将文本、图像和音视频等数字化信息，通过互联网传输，从而做到信息资源共享。拥有数字终端的用户通过互联网，访问数字图书馆网站，可以方便快捷地享用数字化信息资源。图 1.2.6 展示的是中国国家数字图书馆首页。

图 1.2.6　中国国家数字图书馆首页

信息技术的革新使得数字图书馆从基于信息的处理和简单的人机界面逐步向基于知识的处理和广泛的智能应用发展，可以帮助人们利用计算机和网络更大范围地拓展智力活动，在信息交流、传播、存储等方面，发挥极其重要的作用。

1.2.2 二进制与数制转换

二进制在计算机等数字设备中被广泛应用，其基本规则如下：

- 二进制的基数为2，两个基本数码是0和1；
- 采用逢二进一的进位规则，例如1+1=10；
- 不同的数位对应不同的权值，权值用基数的幂表示。

由于计算机中的操作需要通过二进制来实现，因此计算机中的数值计算就需要将十进制数转换成二进制数。十进制整数转换为二进制数可采用除2反向取余法，即将十进制整数除以2，得到商数和余数，用商数再除以2，依此类推直到商数为0为止，将每次得到的余数按照逆序排列，即为换算的二进制数的结果。

例1：将十进制数19转换成二进制数。

将十进制数19除以2，记录商和余数，再将得到的商继续除以2，依此类推，直至商为零（如图1.2.7所示）。将余数逆序排列便可得到结果：$(19)_{10}=(10011)_2$，其中下标数字表示进制，也可记为19D=10011B，字母D表示十进制，字母B表示二进制。

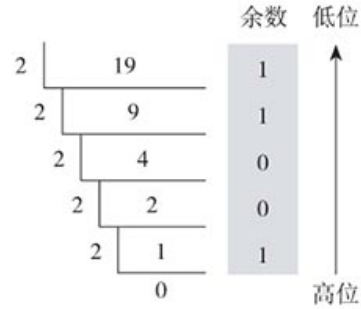


图1.2.7 除2反向取余法

将二进制数转换成十进制数，常用的方法是按权展开求和。

例2：将二进制数10011转换成十进制数。

二进制数中每位数字的权是以2为底的幂，按权展开求和即：

$(10011)_2=1\times2^4+0\times2^3+0\times2^2+1\times2^1+1\times2^0=16+0+0+2+1=(19)_{10}$

在计算机科学中，除了使用二进制外，人们还经常使用八进制和十六进制。不同的进制，常在数的右下角标明基数。基数是可使用数码符号的数目，R进制的基数就是R，遵循“逢R进一”原则。例如，$(27)_8$是八进制数，$(FFCC66)_{16}$是十六进制数。八进制数和十六进制数还可以分别用O和H表示，例如27O、FFCC66H。

八进制的基数为8，逢八进一，1位八进制数可以用3位二进制数表示。

十六进制的基数为16，逢十六进一，1位十六进制数可以用4位二进制数表示。

通常，十进制整数转换成R进制数采用“除R反向取余法”。R进制数转换成十进制数采用“按权展开求和法”。

实践活动

颜色码的数制转换

在计算机中，RGB（red，green，blue）颜色值可以表示为十六进制颜色码。例如，颜色值RGB（64,224,208）可记为#40E0D0，其中表示红色的64、表示绿色的224和表示蓝色的208分别对应十六进制数40H、E0H、D0H。

1. 将#9400D3、#D2B48C表示成相对应的RGB颜色值。
2. 使用计算器程序，验证上述结果。

人民教育出版社

1.2.3 数据编码

通过数据编码，人们可以方便地存储、检索和使用数据。生活中有许多数据编码的实例，例如居民身份证号码、车辆的车牌号、电影院的座位号等。计算机作为数据处理的一种工具，无论处理的是字符、图像、声音，还是其他形式的内容，都需要转换成二进制形式的编码，这样才能够处理。

字符编码

字符是人与计算机交互过程中不可或缺的重要内容，它是多种文字和符号的总称。由于计算机只识别0和1，因此在处理各种字符时，就需要将字符转换为计算机可以识别的二进制数据。

在计算机系统应用的早期，计算机只能处理英文字母、数字和常用的英文标点符号，这时期基础的编码标准是美国信息交换标准码（American Standard Code for Information Interchange，ASCII）。ASCII编码后来被国际标准化组织（International Organization for Standardization，ISO）接受为国际标准ISO 646。

为了扩充ASCII编码，以用于显示更多的语言文字，不同的国家和地区制定了不同的字符集和字符编码标准。但是，由于缺乏统一规划，这些字符编码有可能存在冲突，同一个编码在不同的字符集中可能会代表不同的字符。

1991年，国际标准化组织和Unicode组织联合制定统一码，也就是unicode。unicode致力于将世界上所有的文字和符号都纳入其中。unicode作为一个字符集，并没有规定字符在计算机中存储和传输的编码。为了更好地用计算机处理这些字符，还需要规定具体的字符编码格式。在我国，常用的编码方案有UTF-8、UTF-16、UTF-32及我国的国家标准GB 18030—2005等。

计算机中存储一个“0”或“1”占用1个二进制位（bit）。8个二进制位组成1字节（byte）。ASCII编码用1字节表示英文字母、数字和常见字符。在GB 18030—2005中，大部分常用汉字采用2字节编码。计算机中常见的存储单位与换算关系如表1.2.1所示。

表1.2.1　计算机中常见的存储单位与换算关系

存储单位	换算关系
B（byte），字节	1 B=8 b（bit，位）
KB（kilobyte），千字节	1 KB=1 024 B=2^{10} B
MB（megabyte），兆字节	1 MB=1 024 KB=2^{10} KB=2^{20} B
GB（gigabyte），吉字节	1 GB=1 024 MB=2^{10} MB=2^{30} B
TB（terabyte），太字节	1 TB=1 024 GB=2^{10} GB=2^{40} B

中国地图出版社
人民教育出版社

表1.2.1 续表

存储单位	换算关系
PB（petabyte），拍字节	1 PB=1 024 TB=2^{10} TB=2^{50} B
EB（exabyte），艾字节	1 EB=1 024 PB=2^{10} PB=2^{60} B

阅读拓展

我国常用的字符编码方案

我国在1980年公布了国家标准GB 2312—1980，其中包含了6 763个汉字。后来为了处理更多文字，1995年曾公布过一个过渡性质的行业规范GBK，全称为《汉字内码扩展规范》。GB 18030—2005是我国目前使用的强制性国家标准，采用单字节、双字节和4字节的变长编码，其优点是与GB 2312—1980编码兼容。另一种在互联网中广泛使用的字符编码方案是UTF-8，它采用1～4字节的变长编码方案，与ASCII兼容，英文仍然是单字节，且有编码自纠错等优点，但很多常用汉字采用3字节编码。因此与GB 18030—2005相比，用UTF-8编码后的汉字文本所占的存储空间会大不少。

声音编码

声音是人们用来表达和传递信息的一种重要方式，从物理学上看，声音是一种波。声波的振幅反映了声音响度的强弱，声波的频率反映了声音音调的高低。应用计算机处理声音时，需要将声波的模拟信号转换为数字信号，也就是声音的数字化，编码是其中重要的一步。通常，声音数字化的基本方法是按照一定的时间间隔采集声波的振幅，并将其转换为二进制数序列，即通过采样、量化和编码来实现。

采样是以相等的时间间隔来测得声音模拟信号的模拟量值，对其进行离散化提取，如图1.2.8所示。采样频率指每秒声音被测量的次数，以Hz（赫兹）为单位。例如，高保真音乐采样频率一般为44.1 kHz，即44 100次/s。

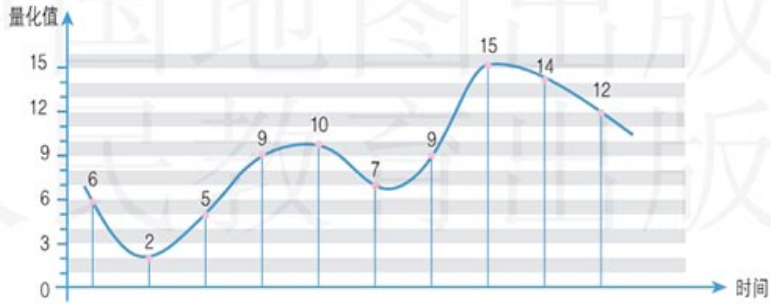


图1.2.8 声音的采样和量化示意图

对于采样所获得的量值，要进行分级量化，就是将采样值变换到最接近的数字值，即用有限个数的数值近似地表示原来连续变化的值。图1.2.8中共有16个量化级别（0～15），因此每个量化值可以用4位二进制数表示。量化位数越多，量化值就会越接近

采样值，音频的精度就越高。如果对一段音乐用256个量化级别进行量化，则每个量化值可以用8位二进制数表示。

通过采样和量化，将一个连续的波形转换成由一系列二进制数表示的数据，形成二进制编码。表1.2.2是对图1.2.8中的量化值用4位二进制数表示的记录。

表1.2.2　音频编码

样本序号	1	2	3	4	5	6	7	8	9	10
量化值（十进制）	6	2	5	9	10	7	9	15	14	12
二进制编码	0110	0010	0101	1001	1010	0111	1001	1111	1110	1100

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

音频所占的存储容量＝采样频率 × 量化位数 × 声道数 × 时长/8

例3：一首时长为100 s的双声道音乐，采样频率为44.1 kHz，量化位数为16，计算该音乐的音频所占的存储容量。

音频所占的存储容量=44 100 ×16×2×100/8=17 640 000 B

17 640 000/1 024/1 024 ≈ 17 MB

阅读拓展

常见的音频文件格式

WAV格式：是声音经过数字化后的文件格式。这种格式记录了经过离散化的声音波形数据。其优点是与原声基本一致，声音质量较高；缺点是文件相对较大。

MP3格式：是一种通过有损压缩存储声音的文件格式。其原理是去除了人耳不敏感的高频部分声音，特点是声音失真较小，文件较小。

AMR格式：主要用于移动设备的音频压缩，压缩比例较大，相对其他的压缩格式质量较差。该格式传输量较小，实时性较强，因此多用于语音通话。在移动网络中，聊天软件常使用AMR格式进行语音压缩。

APE格式：是一种无损压缩格式，以更精炼的记录方式来缩减体积，还原后数据可与原始数据保持一致，但压缩比例通常较小。

图像与视频编码

一幅图像可以看作由许多彩色或各种级别灰度的点组成的，这些点按横纵进行排列，被称为像素，如图1.2.9所示。图像中每个像素的颜色值都用一个或多个二进制位来存储。

通过扫描仪、数字照相机、智能手机和连接在计算机上的摄像头等设备，可以把图像数字化，再经过相关软件处理，就可以把图像转换成相应格式的文件进行保存。BMP格式是图像文件的一种基本格式，由于未经压缩，BMP文件占用的存储空间通常会比较大。

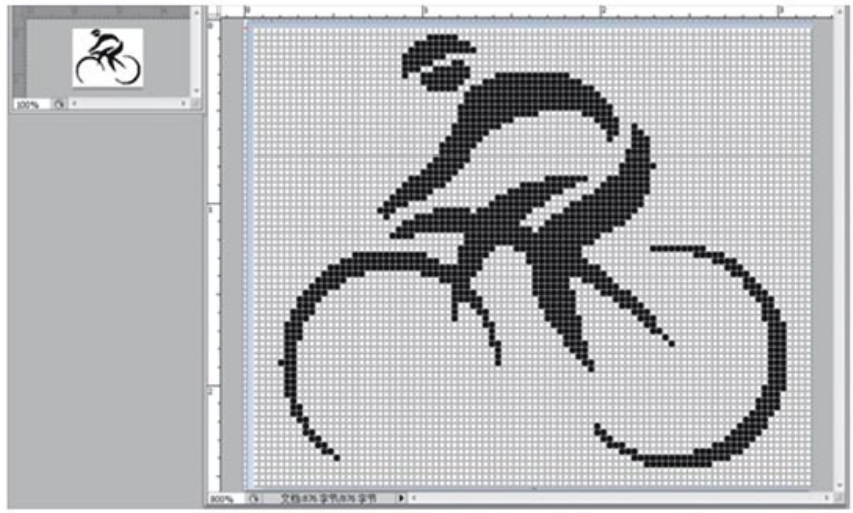


图1.2.9　图像与像素

图1.2.9的左上角是一幅BMP文件原图，右侧是放大8倍的效果图。图中每个网格对应一个点，代表组成图像的像素。图像尺寸可用像素个数来表示，即“水平像素数 × 垂直像素数”。例如，图像尺寸为90×72像素，说明该图像水平方向有90个像素，垂直方向有72个像素。通常，图像所占的存储容量可用该图像所有像素的字节数来表示，其计算公式为：

图像所占的存储容量＝水平像素数 × 垂直像素数 × 颜色深度/8

公式中，图像的颜色深度指图像中每像素的颜色所占的二进制位。例如，图1.2.9只有黑白两种颜色，只需要1个二进制位就能表示图中的颜色。可将黑点用编码1表示，白点用编码0表示。

例4：图像尺寸为90×72像素，颜色深度为1，计算该图像所占的存储容量。

图像所占的存储容量=90×72×1/8 =810 B

常见的彩色图像颜色深度有8 位、16 位、24 位和32 位等，颜色深度为n，能表示的颜色就有2^n种。一般来说，颜色深度越大，图像的色彩就越丰富，图像占用的存储空间也就越大。

例5：图像尺寸为90×72像素，颜色深度为24位，计算该图像所占的存储容量。

图像所占的存储容量=90×72×24/8=19 440 B

19 440/1 024 ≈ 19 KB

高质量的BMP图像数据量很大，占用存储空间较大，所以常用图像压缩技术来减少图像的数据量，提高图像的网络传输速度和效率。

实践活动

常用的图像文件格式

图像文件有多种存储格式，常见的有JPG/JPEG、PNG和BMP等。用户可根据自己的需求，选择合适的图像文件格式。

1. 以体质健康为主题，选用一种“图像编辑软件”创作一幅图像，分别保存为JPG/JPEG、PNG和BMP文件格式，试从图像尺寸、大小和清晰度等方面比较它们的差异。查阅资料，分析这三种图像文件格式的特点。

2. 尝试下载网页中的图像文件，描述它们所采用的文件格式，试说明选用相应文件格式的原因。

视频画面是由以一定的速度连续播放的一组静态图像形成的，这些静态图像被称为帧。当连续的图像变化超过 24 帧/ s时，根据视觉暂留原理，就会形成比较流畅的视频画面了。

当前，通过智能手机、平板计算机和数字摄像机等设备可以直接录制数字视频。此外，通过视频采集卡的输入端口可以采集模拟视频信号，对其进行数字化处理后，就可生成数字视频文件。常见的视频文件格式有AVI、WMV和MP4等。

1.2.4 数据压缩

很多信息，尤其是图像、音频和视频等类型的信息，经过数字化后生成的数据量较大，包含了许多冗余信息。例如，一幅图像中的建筑背景、蓝天和绿地，其中许多像素信息是相同的；一段视频的相邻帧之间，除了运动物体有少许变化外，其他静止的像素信息也都是重复的。因此，在不损失有用信息的前提下，可以按照一定的编码规则对数据进行重新组合，以去除数据冗余。数据压缩的目的，是使文件更少地占用存储空间和缩短传输时间。

常用的压缩方法分为无损压缩和有损压缩。无损压缩指对压缩后的数据进行还原后，得到的数据与压缩前完全相同。例如，AAAAAAAA可以用A08来代替。图1.2.10中显示了这种简单压缩方法的一个示例。从图中可以看出，压缩后的编码在解压后，可以完全复原至压缩前的状态，因此这种压缩方法属于无损压缩。

图1.2.10 简单的无损压缩编码示意图

常用的无损压缩算法有zip压缩算法和7z压缩算法等。压缩后所生成的文件称为压缩包，可能只有原来文件的几分之一，甚至更小。压缩包中的数据可以用压缩软件还原，即恢复到原始状态，这个过程称为解压缩。

相对于无损压缩，有损压缩指在压缩过程中会损失一定的信息，压缩后的数据无法还原成压缩前的样子。例如，音频、图像和视频数字化后存在很多冗余信息，而损失一些冗余信息并不影响视听效果，因此有损压缩被广泛应用于音频、图像和视频文件。常见的有损压缩格式有MP3（音频数据压缩格式）、JPEG（图像数据压缩格式）和MPEG（视频数据压缩格式）等。

技术支持

多媒体素材的采集与加工

1. 文本：可以通过键盘、手写、语音和文字识别软件等方式输入计算机。常用的文本加工操作有设置字体、字号、调整段落的行距、缩进和段间距等。

2. 图像：可以通过数字照相机、平板计算机等拍摄获取，也可以借助扫描仪扫描实物图像来获取。常用的图像加工操作有修改图像尺寸、格式转换、调整亮度和对比度、裁剪、修饰与合成等。

3. 音视频：音视频一般可以通过计算机、智能手机、平板计算机中的录音、摄像功能来录制，还可以使用数字摄像机拍摄来获取。常用的音视频加工操作主要有降噪、剪辑、合成和格式转换等。

此外，多媒体素材还可以通过互联网下载获得。更多的学习支持可访问教学资源平台。

中国地图出版社
人民教育出版社

项目实施

收集体质健康相关素材，初步完成电子作品

一、项目活动

1. 各小组依据绘制的“促进体质健康”的思维导图，从“个人健康评估、运动习惯养成、体质改善措施”等方面获取相应的多媒体素材，并进行必要的加工和处理，保存成合适的文件格式。

2. 小组合作选用一款多媒体集成工具，例如演示文稿或网页制作软件等，形成汇报思路，初步完成多媒体作品。

二、项目检查

1. 依据思维导图，收集制作多媒体作品所需的素材，并进行合理的数字化处理和加工。

2. 小组合作完成多媒体作品的初稿，并进行完善。

练习提升

1. 文件夹中存放了三个文件。一个文件是存放了3万个汉字的纯文本文件（编码方案为GB 18030—2005）；另一个文件是尺寸为1 024×768像素、颜色深度为24位的BMP格式的社团徽章图像；还有一个文件是一首时长2 min的WAV格式的校歌，采样频率是44.1 kHz，量化位数为16位，双声道。如果想把这三个文件通过电子邮件一并发送（假设邮件附件总容量不能超出20 MB），分别计算这三个文件的大小，并且判断它们能否一次性发送成功。（注：忽略因其他因素增加的文件存储容量。）

2. 学校档案室里有很多相纸照片资料，有些照片已泛黄。如何将这些资料存储到计算机中？试说出具体的方法。保存后，若要对图像文件中泛黄的部分进行修复，有什么好的方法？

3. 小明是学生会的宣传干事，学期初为了招募社团新成员，他需要组建一个小团队，策划制作一部社团宣传片。试帮助小明考虑以下两个问题：

（1）制作团队需要哪些成员，分别担任什么角色；

（2）制作宣传片需要哪些素材，如何获取、分类和处理等。

最后将以上方案绘制成思维导图，供小明参考。

中国地图出版社
北京教育出版社

1.3
数据科学与大数据

学习目标

- 认识数据科学，体验大数据对学习、生活和社会发展的影响。
- 了解大数据在社会各领域中的应用，感受大数据在社会变革中的作用。

体验探索

体验大数据

大数据技术的广泛应用，正在悄然改变着我们的生活。有网站利用电子地图的定位功能进行大数据计算与分析，并采用可视化呈现方式，反映人口流动的轨迹与特征。通过它可以了解当前及过往时间段内，全国及各省、市的人口流动情况，以便直观地确定流入人口的来源和流出人口的去向。图 1.3.1 呈现的是某日迁入或迁出北京的人口数据情况。这种利用大数据技术生成的可视化地图可以为交通部门制定政策、提升服务质量提供参考，也可以为百姓生活、企业生产等提供参考，从而实现更大的社会价值。

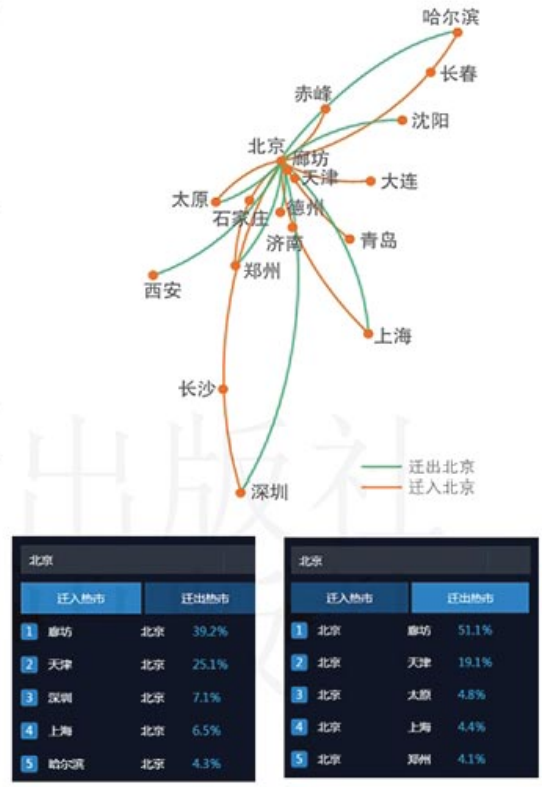


图 1.3.1　可视化“迁徙”地图

选择一座城市，通过网站了解与其有关的人口迁入和迁出的热点城市分别有哪些。

思考：

1. 试分析这些人口流入和流出的热点城市有何关联？从中有何发现？

2. “迁徙”地图中的数据是如何获取的？数据量会有多大？试说明大数据与传统意义上的数据的区别。

1.3.1 数据科学的兴起

互联网、智能终端和行业信息化的快速发展推动了社会数据量的迅速增长。海量数据引发了人们对数据研究思路和方法的转变与飞跃，数据科学成了一个新兴的研究领域。

思考活动

数据与科学发现

在人类文明的历史长河中有很多重要的科学发现。伽利略通过望远镜观测到木星的4颗卫星，为“日心说”找到了证据；居里夫妇在科学实验过程中发现了放射性元素钋和镭；某科学合作组织利用激光干涉引力波天文台探测装置获取的大数据首次发现引力波。

查阅资料，小组讨论上述科学发现在所使用工具、方法上的异同，分析数据在这些发现过程中所起的作用。

人类认识世界的方式多种多样，人们可以通过观察、感知的方式分析自然现象，寻求其中的发展规律；可以通过实验探究的方式研究问题，对问题假设进行验证和判定；如今数据科学为人们提供了认识事物的新方法。例如，科学家通过信息技术工具采集来自太空的相关数据，并及时对它们进行处理，成功探测到引力波的存在。再如，水利部门通过长期收集和分析河流的水流量，就可以预测未来水量的变化。

计算机技术的革新提升了人们处理数据的能力，在计算机的帮助下，人们可以通过多样的计算方式和应用软件快速地处理数据。我国制造的“神威·太湖之光”巨型计算机（图1.3.2），峰值运算速度达到了每秒12.54亿亿次，大大提升了数据处理的速度。

图1.3.2 “神威·太湖之光”巨型计算机

物联网传感器、智能终端的广泛应用，产生和传播着大量的数据，它包括了人们发

布的文本、图像、音频和视频等文件，也包括了人们在上网过程中所产生的浏览日志、点击流量，还包括通过传感器所采集的声、光和温度等数据。计算工具的革新、数据量的持续增长推动了数据科学的发展。

如今，数据科学既包括对数据本身的研究，又包括为社会各个领域的发展提供了管理和研究的新思路和新方法。因此，生活在信息社会，就需要学习一些数据知识，掌握数据处理方法，挖掘数据价值，通过采集、分析和利用数据可以让生活和学习更美好。

阅读拓展 引力波与数据

2017年的诺贝尔物理学奖颁发给了对引力波研究做出贡献的三位科学家，以表彰他们利用激光干涉引力波天文台探测装置（图1.3.3）探测到引力波的存在所做出的重要贡献。引力波的发现是数据技术在科学领域的一个具体应用，捕捉引力波最大的挑战在于其数据的采样频率比较高，超过每秒16 000次，采样信道达上万个，数据量特别大，每年至少能采集到500 TB的数据。通过这些数据，科学家可以更高效地开展研究。随着数据采集与处理技术的发展，数据在科学研究中起着越来越重要的作用。

图1.3.3 探测引力波实验装置

中国教育出版社
人民教育出版社

1.3.2 大数据及其应用

信息技术与经济社会的交汇融合引发了数据的迅猛增长，数据已成为国家基础性战略资源，大数据正日益对全球生产、流通、分配、消费活动及经济运行机制、社会生活方式和国家治理能力产生重要影响。

大数据特征

生活在信息社会，数据伴随着每一个人。从人一出生，个人身份数据就会被采录，之后从上学到就业和生活，各种数据不断产生、传送和接收。信息技术飞速发展的今天，人们在利用社交、教育、电商和金融等平台进行交流、学习、购物和理财等活动的过程中，产生了海量的数据，这些数据正在快速流动、急剧增加，深刻影响着人们的生活、学习和工作。例如，体质状况大数据服务于人们的健康；智能交通大数据有利于人们出行；环境资源大数据助力政府决策；教育教学大数据使我们的学习更加个性化。随着研究的不断深入，通常认为大数据具有巨量性、多样性、迅变性和价值性等特征，如图1.3.4所示。

■ 巨量性

巨量性指数据体量巨大。一旦拥有了事物全部或几乎全部的数据，就在一定程度上提高了数据对事物描述的完整性，这使大数据应用中的全样本分析成为可能。

■ 多样性

多样性指数据种类繁多。大数据包括传统的关系数据，还包含网页、搜索索引、图像、音视频和地理位置等数据。大数据之所以具有如此强的多样性，其主要原因在于：首先，随着互联网的发展，人在上网时，不但会获取数据，还会生成和传播各种数据；其次，各种设备通过网络连成了一个整体。这些设备也在不停地产生、采集和传递各种各样的数据。在互联网上，人类不仅是网络数据的获取者，也成了数据的制造者和传播者。

图1.3.4　大数据的特征

■ 迅变性

迅变性指数据生成速度快，而且要求在短时间内处理完毕。只有快速地从庞杂的数据中获取有价值的信息，才能更好地服务于人，而数据量的快速增长对数据处理速度提出了更高的要求。例如，个性化推荐算法会尽可能完成实时推荐；搜索引擎尽快处理网络中最新信息等。

■ 价值性

价值性不仅指大数据能产生价值，更是指大数据所创造的价值密度相对偏低，即庞

中国地图出版社
北京教育出版社

大的数据量中可能产生价值的只是其中非常小的部分。例如，一段几小时连续不间断的监控视频，可能有使用价值的数据只有事件发生前后的几分钟。

随着人们对大数据认识与应用的深入，大数据也表现出更多样的特征，例如，大数据的真实性越来越为人们所关注，只有高质量和真实的大数据才有助于人们进行预测与抉择。“大数据”与“传统数据”相比不仅在规模上不同，在采集方式，特别是分析方法上也有着明显的差别，包括数据全样本（或近乎全样本）采集、数据实时处理、语义分析和深度学习等。正是由于采用了各种新的数据处理方法和技术，才有可能挖掘出大数据中蕴含的丰富价值。

大数据技术

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

■ 大数据采集技术

大数据采集技术通过物联传感、社交网络等方式获得各种类型的海量数据。例如，随着全国各地平安城市项目的大规模推进，视频采集在智能安防领域发挥着越来越重要的作用，很多机场、车站、码头、办公大楼和住宅小区等场所都安装有视频采集系统（图1.3.5）。

图1.3.5　视频采集摄像头

■ 大数据预处理技术

大数据预处理技术不仅能提高数据质量、降低数据计算的复杂度，还能降低数据规模、提升数据处理的准确性。例如，社交大数据中有些数据涉及用户的隐私，也可能存在一些异常或错误数据，因此要对这些数据进行预处理，这样才能更好地帮助我们进行后期分析以便获得有价值的信息。

■ 大数据存储与管理技术

大数据存储与管理技术能通过相应的数据中心把采集到的数据存储起来，并进行管理和调用。例如，使用云存储和分布式管理技术能方便且快速地处理巨量的数据，实现大数据应用，如图1.3.6所示。

中国地图出版社

图 1.3.6　云存储与大数据应用示意图

■ 大数据分析与挖掘技术

大数据分析与挖掘技术能从规模巨大的数据中，分析并提取出有潜在价值的信息。例如，我国政府部门开发的“精准扶贫大数据云平台”通过对数据的提取分析，展示贫困人口的致贫原因，深度挖掘数据的价值，为扶贫工作提供真实可靠、及时、全面的决策数据，为实现精准扶贫和精准脱贫保驾护航，如图 1.3.7 所示。

图 1.3.7　精准扶贫大数据云平台

■ 大数据可视化与应用技术

大数据可视化与应用技术能将分析或挖掘出来的数据进行直观、形象地呈现，为人们的社会活动提供依据，提高各个领域的运行效率，以发挥出更大的效益。例如，利用手机基站注册数据，不但可以获得居住地的信息，还能了解人口密度。已有数据证明，用手机记录得到的人口数据质量比较高。基于同样原理的景区“大数据客流分析系统”，可以获取游客来源，分析游客密度、景点的人流趋势等，从而为景区智能化管理提供大数据参考，确保景区有序的活动和良好的交通秩序，如图 1.3.8 所示。

图 1.3.8　大数据客流分析系统

实践活动　解读导航地图大数据

利用智能手机上的导航地图可以了解即时路况信息，以查找“捷径”顺畅出行，如图1.3.9所示。

现在的计算机技术已能很方便地将路上车辆的行驶轨迹数据充分利用起来，自动计算出每条道路的实时交通流量。导航地图平台还能根据用户实时上报的交通事件信息，通过大数据平台，将各种信息整合，得出相应的交通路线。

大数据平台综合考虑道路环境、天气情况和节假日等多种因素，基于大数据分析得出每条道路在不同环境或不同时间的路况规律，为交通预测和路径规划提供数据依据。

体验利用导航地图查找出行“捷径”，结合案例，说出其中的大数据从何而来，又是如何处理的。

图1.3.9　导航地图

大数据应用

随着互联网的发展，大数据已经渗透到很多行业，成为重要的生产要素，并通过各行各业的不断创新，逐步为人类创造着更多的价值和财富。

■ 生活服务

目前大数据在生活服务方面的应用较为广泛，通过分析客户的爱好和消费行为及其趋势等，提供更为精准的服务。例如，电商网站会搜集客户的社交数据、浏览器的日志文本及各类传感器采集的数据，通过跟踪分析这些数据，针对客户的个人喜好和消费能力的统计，推荐不同的商品，引导消费，以实现针对客户的个性化服务。随着大数据与人工智能技术的紧密结合，企业以互联网为依托，线上服务、线下体验及现代物流深度融合，这种“新零售”方式正在给人们带来全新的购物体验，如图1.3.10所示。

图1.3.10　新零售

人民教育出版社

■ 智慧城市

大数据可以用来改善城市生活，提升城市管理水平，促进智慧城市的建设。例如，目前很多城市都在进行大数据的分析和应用试点。图 1.3.11 是一个城市数据活化实例的简图。数据活化是把隔离的数据变为整体，恢复数据之间的关联性，以突破数据利用的局限。数据采集的工作可由企业进行，数据经过活化后提供给数据中心。开发者利用这些数据开发新的应用提供给不同需求的个人、企业和政府部门，形成了一个新兴的产业。当智慧城市和产业结合起来，就可以形成健康良性的循环，推动智慧城市的可持续发展。智慧城市的创建已经成为今后市政规划与建设的重要方向。

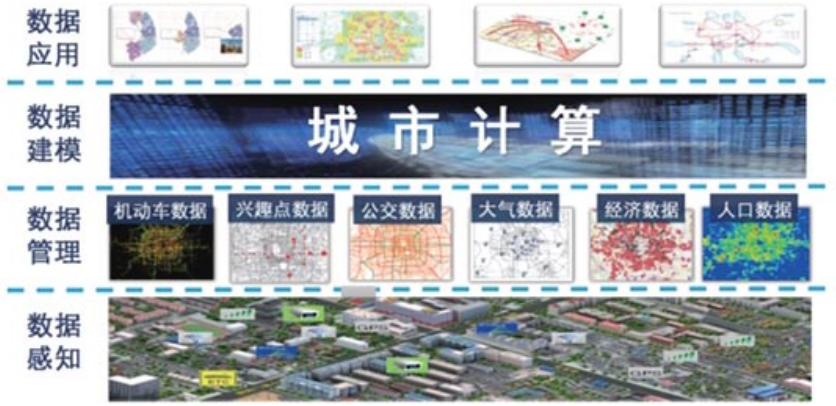


图 1.3.11　城市数据活化实例

■ 医疗健康

大数据在医疗健康方面的应用改变了传统的医疗与健康服务模式，提高了服务的针对性（图 1.3.12）。例如，健康类应用通过可穿戴设备采集数据，进行分析处理后，为患者提供针对性治疗建议，可让医生的诊断更为精确。通过大数据分析，可能以后服药将不再是通常的“成人每日三次，一次一片”的文字说明，而是自动检测并及时提醒。通过计算机科学与生命科学相结合，可以完成超大样本癌症基因的测序分析，能帮助人们解开疾病成因的秘密，辅助科学家攻克医学领域的难题，将对人们今后的健康与医疗环境产生深远的影响。

中国地图出版社
人民教育出版社

图 1.3.12　智能医疗

■ 社区管理

大数据现在已经被广泛应用到社区管理中，为保护居民安全和维护公众利益发挥着越来越重要的作用，如图1.3.13所示。例如，智能社区在辖区内的场所安装烟雾、水压和有害气体等监测传感器，当监控场所存在安全隐患时，报警装置会及时发出警报，数据将直接传输到社区的大数据平台，并第一时间采取相应措施，如通知附近的消防队，对火灾等安全隐患做到早发现、早预警、早扑灭。同时，诸如报警信息等数据还将传送到移动执法终端，方便安全巡查人员随时了解监控场所的动态，以便他们及时监督整改。此外，利用大数据平台或工具，有关部门还可以监控公共场所的活动秩序，协助搜捕犯罪嫌疑人等。总之，大数据的合理应用可以使我们的生活变得更安全。

图1.3.13　智能社区管理系统示意

大数据作为一种重要的资源，它在推动经济发展、完善国家治理、提升政府服务和监管能力等方面具有重要意义。为了更好地发挥大数据的价值，就需要建立“用数据说话、用数据决策、用数据管理、用数据创新”的管理机制，以实现基于数据的科学决策。

大数据深刻改变着人类的思维、学习、生活和生产方式，引领着生活新变化，孕育着发展新思路。大数据涉及每个人、每个企业，乃至整个国家。随着我国大数据战略的实施，基于大数据的智慧生活、智慧企业、智慧城市将会越来越普及。

中国地图出版社
人民教育出版社

实践活动　**了解年度“热词”有多热**

国家语言资源监测语料库的专家团队利用语言信息处理技术，根据词语出现的频次，结合人工提取、筛选后，评选出年度中国十大“热词”，以展示当年中国主要的关注点和语言特点。借助一些大数据平台的指数分析功能，可以研究关键词搜索的趋势，洞察使用者的兴趣和需求变化，监测媒体及舆情动向等。

1. 尝试将你所了解到的“热词”作为关键词，通过某一大数据平台的指数分析功能进行检索，如图1.3.14所示，并进行指数趋势研究，并填写表1.3.1。

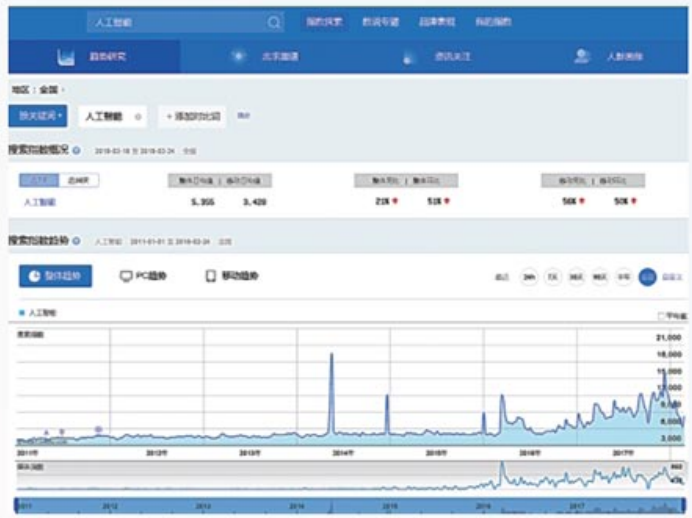


图1.3.14 “热词”检索结果

表1.3.1 大数据应用体验

关键词	搜索指数最高的出现日期	搜索指数的最高值

2. 添加合适的对比词，体验该平台指数分析功能的大数据应用，了解其可视化分析结果。

3. 实践延伸：在互联网上，已经有很多大数据平台，你还知道哪些？如何应用这些平台来支持你的生活和学习？

中国地图出版社
人民教育出版社

项目实施

了解与体质健康相关的大数据应用，完成多媒体作品

一、项目活动

1. 了解与体质健康主题相关的大数据应用，收集并整理加工相关的文字、图像和音视频素材。

2. 基于之前完成的素材准备和技术准备，结合大数据特征，提出基于大数据的体质健康新方法，完善项目的多媒体作品并在全班交流。

二、项目检查

在“体质数据促健康”项目中，小组展示结合大数据特征的项目作品，并完成各组互评。

练习提升

1. 在我国很多城市共享单车成为解决短距离出行不便问题的新选择。试分析这个典型的“互联网+”应用中，大数据的采集是如何实现的，又是如何存储和传输的。

2. 阅读分析。

大数据对人类社会的影响是全方位的。它改变的不仅仅是信息生产力，更是信息生产关系；不仅是知识生产和传播的内容，更是其生产与传播方式。以互联网为代表的媒介技术颠覆了印刷时代的知识生产与传播方式。新媒体技术改写了静态、单向、线性的知识生产格局，改变了自上而下的知识传播模式。在这个时代，我们的知识生产若再固守印刷时代的知识生产理念，沿袭此前的知识生产方式，就会被远远地甩在时代后面。

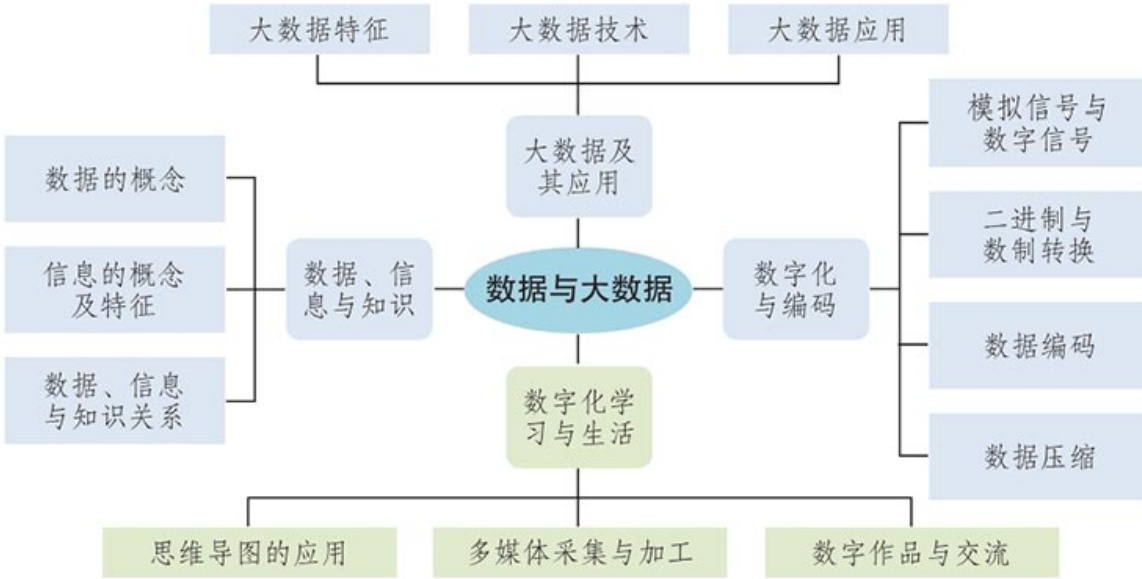
（节选自2013年2月22日《文汇读书周报》，有删改）

在小组研讨的基础上，试分析大数据与互联网的关系，讨论如何更好地应用大数据。

中国教育出版社
人民教育出版社

总结评价

1. 下图展示了本章的核心概念与关键能力，请同学们对照图中的内容进行总结。



2. 根据自己的掌握情况填写下表。

学习内容	掌握程度
数据与信息的概念	□不了解 □了解 □理解
数据、信息和知识的关系	□不了解 □了解 □理解
数字化及其作用	□不了解 □了解 □理解
二进制及数制转换	□不了解 □了解 □理解
字符编码	□不了解 □了解 □理解
声音编码	□不了解 □了解 □理解
图像与视频编码	□不了解 □了解 □理解
数据压缩	□不了解 □了解 □理解
数据科学	□不了解 □了解 □理解
大数据特征	□不了解 □了解 □理解
大数据技术	□不了解 □了解 □理解
大数据应用	□不了解 □了解 □理解

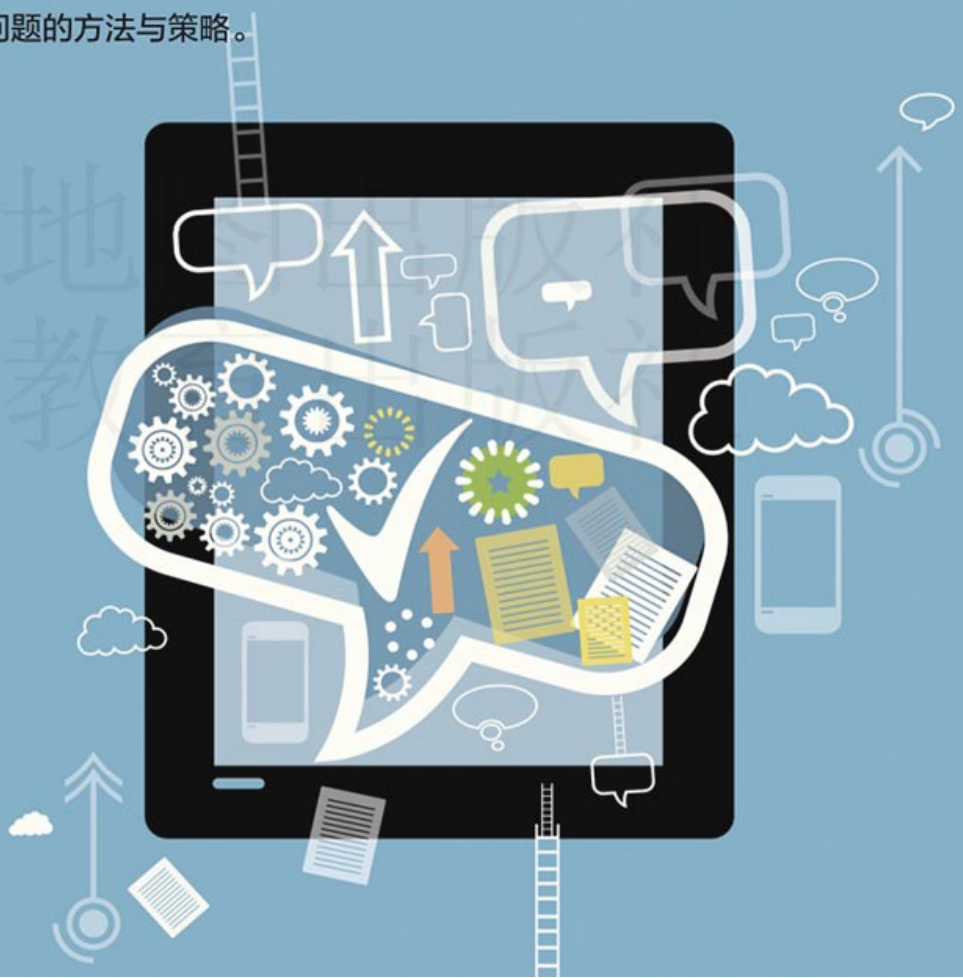
中国地图出版社
人民教育出版社

第 2 章
算法与程序实现

2017 年 10 月的一天，在杭州市中心路段开展了一场救护车施救演练。在全程近 7 km 的路段中，救护车获准优先通行 21 次，平均行驶速度达 36 km/h，相较于该路段常规通行时间，省时近 900 s。优先通行“抢”出的这十几分钟，可能就是挽回病患生命的“黄金”时间。在这个现实版的生死时速演练案例中，批准救护车优先通行的正是杭州“城市大脑”工程的“交通模块”。截至 2018 年初，杭州的“城市大脑”已接管了市区内的主要路口信号灯，通过各类数据感知交通态势，进而优化信号灯配时，使车辆通行速度提升近 15%。这就是通过算法与程序设计对大数据进行综合应用的奇妙之处。

算法与程序浸润在我们生活的各个方面。计算机与移动终端已成为生活中不可或缺的工具，它们之所以能够帮助人们处理各种复杂的事情，主要借助于其中功能各异的程序。

在本章的学习中，我们将通过“编程控灯利出行”项目活动，学习如何利用编程的方式实现算法并解决问题，从而发展计算思维，掌握利用计算思维解决问题的方法与策略。



中国地图出版社
人民教育出版社

2 主题学习项目：编程控灯利出行

项目目标

智能交通利用先进的信息技术手段，对交通进行综合管理，为公众提供智能、便捷的交通出行服务。交通信号灯作为交通控制、疏导的重要手段，是智能交通研究的一个重要内容。本章通过“编程控灯利出行”项目，探讨智能交通环境下交通信号灯的控制策略和编程实现方式。

1. 围绕项目问题，进行需求分析和功能分解，设计问题求解方案。

2. 掌握Python语言基本知识与编程方法，能编写程序实现简单问题的求解。

3. 通过解决项目问题，掌握用计算机解决问题的基本方法，并能迁移解决其他问题。

项目准备

为完成项目，需做以下准备。

- 寻找一名同伴，在项目实施中通过互助合作方式完成任务。
- 调查学校或居住地周边道路的通行状况，汇总问题，为开发项目程序做好资源准备。
- 为“编程控灯利出行”项目实施搭建Python语言环境。

项目过程

在学习本章内容的同时开展项目活动。为了保证本项目的顺利完成，要在以下各阶段检查项目的进度。

设计方案 ①

完成项目问题的需求分析，设计解决问题的方案。

P45

描述算法 ②

设计项目问题求解的算法，用流程图对算法进行描述。

P53

编写程序 ③

按照算法编写程序，实现“自助式人行过街红绿灯”的功能。

P68

调试运行 ④

调试运行项目程序，结合实际需求完善功能。

P81

项目总结

完成“编程控灯利出行”项目，并进行展示交流与评价。掌握编程解决问题的基本方法，理解程序设计中的算法思想，并能迁移解决其他问题，发展计算思维。

2.1
解决问题的一般过程和用计算机解决问题

学习目标

- 经历用计算机解决问题的基本过程，并能应用于实际问题的解决。
- 体验编程解决问题的基本方法，提高学习算法与程序的兴趣，发展计算思维。

体验探索

两种不同的叫车方式

小明同学准备和父母一起乘飞机外出旅游，为了尽快赶到机场，父母决定到小区门口招手叫辆出租车，而小明则建议通过叫车软件网络约车。

思考：

1. 针对上面的情境，比较“招手叫车”和“网络约车”两种方式的不同，完成表2.1.1。

表2.1.1　两种叫车方式的分析

叫车方法	叫车地点	叫车步骤与过程	便捷程度
招手叫车			
网络约车			

2. 从问题解决的过程和效率等方面，分析这两种方法的异同。

相同点：_____。

不同点：_____。

中国地图出版社
人民教育出版社

2.1.1 解决问题的一般过程

生活中人们经常会面对各种各样的“问题”，诸如探索自然、了解社会、认识自我等。在解决这些问题的过程中，人们逐渐把握规律，认识真理。解决问题就是要在已知条件和可能的结果之间寻求具体的途径与方法，并应用它们实现目标。通常，解决问题的过程需要经历一系列的思维和实践活动。

思考活动

交通指挥问题

在无信号灯或者信号灯不能正常工作的路口，经常会有交通警察在现场指挥交通，如图2.1.1所示。

图2.1.1 交通警察指挥交通

思考：

描述交通警察为保证道路畅通进行交通疏导的一般过程和方法。

人们解决问题的过程通常包括分析问题、寻找解决问题的途径与方法，以及解决问题并验证结果等阶段，如图2.1.2所示。

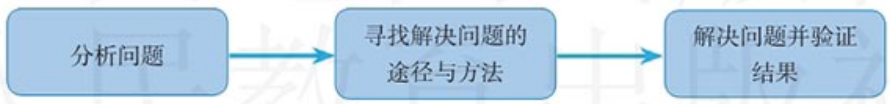


图2.1.2 解决问题的主要阶段

1. 分析问题

分析问题主要包括调查、收集与问题相关的资料，明确问题的目标、条件和所涉及的相关知识与规则等，找出已知与未知之间的联系。

例如，交通警察在指挥交通时，要对道路的交通流量情况进行观察，及时了解各条道路的路况信息，明确问题求解的目标：维护交通秩序，提高通行效率，保障交通安全。

2. 寻找解决问题的途径与方法

明确要解决的问题后，需要将待解决的问题与已知条件或已有的规则相关联，设计问题求解的方案，包括具体的途径和方法等。

例如，交通警察依照工作规范，通过对道路通行状况的分析，结合已有经验，“想”出问题解决的具体过程和方法，指挥来往车辆和行人有序通行。

3. 解决问题并验证结果

最后，要依据确定的求解方法进行问题解决，考查所得到的解答，通过检验答案和评估效果，实现问题的最终解决。在方案执行过程中，如果发现结果没有达到预期，就需要调整解决方案。

例如，交通警察指挥交通时，使用手势信号进行指挥和疏导，协调与控制道路交通。他（她）还要根据道路实际状况，适时调整指挥方案，以便提高通行效率。

2.1.2 用计算机解决问题的过程

计算机具有运算速度快、计算精确度高、逻辑运算能力强、存储容量大和自动化程度高等特点。因此，利用计算机解决问题，能在一定程度上提高问题解决的效率。例如，交通信号控制系统就是利用计算机程序控制红绿灯时长变化来指挥交通的，可以在节省人力的前提下，科学分配道路上车辆、行人的通行权，使之有秩序地顺利通行。随着物联网、大数据和人工智能等技术的发展，交通信号灯控制系统还可以根据联网数据精准地自动调节红绿灯的时长，实现多路口信号灯的智能联动控制。

思考活动

用计算机程序控制交通信号灯“红灯变绿灯”

交通信号灯是国际通用的道路交通信号设施。我们可以用计算机程序模拟控制红绿灯的变化，例如编程实现一个“灯”由红色变为绿色的效果。

思考：

结合前面交通警察指挥交通的过程，分析用计算机编程解决该问题的过程，思考二者有何不同。

中国地图出版社
人民教育出版社

用计算机解决问题即让计算机按照程序执行指令。人们可以使用已有的程序，也可以根据需要自己编写程序。利用计算机编程的方式进行问题求解时，通常需要经历分析问题、设计方案、编程调试和解决问题等环节，如图2.1.3所示。根据问题求解的需要，中间过程可能要反复修正，直至问题得到有效解决。

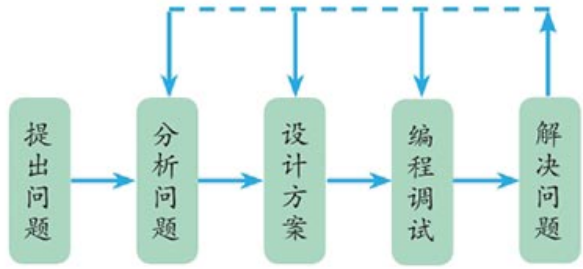


图2.1.3　用计算机解决问题的一般过程

分析问题

用计算机编程解决问题时，需要对问题进行分析，明确问题的目标和条件等，需要把问题进行抽象，通过建模的方式，界定和描述问题。问题描述的方式并不唯一，有的问题可用数学模型描述，有的问题可用文字、表格或图形等形式描述。

例如，“红灯变绿灯”问题中，信号灯的初始状态为红灯变亮（禁止通行），并且红灯要持续显示一段时间，然后绿灯变亮（准许通行），如表2.1.2所示。

表2.1.2　分析“红灯变绿灯”问题

初始状态	中间状态	终止状态
红灯变亮	红灯持续显示	绿灯变亮

设计方案

计算机编程解决问题的设计方案一般包括以下两个环节：

首先，根据需求分析，将问题按照求解过程分解为若干相对独立的功能，每个功能完成一个特定的任务。如果分解的某些功能仍然比较复杂，还可再进行细分，如图2.1.4所示。设计方案的过程中，人们进行问题分解的角度并不唯一，不同的角度所得的功能分解图也会不尽相同。

中国地图出版社
人民教育出版社

图2.1.4　问题功能分解示意图

例如，在设计“红灯变绿灯”问题的解决方案时，根据前面的问题分析，需要考虑

信号灯由初始状态到终止状态的变化过程，因此该问题可以分解为如图2.1.5所示的三个功能。

图2.1.5 “红灯变绿灯”功能分解示意图

然后，针对分解的各个特定功能分别进行详细的操作步骤设计，给出问题求解的具体过程和方法，即进行算法设计。

编程调试

1. 编写程序

编写程序就是利用计算机程序设计语言描述算法，实现问题求解的过程。在计算机领域，程序是一组可执行的计算机指令。程序设计语言是编写程序所使用的计算机语言。随着计算机技术的发展，程序设计语言经历了从机器语言、汇编语言到高级语言的发展历程。

高级语言的出现为计算机的应用开辟了广阔的前景。目前，很多人都在使用高级语言编写程序。高级语言有很多种，虽然它们的特点各不相同，但编程解决问题的过程同一些基本的程序设计规则和方法却是相似的。因此在学习某种语言后，应该具有将其中共性的思想和方法迁移到其他语言环境中进行问题求解的能力。

本章使用Python语言编写程序解决具体问题。Python语言具有简洁、明确等特点，拥有丰富的扩展功能模块（有些时候一个模块不能独立地解决某一类问题，软件工程师就把这些相关的模块组织到一起形成“库”），可以支持复杂的数据处理，在数据分析和人工智能等领域都有广泛的应用。

2. 调试程序

编写完成的程序，需要进行调试运行。一般，调试程序不仅要发现错误，分析其产生的原因，并进行改正；还要对运行的结果进行分析和验证，判断其是否正确和完整。

在编写程序时，为了便于阅读和理解程序，通常会使用注释。注释语句是对程序代码的解释和说明，在运行程序时不会被执行。在编写程序时，要养成添加注释的良好习惯，以提高程序的可读性。

Python 语言中的单行注释是以“#”开头的，或以独立行的形式单独存在，或写在程序语句后。例如：

```
# 初始化红绿灯——独立行注释
import turtle # 导入可以绘制图形的Python内置模块turtle——语句后的注释
```

实践活动

Python程序的修改与运行

运行教学资源平台提供的Python程序“红灯变绿灯.py”，阅读程序代码，参照注释语句分析程序功能。程序代码如下：

```
import turtle                    # 导入可以绘制图形的Python内置turtle模块
import time                      # 导入与时间处理有关的Python内置time模块
# 初始化红绿灯
light = turtle.Turtle()
light.hideturtle()
light.screen.delay(0)            # 禁用绘制过程的动画显示
# 红灯变亮
light.color("red", "red")
light.begin_fill()
light.circle(20)
light.end_fill()
# 红灯持续显示
time.sleep(15)
# 绿灯变亮
light.color("green", "green")
light.begin_fill()
light.circle(20)
light.end_fill()
```

1. 根据图2.1.5所示的“红灯变绿灯”功能分解，分析每个功能所对应的程序代码。

2. 如果将红灯持续显示时长设为30 s，应该如何修改程序？

阅读拓展

程序设计语言

在用计算机编程解决问题的过程中，确定问题求解的算法后，还需要通过程序设计语言把算法转换成计算机可运行的程序。程序设计语言经历了从低级语言到高级语言的发展，其中，机器语言、汇编语言和高级语言各有不同的特点，如表2.1.3所示。

表2.1.3　机器语言、汇编语言与高级语言的特点比较

程序设计语言	特点
机器语言	是用二进制数表示的机器码集合，可以被计算机直接执行
汇编语言	是面向机器的程序设计语言，用“助记符”来代替机器码进行编程
高级语言	是独立于机器的面向过程或对象的程序设计语言，语法结构接近人类自然语言，便于编写和理解，其表达方式接近于被描述的问题，通用性和可移植性好

中国地图出版社
人民教育出版社

项目实施

确定项目问题的解决方案

一、项目活动

随着经济的发展与社会的进步，智慧交通成为智慧城市建设的重要组成部分。“编程控灯利出行”项目通过程序设计，研究改善和优化交通信号灯的控制策略，以提高交通出行效率，实现交通管理的人性化和智能化。

1. 明确项目问题。很多城市为了便于行人出行，会在道路上配置一种“自助式人行过街红绿灯”，如图2.1.6所示。行人按下过街按钮，当信号灯变为绿灯后，便可通过。这种装置实现了行人和交通信号灯之间的交互，体现了人性化的交通管理，可在一定程度上提高道路通行率。本章项目研究的主要任务就是通过编程，模拟实现这种“自助式人行过街红绿灯”的效果，体验程序设计的一般方法。

图2.1.6　自助式人行过街红绿灯

围绕“自助式人行过街红绿灯”问题，进行调查分析。查阅资料，了解交通信号灯配置方面的相关知识，确定项目问题求解模型并填写表2.1.4。

表2.1.4　“自助式人行过街红绿灯”问题分析表

信号灯初始状态	单击按钮后，信号灯的状态变化	信号灯终止状态

中国地图出版社
人民教育出版社

2. 分解项目功能。根据对问题的分析，画出“自助式人行过街红绿灯”功能分解图，制订解决问题的方案。

二、项目检查

分析项目活动的主要问题，梳理项目功能，制订解决问题的方案。

练习提升

1. 随着移动设备的普及，越来越多的人使用移动应用程序解决问题。与同伴分享一款自己感兴趣的移动应用程序，分析它能解决什么类型的问题，是如何解决问题的，完成表2.1.5。

表2.1.5　例举移动应用程序解决问题

移动应用程序	可以解决的问题	问题解决的过程

2. 查找Python和C语言（或其他高级语言）的资料，用思维导图描述它们的特点。

3. 设计拓展活动项目，实现更多个性化问题的求解。结合生活或学习需求，提出想要解决的问题，如编写24点游戏程序、制作古诗查询助手和开发选课系统等。

拓展活动项目名称：

① 提出问题：

② 分析问题：

③ 初步规划解决方案，进行功能分解：

中国地图出版社
人民教育出版社

2.2
算法的概念及描述

学习目标

- 描述算法的特征，理解算法在问题解决中的作用。
- 选用恰当的描述方法和控制结构表示简单算法，增强用算法解决问题的意识。

体验探索

规划乘车路线

小明同学所在城市的地铁线路局部图，如图2.2.1所示。他计划从A站出发去B站附近的图书馆学习。假设地铁各线路每两站间行车用时相等，记为t_1；换乘地铁的用时也都相等，记为t_2。

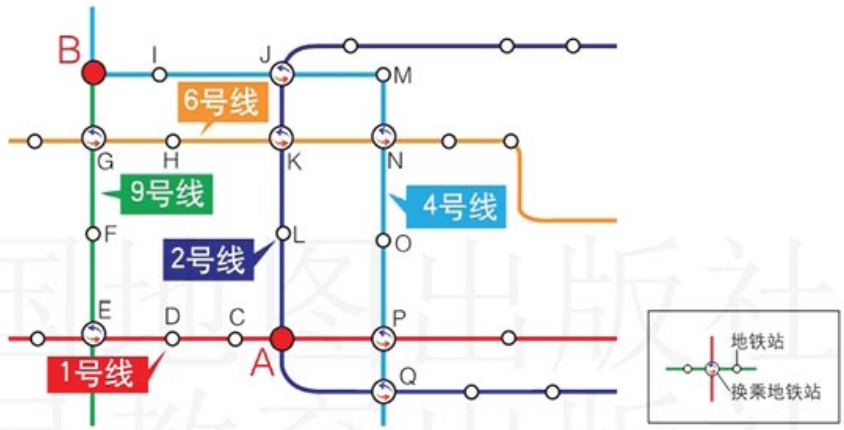


图2.2.1　某城市地铁线路局部图

思考：

1. 列举出由A站出发到达B站的所有换乘次数最少的乘车路线。

2. 如果小明同学希望尽快到达B站，试为他推荐一条最佳乘车路线，并说明理由。

中国地图出版社
人民教育出版社

2.2.1 认识算法

在生活和学习中，人们经常会运用到算法（algorithm）知识，只是很少意识到。从广义上讲，算法是为解决一类特定问题而采取的确定的、有限的步骤。它描述出某类问题求解的方法和过程，在整个问题解决过程中起着重要的作用。例如，表2.2.1所示的求两个正整数最大公约数的问题，应用的是我国古代《九章算术》中记载的“更相减损术”，其方法为“可半者半之，不可半者，副置分母、子之数，以少减多，更相减损，求其等也。以等数约之。”

表2.2.1　用“更相减损术”求两个正整数的最大公约数

两种情形	步骤
“不可半者”（如153和119）	直接辗转相减，直至减数和差相等： ① 153−119=34 ② 119−34=85 ③ 85−34=51 ④ 51−34=17 ⑤ 34−17=17 所以，153与119的最大公约数为17
“可半者”（如24和16）	第一步，先将24和16用2反复约简，直至不都是偶数，约数最后为8（用2约简了三次） 第二步，分别将两个约简后的数辗转相减，直至减数和差相等： ① 3−2=1 ② 2−1=1 第三步，求差和约数的乘积：1×8=8 所以，24与16的最大公约数为8

中国地图出版社
人民教育出版社

在计算机领域，算法作为一个精心设计的运算序列，描述了计算机如何将输入转化为输出的过程。

算法一般具有如下特征：

■ 有输入。一个算法一般要求有0个或多个输入，以描述运算对象的初始情况。例如，在用“更相减损术”求最大公约数的过程中，输入的是两个正整数。

■ 有输出。一个算法可以有一个或多个输出，以反映对输入数据加工后的结果。例如，用“更相减损术”求得的最大公约数就是该算法的输出。

■ 有穷性。算法的有穷性指算法必须能在执行有限个步骤之后终止，也就是算法步骤不能是无限的。例如，在用“更相减损术”求最大公约数的过程中，要求重复辗转相减直

至减数和差相等。实际上，对于任意给定的两个正整数，无论原始值有多大，这个过程的执行次数都是有穷的。

■ 可行性。算法中的每一步操作都是可以执行的，或者都可以分解成计算机可执行的基本操作。

■ 确定性。算法的每个步骤都具有确定的含义，没有歧义。模糊不清、模棱两可或带有二义性的描述都会影响算法的确定性。

阅读拓展

为什么要学习算法

我们在生活和学习中会经常遇到算法。例如，楼宇电梯通常按照一定的算法响应用户请求，合理停靠到相应的楼层；铁路网络订票系统按照一定的算法设置订票模式，高效服务用户。

智能时代，算法已经广泛应用于各领域。许多专家通过分析行业的运行规律，界定问题，有针对性地建立模型、设计算法，并应用信息技术实现模型，从而创造出新的产品，创生出新的产业。例如，通过设计算法，在网络环境下实现“互联网＋单车”，催生出“共享单车”这个新产业，解决短距离出行不便问题。

中学生学习一些算法知识，了解算法的基本设计方法，可以深入理解身边数字化工具的特征，能够利用算法思想解决实际问题，提高学习和生活效率，更好地融入信息社会。

2.2.2 描述算法

描述算法就是将解决问题的步骤，用一种可理解的形式表示出来。常用的描述算法的方法有自然语言、流程图和伪代码等。

中国地图出版社
人民教育出版社

思考活动

描述“红灯变绿灯”问题的算法

完善本章第一节思考活动“红灯变绿灯”功能，在交通信号灯下方增加一个15 s的“倒计时器”，提示过往的行人和车辆。

思考：

如何将“倒计时15 s”的算法描述出来？

用自然语言描述算法

自然语言指人们日常所用的语言。用自然语言描述算法就是使用人们能读懂的简短语句对算法的步骤进行描述。其中，“倒计时15 s”算法可用自然语言描述为：

步骤1：将计数器t设为15；

步骤2：如果t大于或等于1，执行步骤3，否则倒计时结束；

步骤3：输出t，并保持显示1 s，然后清除显示；

步骤4：将t的值减1，跳转至步骤2。

用自然语言描述算法易于理解，它既可以描述生活中的算法，也可以描述在计算机中执行的算法。但是，自然语言的描述方法存在容易产生二义性的缺点，有可能干扰后续的编程实现。

用流程图描述算法

流程图是一种常用的表示算法的图形化工具。用流程图描述的算法直观易读，问题解决的步骤清晰简洁，算法结构表达明确，很适合初学算法的人员使用。流程图中常用的符号及其功能如表2.2.2所示。

表 2.2.2　流程图常用符号及其功能

流程图符号	名称	功能
	开始/结束框	表示算法的开始或结束
	输入/输出框	表示输入或输出数据
	处理框	框中指出要处理的内容，此框有1个入口和1个出口
	判断框	用于表示条件判断及产生分支的情况，判断框有4个顶点，通常上面的顶点表示入口，视需要用另外3个顶点来表示出口
	流程线	用于控制流程方向
	连接点	用于连接因页面写不下而断开的流程线

中国地图出版社
人民教育出版社

绘制流程图的方法很多，可以手工绘制流程图，也可以用软件制作，如使用文本编辑软件中的“流程图”对象绘制，或使用专门的流程图绘制软件，还可以到在线绘制流程图网站进行制作。

算法有顺序结构、选择结构和循环结构三种基本控制结构，可用流程图描述，如图2.2.2所示。

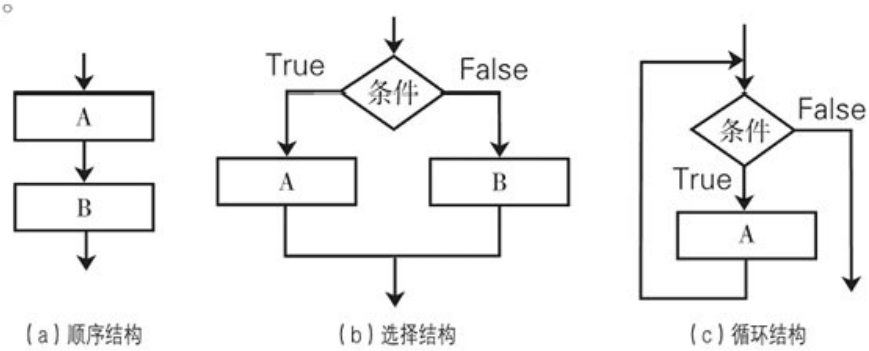


图2.2.2　用流程图描述算法的三种基本控制结构

■ 顺序结构。每一个步骤按先后次序被执行，即执行处理A，然后执行处理B，如图2.2.2(a)所示。

■ 选择结构，又称分支结构。根据条件的成立与否，选择执行不同的分支处理，如图2.2.2(b)所示。当条件成立时(用True表示)，执行处理A；当条件不成立时(用False表示)，执行处理B。

■ 循环结构。当条件成立时，反复执行处理A，一旦条件不成立就立即结束循环，如图2.2.2(c)所示。

在实际问题解决中，经常会综合使用这三种结构。例如，“倒计时15 s”的算法可用图2.2.3所示的流程图描述。

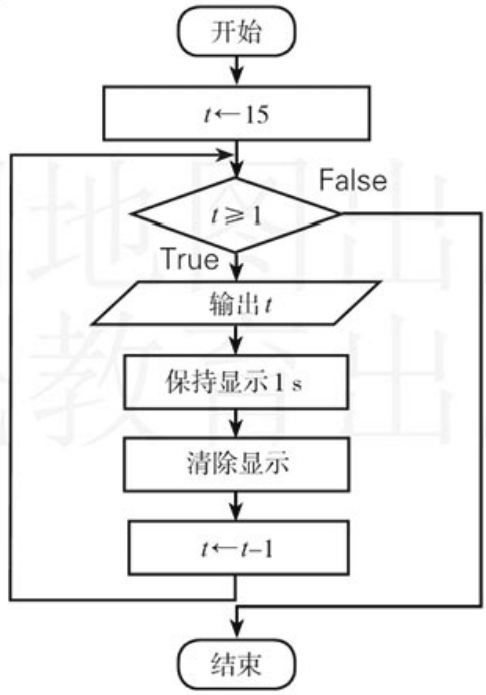


图2.2.3　“倒计时15 s”算法流程图

中国地图出版社
人民教育出版社

用伪代码描述算法

用伪代码描述算法就是采用一种类似于程序设计语言的代码来表示算法。伪代码没有固定的、严格的语法规则，只要定义合理，没有矛盾即可。例如，“倒计时15 s”的算法用伪代码可以描述为：

```
t ← 15
while t≥1
    output t
    sleep 1s
    clear
    t ← t-1
end while
```

用伪代码描述算法回避了程序设计语言严格的书写格式，保持了语言叙述准确、无二义性的优点，结构性强，比较容易书写和理解。

算法描述了问题求解的具体步骤，决定着问题解决的过程。解决同一问题可能会有不同的算法，不同算法求解的过程或有不同。例如，在“体验探索”的“规划乘车路线”过程中有多种解决方法，由于换乘的次数和乘坐的站点数不同，采用不同的方法乘车，总用时并不相同，但要想“尽快到达B站点”，其中用时最少的一组应该是该问题求解的最佳方法。

实践活动

找出质量较轻的零件

已知有10个一模一样的零件，其中9个零件的质量相同，只有1个质量略轻，不符合规格要求。现在有一台天平，请设计算法找出该零件。

1. 如果采用一一比较的方法，逐一称重对比，最多需要比较多少次才能找出这个质量较轻的零件？试着描述该算法，想一想还有哪些方法可以解决该问题？

2. 如果有n个零件（$n>10$），要找出其中质量较轻的一个，以上方法是否仍然可用？试分析n=10 000时，这些算法在问题解决效率上的不同。

中国地图出版社
人民教育出版社

阅读拓展

算法与数字生活

算法与数字生活息息相关，如使用机器人清理房间，通过智能运动手环进行科学健身，输入用户口令安全登录账户等，算法都在其中起着重要的作用。借助高效的算法和优化的程序，我们可以更加快捷方便地解决越来越多的问题，提高生活、工作和学习效率。

用数字化工具解决问题离不开程序的控制，而程序的核心是算法设计。算法描述了解决问题的步骤，程序则是用某种程序设计语言将设计好的算法进行具体实现，从而实现自动化的问题求解。

例如，空调具有的定时关机功能，就可以通过倒计时算法来实现。在设定关机时间后，系统将进行倒数计数，当计数结束时就会发出关机指令，关闭空调。

再如，某城市公交车票价2元，乘客可以刷卡乘车。刷卡时，若公交卡余额不足2元，提示“请投币”；若余额大于或等于2元但小于10元，提示“余额即将不足”；若余额大于或等于10元，提示“欢迎乘车”。用流程图描述该功能实现的算法如图2.2.4所示。

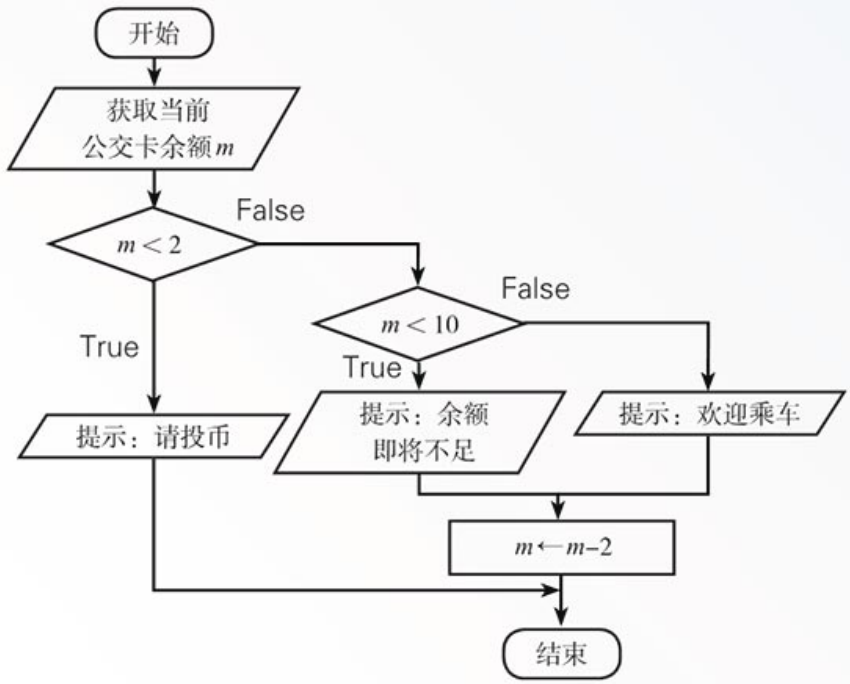


图2.2.4　某城市刷卡乘坐公交车的算法流程图

中国地图出版社
人民教育出版社

项目实施

描述项目问题的算法设计

一、项目活动

结合“编程控灯利出行”项目，在完成“自助式人行过街红绿灯”的问题分析和功能分解之后，对各功能进行算法设计，并用流程图的方式描述出来。

二、项目检查

设计项目任务实现的算法，并用流程图描述出对应的算法。

练习提升

1. 小明早上从起床到出门上学需要洗漱（5 min）、烧水（10 min）、准备早餐（10 min）、吃早餐（15 min）和听新闻（10 min），试帮他统筹安排各项事务，设计一个合理的算法步骤。

2. 某地有两种不同类型的出租车，其计费标准分别为：

- 甲车3 km起步，价格为10元，3 km以上（含3 km）为2元/ km；
- 乙车3 km起步，价格为8元，3 km以上（含3 km）为2.2元/ km。

设计算法，在不同里程时给出最优资费的用车选择。选用一种算法描述方法对该算法进行描述。

3. 欧几里得算法又名辗转相除法，其算法可用图2.2.5所示的流程图描述（“%”为取模运算符，可返回除法的余数），认真阅读流程图，分析该算法实现的功能。

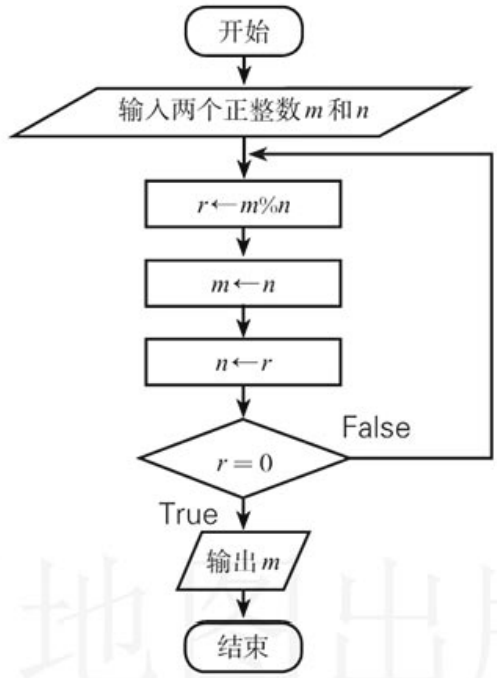


图2.2.5　欧几里得算法流程图

中国地图出版社
人民教育出版社

2.3
程序设计基本知识

学习目标

- 掌握Python语言的基本知识，体验通过编程解决问题的基本过程，感受计算机编程的魅力。
- 运用顺序、选择和循环三种结构编写程序，提高利用计算机解决问题的能力。

体验探索

编程绘制交通信号灯

我们可以编写Python程序绘制各种图形。例如，绘制一个圆形红色信号灯的程序代码如下：

```
import turtle
light = turtle.Turtle()
light.hideturtle()
light.color("red", "red")
light.begin_fill()
light.circle(20)
light.end_fill()
```

参考教学资源平台中提供的“turtle模块帮助.docx”文件，理解各语句的功能，编程绘制交通信号灯。

思考：

1. 如果要绘制如图2.3.1所示的“红绿”信号灯，应该如何编写程序代码？分析其中两盏信号灯的位置坐标有何关系，描述出对应的算法。

2. 如果要绘制如图2.3.2所示的“红黄绿”信号灯，又应该如何修改程序代码？分析其中哪些语句是可以重复使用的。

中国地图出版社
人民教育出版社

图2.3.1　红绿信号灯

图2.3.2　红黄绿信号灯

2.3.1 数据类型

在编写程序解决问题的过程中，为了更好地处理各种数据，程序设计语言提供了多种数据类型。Python语言中常见的数据类型有整型、浮点型、字符串型、布尔型和列表等。表2.3.1中列出了Python语言的部分常用数据类型。

表2.3.1 Python语言的部分常用数据类型

数据类型	类型标识符	类型说明及示例
整型	int	Python中的整数理论上可以任意大（不存在计算机位数不够的情况），如3、−99、988 776 635 547 347 784
浮点型	float	可以使用浮点型表示小数，如3.141 59、0.008 1、2.318E+6（表示2.318×10^6）
字符串型	str	字符串通常是用一对单引号（''）或双引号（""）括起来的一串字符，如：'中国'、"China"、"300451"、"\t"（制表符）
布尔型	bool	布尔类型通常用于逻辑判断，只有两个值：True和False，分别代表真和假
列表	list	列表是用来存放一组数据的序列。列表中存放的元素可以是各种类型的数据，它们被放置在一对中括号“[ ]”中，以逗号分隔，如：[1, 2, 3, 'A ', 5]、["《红楼梦》", "《水浒传》", "《三国演义》", "《西游记》"]

实践活动

确定数据类型

根据表2.3.2中描述的现实事物或现象，列举出具体的数据，说明其对应的数据类型，完成表格填写。

表2.3.2 对应的数据类型

现实事物或现象	列举的数据	数据类型
倒计时的数字	15	整型
列出信号灯的三种颜色名称		
按钮是否可用		

在“编程控灯利出行”项目中，编程完成“自助式人行过街红绿灯”的任务还需要使用哪些数据？分析其对应的数据类型。

中国地图出版社
人民教育出版社

在程序设计中，可以通过强制类型转换操作，把某个数据从一种数据类型转换成为另一种数据类型。Python语言提供了可以进行强制类型转换的函数，如表2.3.3给出的部分数据类型转换函数。

表2.3.3　Python语言中部分数据类型转换函数

数据类型转换函数	功能说明
float(*x*)	将*x*转换为一个浮点型数据
int(*x*)	将*x*转换为一个整型数据
str(*x*)	将*x*转换为字符串型数据

2.3.2　常量、变量及表达式

在程序设计中，确定了数据的类型，也就确定了对该数据可以进行的处理操作。程序设计语言利用运算符来操作数据，运算符将数据连接成表达式，单个的常量或变量也是表达式。

常量与变量

1. 常量

常量指程序运行过程中，其值不能改变的量，例如1024、"red"等。

2. 变量

变量指程序运行过程中，其值可以改变的量。变量是内存中存放数据的存储单元，用变量名来表示。Python语言中，变量命名需要遵循以下规则：

■ 变量名可以包括字母、汉字、数字和下划线（_），但不能以数字开头。例如，name2是合法变量名，而2name是非法变量名。

■ 变量名区分大小写，例如，myName、MyName、myname和Myname是不同的变量。

为了便于理解，变量名应具有一定含义，即“见名知意”。需要注意的是，不能用系统关键字作为变量名。系统关键字是程序设计语言中保留下来的用于特殊用途的标识符，每个关键字都有特殊的含义。在Python语言中，常用的系统关键字有for、in、if、continue、break、else、and、not和or等。

Python语言中的变量是在首次赋值时创建的。赋值语句是最基本的程序语句，其格式为：

变量名＝表达式

例如，i = 1、colors = ["red", "yellow", "green"]和light = turtle.Turtle( )等都是赋值语句。需要注意的是，Python语言中的变量必须在使用前被赋值，不能使用未被赋值的变量。在使用过程中，变量还能够被重新赋值。

技术支持　访问Python语言列表中元素的值

Python 语言中的变量可以被赋值为列表类型的值，如colors = ["red", "yellow", "green"]。若要访问列表中某个元素的值，可用变量名加索引值的方式。例如，列表变量colors中的3个元素为：

colors[0]	colors[1]	colors[2]
"red"	"yellow"	"green"

colors[*n*]中的*n*表示列表的索引值（*n*=0,1,2），其中，colors [0] 的值为"red"，colors [1] 的值为"yellow"，colors[2]的值为"green"。

运算符与表达式

运算符标明了对操作数（参与运算符计算的数据）所进行的运算，它与操作数连接而成的有意义的式子被称为表达式。表达式在进行运算处理后得到的结果称为表达式的值。例如，2+3就是一个表达式，其中的操作数是2和3，而运算符是“+”，表达式的值是5。表达式的值可以赋给变量，也可以作为程序语句中的判断条件。

不同类型的数据可以进行不同的运算，如整型数据有取模运算（%），而字符串没有。不同类型的数据使用同一种运算符运算时，常会产生不同的结果。例如，“+”除了可以进行数值加法运算，还可以对字符串、列表等数据进行连接运算，如"China"+"521"的运算结果是"China521"，[11, "张明"] + [90.5]的运算结果是[11, "张明", 90.5]。

常用的运算符有算术运算符、关系运算符和逻辑运算符等。

1. 算术运算符

算术运算符主要进行算术运算，运算的结果为整型或浮点型。常见的算术运算符有加（+）、减（-）、乘（*）、除（/）、取模（%）、幂（**）和整除（//）等。例如，6%4的结果为2，2**3的结果为8，9//2的结果为4。

2. 关系运算符

关系运算符主要用于比较两个值的大小，运算的结果为布尔值True或False。Python语言中常见的关系运算符如表2.3.4所示。

表2.3.4　Python语言中常见的关系运算符

关系运算符	描述	用法	举例
==	等于	*a*==*b*	2==3的结果为False
>	大于	*a*>*b*	2>3的结果为False
<	小于	*a*<*b*	2<3的结果为True
>=	大于或等于	*a*>=*b*	2>=3的结果为False
<=	小于或等于	*a*<=*b*	2<=3的结果为True
!=	不等于	*a*!=*b*	2!=3的结果为True

中国教育出版社
人民教育出版社

3. 逻辑运算符

逻辑运算符可以对关系式或布尔值进行逻辑运算，运算结果仍为布尔值True或False。Python语言中常见的逻辑运算符如表2.3.5所示。

表2.3.5　Python语言中常见的逻辑运算符

逻辑运算符	描述	用法	举例
not	非	not *a*	not (2==3) 的结果为 True
and	与	*a* and *b*	只有当两个操作数都为 True 时，其结果才为 True 如：(2<3) and (2>1) 的结果为 True
or	或	*a* or *b*	只要有一个操作数为 True，其结果就为 True 如：(2<3) or (1>2) 的结果为 True

如果两个或多个运算符出现在同一个表达式中，则要按照优先级确定运算顺序。优先级高的运算符先运算，优先级相同的从左向右依次运算。需要注意的是：

- 当表达式中出现“()”时，它的运算级别最高，应先运算“()”内的表达式；
- 运算符的优先级为：算术运算符 > 关系运算符 > 逻辑运算符；
- 在同类运算符中也要注意不同的优先级。例如，逻辑运算符not、and 和 or的优先级为：not > and > or。

实践活动　写出Python语言表达式

根据问题需求，写出对应的表达式，完成表2.3.6。

表2.3.6　问题描述对应的Python语言表达式

问题描述	Python语言表达式
数学中的代数式：$\frac{2a}{b+c}$	
将两个字符串"加油！"和"中国"连成一个字符串	
表示110米栏成绩为16.24 ～ 18.62 s（包括16.24 s和18.62 s）	

中国地图出版社
人民教育出版社

2.3.3　语句与程序结构

一般来说，程序是由语句组成的，执行程序就是按特定的次序执行程序中的语句。程序语句包括赋值语句、控制语句和输入输出语句等。其中，控制语句通过对程序流程

的控制，决定了程序执行的路径，也决定了程序的结构。程序设计的三种基本结构包括顺序结构、选择结构和循环结构。

顺序结构

顺序结构指程序的执行按语句的排列顺序从上到下依次执行，直至结束。例如，“体验探索”中绘制红灯或绿灯的过程就是顺序结构。

程序通常包括输入数据、处理数据和输出结果三部分。Python 语言主要用函数 input( ) 实现数据输入，用函数 print( ) 实现数据输出。

■ 输入函数 input( )

函数 input( ) 主要用来接收键盘的输入，返回值为字符串型数据。通常，在输入时可以给出提示信息，例如：x = input(" 请输入一个正整数：")。

■ 输出函数 print( )

函数 print( ) 主要用于在屏幕上输出一个或多个输出项的值，多个输出项中间用逗号隔开，例如：print(x, " 是奇数 ")。

实践活动

编程计算体重指数

根据键盘输入的身高和体重值，编写程序计算体重指数，以 bmi.py 为名保存文件。

1. 分析问题，找出已知条件和求解的目标，并确定二者之间的关系。

2. 用流程图描述该问题求解的算法，并编程实现。想一想能否调整这些语句的书写顺序，为什么？

技术支持

函数

程序中的函数是预先定义好的一段逻辑相对独立、功能相对单一的代码块。在需要时可以直接调用函数，而不必重写代码，这样提高了代码的可读性和可重用性。函数调用的一般格式为：

函数名(参数)

在编写程序时，既可以调用系统提供的内置函数，也可以根据功能需求自定义函数并调用。

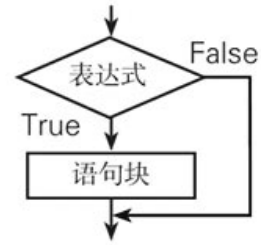
中国地图出版社
人民教育出版社

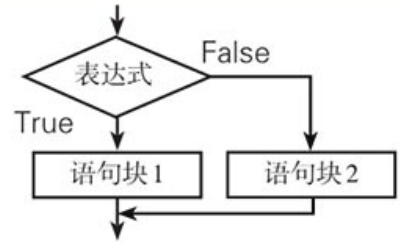
选择结构

事实上，很多问题的解决并不是简单地依次顺序执行，有时需要根据条件有选择地处理。例如，在“自助式人行过街红绿灯”问题中，红绿灯在“按钮被按下”时，就会被启用，即红灯持续一段时间后变绿灯。因此，程序设计时就需要利用选择结构，对“按钮是否被按下”这一状态进行判断。

在程序的选择结构中，某些语句会受到条件的制约，根据条件成立与否有选择地执行。选择结构利用条件语句，通过判断表达式的值（True或False），决定程序执行的分支。单分支结构和双分支结构是常见的两种选择结构，在Python语言中是由if语句实现的，基本格式见表2.3.7，注意其中的冒号不能省略。

表2.3.7　Python语言单分支、双分支结构格式

分支类型	基本格式	对应算法流程图
单分支	`if 表达式:` `　　语句块`	

双分支	`if 表达式:`
	`　　语句块1`
	`else:`
	`　　语句块2`
		

例如，判断一个正整数奇偶性的方法通常是：如果一个正整数x能被2整除，那么这个数为偶数，否则为奇数。其流程图如图2.3.3所示。

中国地图出版社
人民教育出版社

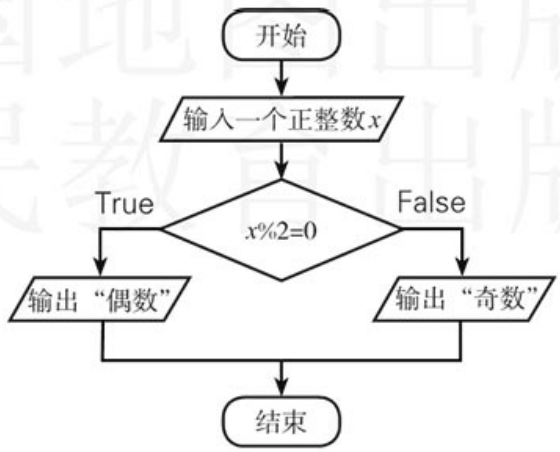


图2.3.3　判断正整数奇偶性的算法流程图

程序示例如下：

```
x = int(input("请输入一个正整数："))
if x % 2 == 0:
    print(x, "是偶数。")
else:
    print(x, "是奇数。")
```

Python语言用缩进的方式标识语句块，表示语句之间的层次关系。同一级语句块中的语句具有相同的缩进量。

选择结构除了单分支和双分支结构，还有多分支结构。生活中很多问题也是需要依据多种不同情况分别进行处理的。

例如，某城市轨道交通实行分段计程票价，起步票价为乘坐5站以内（含5站）单人票价2元；乘坐5站以上10站以下（含10站），单人票价3元；乘坐10站以上16站以下（含16站），单人票价4元；乘坐16站以上单人票价5元。那么该问题就有4种分支情况，如图2.3.4所示。

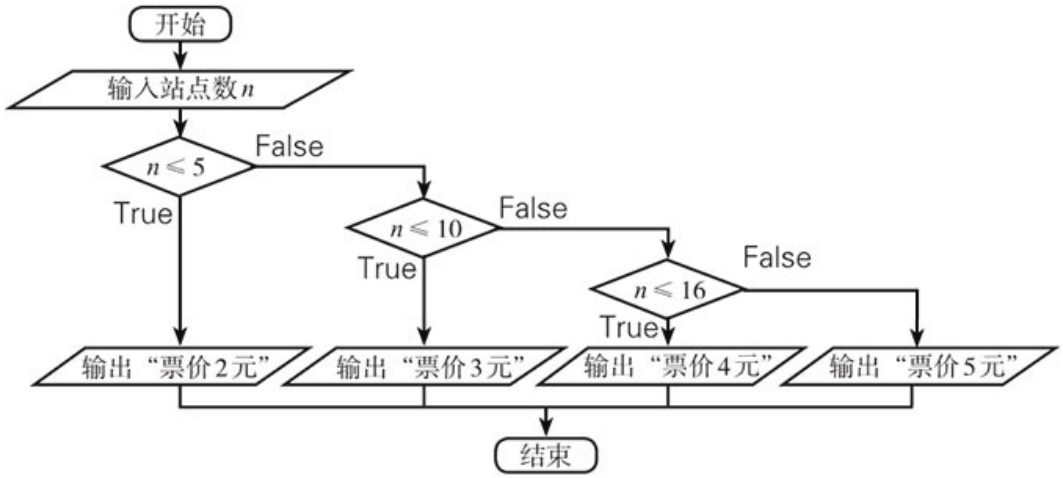


图2.3.4　分段计程票价的多分支算法流程图

Python语言中，可以在if语句中配合使用elif语句来实现多分支结构，多分支结构的一般格式及“分段计程票价”程序示例如表2.3.8所示。

表2.3.8　Python语言中多分支结构的一般格式及程序示例

Python语言中多分支结构的一般格式	“分段计程票价”程序示例

`if 表达式1:`
`　　语句块1`
`elif 表达式2:`
`　　语句块2`
`　　⋮`
`elif 表达式n:`
`　　语句块n`
`else:`
`　　语句块n+1`
	`n = int(input("请输入乘坐的站点数:"))`
	`if n <= 5:`
	`　　print("票价2元")`
	`elif n <= 10:`
	`　　print("票价3元")`
	`elif n <= 16:`
	`　　print("票价4元")`
	`else:`
	`　　print("票价5元")`

中国铁道出版社
人民教育出版社

在程序设计中，选择结构是处理分支问题的基本手段。在用选择结构进行问题求解时，需要先确定分支情况，然后再用恰当的if语句编写程序。

实践活动

评定体重指数等级

依据《国家学生体质健康标准》体重指数单项评分表，如果高一年级男生体重指数值在16.5 ~ 23.2之内，则属于“正常”等级。

完善程序（bmi.py）功能，实现输入某个高一男生的身高和体重值，就能输出他的体重指数等级。

1. 用变量bmi表示体重指数值，写出判定bmi是否在“正常”等级范围内的表达式。

2. 结合表2.3.9所示的高一学生体重指数评分表，实现对某个男生体重指数等级的完整评定，即根据输入的身高和体重值，显示对应的等级：“正常”“低体重”“超重”或“肥胖”。用流程图描述该问题求解的算法，并编程实现。

3. 如果继续完善程序，使之能对任意高一学生（男生或女生）进行体重指数等级的判定，需要考虑哪些条件？试用流程图将求解的算法描述出来。

表2.3.9　高一学生体重指数评分表

性别	等级	体重指数/(kg · m^{-2})	单项得分
男生	正常	16.5 ~ 23.2	100
	低体重	≤16.4	80
	超重	23.3 ~ 26.3	
	肥胖	≥26.4	60
女生	正常	16.5 ~ 22.7	100
	低体重	≤16.4	80
	超重	22.8 ~ 25.2	
	肥胖	≥25.3	60

中国地图出版社
人民教育出版社

循环结构

在解决问题过程中，还会遇到需要重复执行某些操作，才能得到最终结果的情况。例如，2.2节中设计的“倒计时15 s”的过程，就是一个反复执行15次“输出t的值—保持

显示1 s—清除显示—$t \leftarrow t-1$”的过程。此时，就需要利用循环结构实现这种重复性的操作。

在循环结构中，一些语句在某一条件成立时，被重复执行，直到条件不成立时，才结束重复执行。这个条件称为循环控制条件，被重复执行的语句块称为循环体。循环结构的流程图如图2.3.5所示。

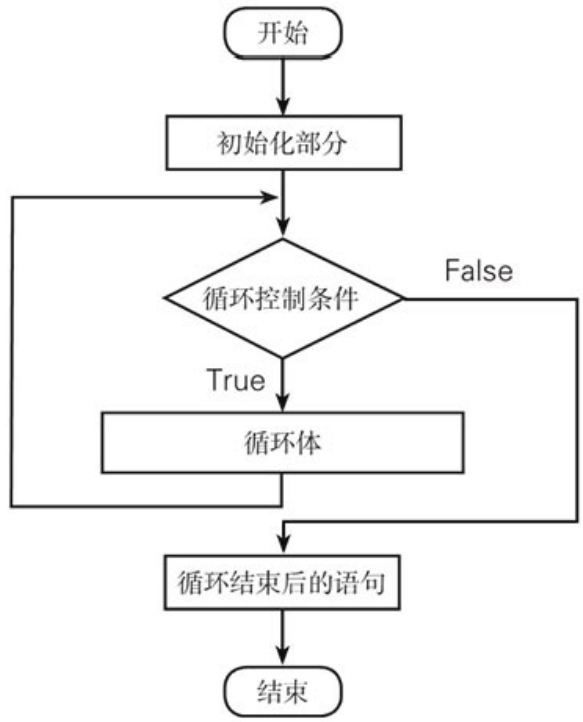


图2.3.5　循环结构流程图

Python语言提供了两种基本的循环结构语句：while语句和for语句。

1. while语句

while语句的基本格式为：

```
while 表达式:
    语句块
```

while语句中的表达式是循环控制条件，其值一般为布尔值（True或False），当表达式的值为True时，执行循环体内的语句，否则就退出循环，执行下一条语句。需要注意的是，while表达式后面的冒号不能省略。

例如，“倒计时15 s”功能实现的程序示例如下：

```
import turtle
import time
text = turtle.Turtle()
text.hideturtle()
i = 15
while i >= 1:
    text.write(i, font=("黑体", 14, "normal"))
    time.sleep(1)
    text.clear()
    i = i - 1
```

技术支持

Python语言中的模块

Python语言中的模块是一个程序文件（扩展名是.py），它将程序代码和数据封装起来。编写程序时，要先导入所需的模块，然后就可以在程序中调用该模块中的函数等。

导入模块一般写在程序的开始位置，其基本格式为：

```
import 模块名
```

Python语言内置了许多模块，常用的有：

1. turtle模块。turtle模块可用于绘制图形。例如，“体验探索”中绘制的红灯和绿灯，就是通过导入turtle模块，利用其中的函数circle()绘制圆形。

2. time模块。time模块主要用于处理时间，其中提供了一系列的控制时间的函数，如sleep(n)实现让程序在该语句位置休眠时间为n（单位为s），即延时n后再执行下一条语句。

3. math模块。math模块可用于数学运算，如平方根、指数、对数和三角函数的运算等。

由于问题解决的复杂性和独特性，当内置模块不能满足需要时，也可使用第三方模块或者自行开发新模块。在使用第三方模块之前，需要通过包管理工具pip进行模块安装。

2. for语句

for语句的基本格式为：

```
for 循环变量 in 序列:
    语句块
```

for语句每次从序列中取出一个元素赋值给循环变量（循环变量初值即为序列中的第一个元素值），当依次访问完序列中所有元素后，循环结束。需要注意的是，for … in后面的冒号不能省略。

例如，在“倒计时15 s”的问题中，使用for语句编写的程序示例如下：

```
import turtle
import time
text = turtle.Turtle()
text.hideturtle()
for i in range(15, 0, -1):
    text.write(i, font=("黑体", 14, "normal"))
    time.sleep(1)
    text.clear()
```

中国地图出版社
人民教育出版社

技术支持

for语句中的序列

1. 在Python 语言中，可以使用函数range() 返回一个等差整数序列，格式为：range(起始值，终值，步长)。需要注意的是，生成的这个整数序列不包括终值，并且起始值和步长（序列中相邻两个元素之间的差）可以缺省，缺省时分别为0和1，如表2.3.10 所示。

表2.3.10　函数range()返回的等差整数序列举例

函数range()举例	功能说明
range(1, 5, 2)	生成从1至5（不包含5），间隔2的连续递增的等差整数序列1，3
range(1, 5)	生成从1至5（不包含5），间隔1的连续递增的等差整数序列1，2，3，4
range(5)	生成从0至5（不包含5），间隔1的连续递增的等差整数序列0，1，2，3，4
range(5, 1, −1)	生成从5至1（不包含1），间隔1的连续递减的等差整数序列5，4，3，2

2. 在for语句中还可以直接使用列表等序列。例如：

```
names = ["张明", "王丽", "李刚", "唐华"]
for name in names:
    print(name)
```

最后执行结果为：

```
张明
王丽
李刚
唐华
```

在for语句中，循环次数由序列中元素的个数决定，常用于确定循环次数的问题解决。对于不能确定次数的问题求解，需要使用 while 语句。实际上，while 语句既可以实现不确定次数的循环，也可以实现确定次数的循环。

3. 循环嵌套

一个循环结构内可以包含另一个循环，这样的结构称为循环嵌套，也称多重循环。常用的循环嵌套是二重循环，外层循环称为外循环，内层循环称为内循环。内循环是外循环的循环体。循环嵌套的执行过程是要首先执行外层循环，外循环每执行一次，内循环则需执行一个完整的循环。例如，输出显示“九九乘法表”，程序示例如下：

```
# 外循环
for i in range(1, 10):
    # 内循环
    for j in range(1, i + 1):
        # 通过end = ""设置函数print()不输出换行符
        print(i, " * ", j, " = ", i * j, "\t", end = "")
    # 设置换行操作(函数print()不输出任何字符，只进行换行)
    print()
```

4. break语句和continue语句

在循环体中，可以使用break语句和continue语句控制程序的流程。break语句用于终止某个循环，使程序跳到循环体以外的第一个可执行语句；continue语句用于终止本次循环进入下一次循环，即跳过循环体内还未执行的语句，进入下一次循环的判定。

break和continue语句的用法比较，可以参考表2.3.11所示的程序示例及其运行结果。

表2.3.11　break和continue程序语句用法及运行结果比较表

程序示例	运行结果

`num = 0`
`for i in range(1, 5):`
`    print("进入第", i, "次循环, i = ", i)`
`    if i == 3:`
`        break`
`    num = num + 1`
`    print("num = ", num)`
	进入第 1 次循环, i = 1
	num = 1
	进入第 2 次循环, i = 2
	num = 2
	进入第 3 次循环, i = 3

`num = 0`
`for i in range(1, 5):`
`    print("进入第", i, "次循环, i = ", i)`
`    if i == 3:`
`        continue`
`    num = num + 1`
`    print("num = ", num)`
	进入第 1 次循环, i = 1
	num = 1
	进入第 2 次循环, i = 2
	num = 2
	进入第 3 次循环, i = 3
	进入第 4 次循环, i = 4
	num = 3

实践活动

编写完整的体重指数“计算器”

我们已经编程实现对高一年级学生体重指数的计算和等级判定，但该程序运行一次只能处理一名学生的信息。试继续完善程序（bmi.py）功能，编写一个完整的体重指数计算小程序，实现对多名学生进行体重指数的等级判断。

1. 要实现多次判定高一学生体重指数等级的程序功能，应如何设置循环条件？该循环的次数是否固定？如何结束该循环？

2. 画出该问题求解的流程图，并编程实现。

程序设计过程中许多问题不可能只通过一种程序结构解决，往往需要将顺序结构、选择结构和循环结构综合起来使用。

阅读拓展　从结构化程序设计到面向对象的程序设计

20世纪60年代，随着计算机应用的普及，人们为了应对规模庞大的应用软件需求，开始研究如何更高效地开发程序。此时，“自顶而下、逐步求精”的结构化模块设计思想开始广泛应用。这种思想要求围绕过程，把解决问题的步骤写出来，程序一步一步执行就能解决问题，因此称为“面向过程”的程序设计。

20世纪80年代，面向对象的程序设计思想开始流行，并逐渐发展成为主流的程序开发技术。C++、Java等都是面向对象的程序设计语言。面向对象程序设计以对象为中心，具有数据与方法的封装、对象继承等特征，代码可重用性比较高。

项目实施

编写程序模拟实现“自助式人行过街红绿灯”

一、项目活动

结合本节程序设计知识的学习，尝试编写程序模拟实现“自助式人行过街红绿灯”的完整功能，参考效果如表2.3.12所示。

表2.3.12　“自助式人行过街红绿灯”各种灯控状态变化表

初始状态	单击按钮后，信号灯状态变化		终止状态
红灯亮，禁止通行	红灯开始倒计时	绿灯亮，行人通行	红灯亮，禁止通行

1. 结合前面对“自助式人行过街红绿灯”项目问题的分析，思考：

（1）如果行人甲按下按钮，在绿灯倒计时结束时刚过完马路，而行人乙此时恰好按下“按钮”请求过马路，那么，程序是否应该响应行人乙的请求？为什么？想一想实现该功能的程序语句。

（2）按下“按钮”后，信号灯的状态变化及对应的倒计时显示等效果可以利用哪些语句实现？

2. 用包管理工具pip安装sinomaps模块，并导入到项目程序中，参考教学资源平台提供的“sinomaps模块中函数和属性功能表.docx”文件，完成“自助式人行过街红绿灯”问题的编程实现。

二、项目检查

编写程序，初步实现“自助式人行过街红绿灯”项目任务的功能。

练习提升

1. 分别指出下面程序中循环体执行的次数，写出运行结果，分析程序功能。

```
i = 100
while i > 1:
    if i % 13 == 0:
        break
    i = i - 1
print(i)
```

```
for i in range(100, 1, -1):
    if i % 13 != 0:
        continue
    print(i)
```

2. 设计算法编写程序实现如下功能：

已知一元二次方程$ax^2+bx+c=0$，从键盘输入a、b和c的值，获取该一元二次方程解的情况。

3. 设计算法编写程序，求任意两个正整数的最大公约数，直至用户输入“q”或“Q”才退出程序。

4. 编写程序和计算机来一场有趣的猜数游戏。要求：由计算机随机生成一个100以内的正整数，用户有5次猜数机会，如果猜中提示“猜中了！”并退出程序；否则，提示“数偏大”或“数偏小”，然后继续猜数。如果5次均未猜中，屏幕显示计算机随机生成的整数，并换行给出提示“5次没有猜中，很遗憾，游戏结束”，退出程序。

提示：内置模块random可以生成一个指定范围内的随机数。参考语句为：

```
m = random.randint(1, 100)    # 生成一个指定范围为[1,100]的随机整数
```

中国地图出版社
人民教育出版社

2.4
常见算法的程序实现

学习目标

- 理解解析算法和枚举算法，根据需要选用这两种算法，编程实现简单问题求解。
- 认识问题解决中不同算法的效率，完成项目程序的调试与运行。

体验探索

绿灯时长的最优设置

通常，行人步行速度约为4.4 km/h，观察到信号灯变化后的反应时间约为2 s。要保证过街行人能走过长为20 m的人行横道（如图2.4.1所示），人行过街绿灯时长至少需要设置为多少？

图2.4.1 “自助式人行过街红绿灯”路口

思考：

1. 写出求解绿灯最短时长的计算公式：_____。

2. 结合实际道路情况，思考在设置人行过街绿灯时长时需要考虑哪些因素，试着给出绿灯时长的最优设置模型。

中国教育出版社
人民教育出版社

2.4.1 基于解析算法的问题解决

解析算法指通过找出解决问题的前提条件与结果之间关系的表达式，并计算表达式来实现问题的求解。

许多问题可以通过分析，抽象成数学模型，借助解析式，用已知条件为变量赋值进行求解。例如，在“体验探索”中求解行人过马路最短绿灯时长时，可以应用行程问题相关公式，先计算行人过马路的时间$t_{过马路}=\frac{S_{马路}}{v_{步行}}$，然后建立数学模型$t_{绿灯长}=t_{过马路}+t_{人反应}$，得到行人过街绿灯最短时长公式$t_{绿灯长}=\frac{S_{马路}}{v_{步行}}+t_{人反应}$，最后只要将已知条件代入公式即可完成该问题求解。

例1：自由落体运动问题。

问题：从离地500 m的高处自由落下一个小球，求从开始落下的时刻起，小球在最后1 s内的位移（重力加速度g以9.8 m/s^{2}计）。

（1）分析问题

已知条件：小球离地高度500 m，重力加速度g为9.8 m/s^2；

求解目标：小球在下落最后1 s内的位移；

已知与未知的关系：可用自由落体运动位移与时间关系的公式$h=\frac{1}{2}gt^2$，求解出下落时间t，以及最后1 s内小球的位移。

（2）设计算法

在该问题中，要计算最后1 s内小球的位移，首先要求出小球的落地时间t，由$h=\frac{1}{2}gt^2$可以得出落地时间$t=\sqrt{\frac{2h}{g}}$；然后计算前$(t-1)$ s小球下落的高度hx；最后求出总高度h（500 m）与hx的差hh，即为最后1 s内小球的位移。

该问题求解算法的流程图如图2.4.2所示。

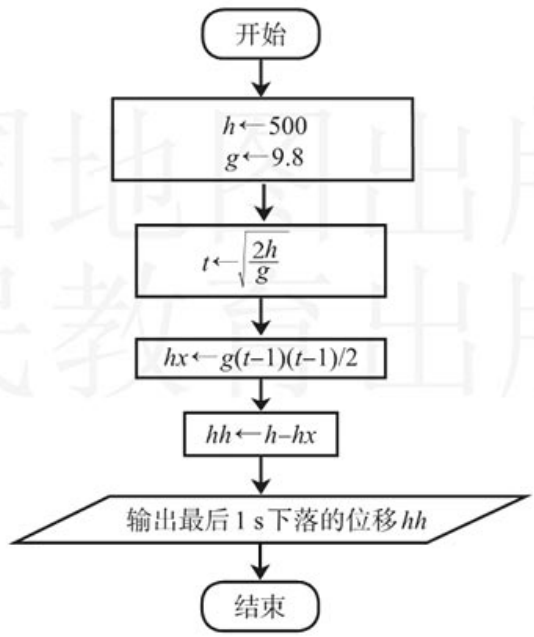


图2.4.2 自由落体运动问题求解流程图

中国地图出版社
人民教育出版社

（3）编程实现与调试

根据算法设计进行编程实现，程序示例如下：

```
import math
h = 500
g = 9.8
t = math.sqrt(2 * h / g)
hx = g * (t - 1) * (t - 1) / 2
hh = h - hx
print("小球最后1 s下落的位移是: ", hh, "m")
```

（4）保存文件，调试运行程序

实践活动　编写程序研究某山地的气温分布

某地区为了开发山区农业，需要了解山地的气候变化。现已知该地山区海拔每升高100 m，气温下降约0.5 ℃，山地最高海拔为1 500 m，山脚下的年平均气温为22 ℃（假设山脚海拔为0 m）。

1. 依据气温随海拔升高而变化的规律，写出计算该山地不同海拔高度的气温的解析式，并编程实现。

2. 某种植物适宜生长在气温为18 ~ 20 ℃的山区，如果要分析这种植物应被种植在该山地多高的地区为宜，需要如何修改算法？试编程实现。

编程解决问题时，要善于综合运用物理、数学、化学等学科的知识和方法来分析问题，寻找问题中各要素之间的关系，用形式化的符号和公式表达要素之间的关系，得出解决问题所需的表达式。例如，我们用自由落体运动的相关公式来求解小球位移的问题，用一元二次方程求根公式求解方程等。生活中许多问题也可以用解析算法解决，例如，通过银行贷款买房后每月还贷金额的计算、打折商品价格的计算及求体重指数等。

阅读拓展

利用“割圆术”求 π 的近似值

“割圆术”是求圆周率的一种算法。我国古代数学家刘徽发现当圆内接正多边形的边数无限增加时，其面积可无限逼近圆面积，即所谓“割之弥细，所失弥少，割之又割，以至于不可割，则与圆合体，而无所失矣”。同时，这些圆内接正多边形每边外有一个余径（线段AB），如图2.4.3所示。用边长乘以余径，加到正多边形的面积上，则大于圆的面积，这样就可以得到圆面积的上限和下限。于是，刘徽采用了“以直代曲、无限趋近、内外夹逼”的思想，创立了“割圆术”。

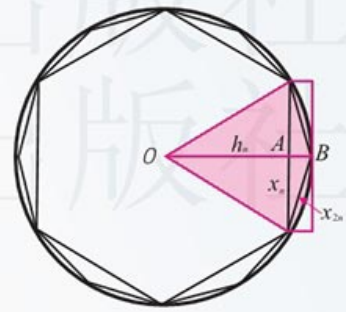


图2.4.3　“割圆术”示意图

中国地图出版社
人民教育出版社

“割圆术”从理论上能够把 π 的值计算到任意精度。刘徽用这种方法首先从圆内接正六边形开始割圆，算到正192边形的面积，得到 π 值为3.14，又算到正3 072边形的面积，得到 π 值为3.141 6。南北朝数学家祖冲之继承并发展了刘徽的“割圆术”，求得 π 的范围为3.141 592 6 < π < 3.141 592 7，这个值在世界上处于领先地位超过1 000年。

现在，我们可以利用计算机来计算圆周率。首先，先分析一下圆内接正六边形、正十二边形、正二十四边形……面积之间的关系，寻求它们的递增规律。如图2.4.3所示，设圆的半径为1，弦心距为h_n；正n边形的边长为x_n，面积为S_n，由勾股定理推导可得：$h_n=\sqrt{1-\left(\frac{x_n}{2}\right)^2}$，$x_{2n}=\sqrt{\left(\frac{x_n}{2}\right)^2+(1-h_n)^2}$（$n\geqslant 6$），其中$x_6=1$。观察发现，正$2n$边形的面积等于正$n$边形的面积加上$n$个等腰三角形的面积，即$S_{2n}=S_n+n\times x_n\frac{(1-h_n)}{2}$（$n\geqslant 6$）。随着$n$的不断增大，$S_{2n}$的值会不断趋近于 π。因此，利用这个推导公式，编写如下程序，即可求解 π 的近似值。

```
import math
n = int(input("请输入正多边形的边数："))
i = 6
x = 1
s = 6 * math.sqrt(3) / 4
while i <= n / 2:
    h = math.sqrt (1 - (x / 2) ** 2)
    s = s + i * x * (1 - h) / 2
    x = math.sqrt((x / 2) ** 2 + (1 - h) ** 2)
    i = 2 * i
print("当正多边形的边数为", n, "时，π的近似值为:", s)
```

2.4.2 基于枚举算法的问题解决

枚举法是依据问题的已知条件，确定答案的大致范围，在此范围内列举出它所有可能情况的方法。在列举过程中，既不能遗漏，也不能重复，通过逐一判断，验证哪些情况满足问题的条件，从而得到问题的答案。

在枚举算法的编程中，首先，要确定枚举对象和枚举范围，验证问题成立的条件；然后，借助循环语句和条件语句进行相应的程序设计，实现问题解决。

例2：票据中模糊数字推断问题。

问题：一张票据上有一个由4位数字组成的编号，甲说数字编号的前两位数字相同，但都不是零；乙说数字编号的后两位数字是相同的，但与前两位不同；丙说数字编号是一个整数的二次方。试根据以上线索推断出编号。

（1）分析问题

已知条件：假设4位数字的编号是$AABB$，其中$A\neq 0$，$A\neq B$，且$AABB$是一个整数的二次方；

求解目标：票据中的数字；

已知与未知的关系：要求解的4位数字的编号必须同时满足所有的已知条件。

（2）设计算法

根据问题分析，只要一一列举出4位数字$AABB$中A与B的所有可能组合，保证$A \neq B$且$A \neq 0$，再验证二次方问题，就可以得到问题的解。因此，该问题可使用枚举算法求解完成，其算法的流程图如图2.4.4所示。

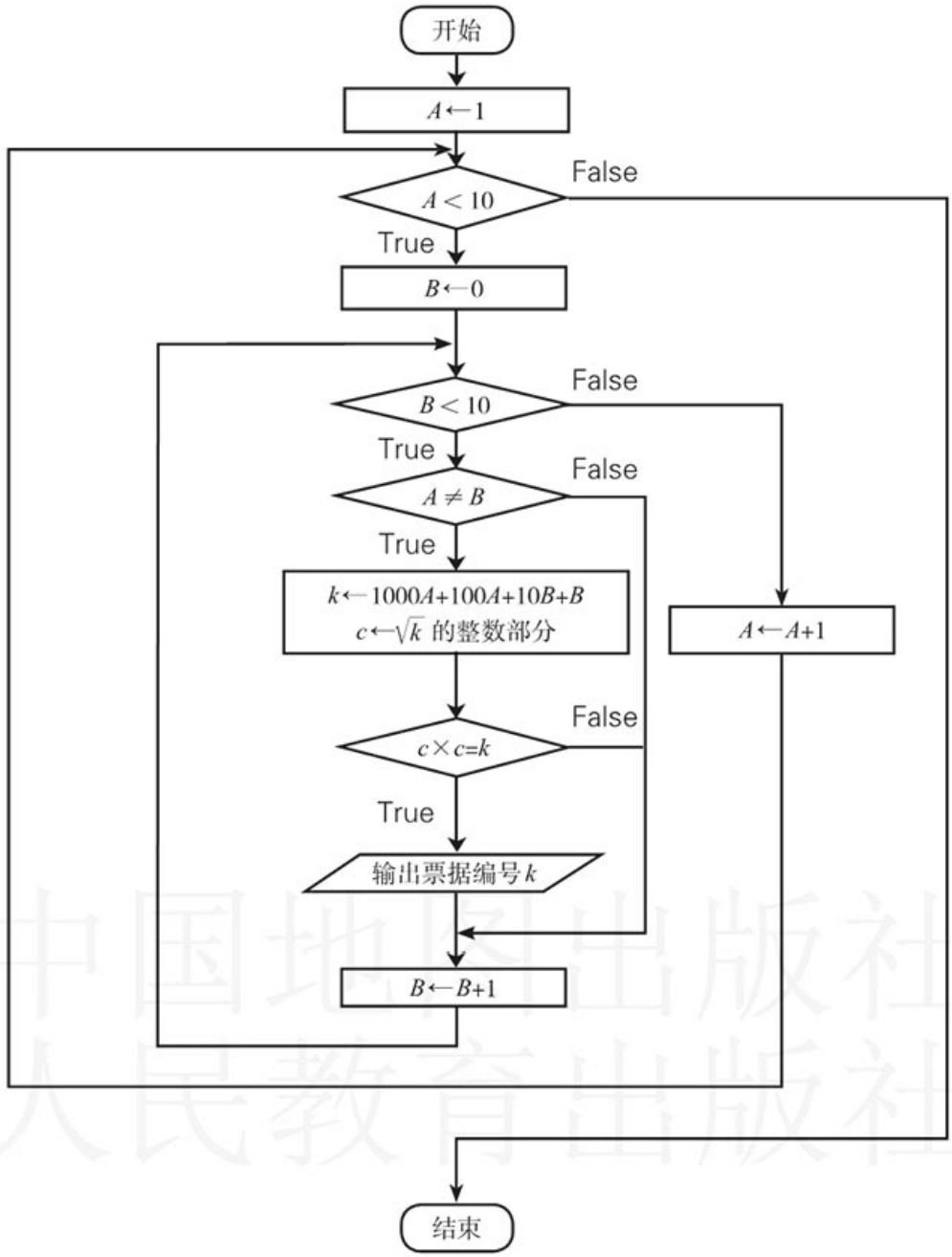


图2.4.4　票据中模糊数字推断问题求解流程图

（3）编程实现与调试

根据算法设计进行编程实现，程序示例如下：

中国地图出版社
人民教育出版社

```
import math
for A in range(1, 10):
    for B in range (0, 10):
        if A != B:
            k = A * 1000 + A * 100 + B * 10 + B
            c = int(math.sqrt(k))     # 求4位数字AABB的算术平方根并取其整数部分
            if c * c == k:            # 若k是完全平方数，则找到该票据编号
                print("票据编号是：", k)
```

（4）保存文件，调试运行程序

枚举算法在生活中有着比较广泛的应用场景，适合解决求解的答案数量有限，并且可能的答案是能按照某种规则列举出来的问题。例如，用枚举法解决一些数学问题（“韩信点兵”“鸡兔同笼”等）、益智游戏和逻辑推理等。

枚举算法需要逐一验证所有可能的情况，运算量比较大，解决问题的效率不够高。因此，在应用枚举法求解问题时，需要考虑优化算法，选择恰当的枚举对象，尽量分析出问题中的隐含条件，缩小枚举范围，以提高解决问题的效率。

实践活动

寻找1 000以内的所有素数

编程求解1 000以内的所有素数。素数是在大于1的自然数中，除了1和它本身以外不再有其他因数的数，如2，3，5，7，11，…

1. 分析该问题的枚举范围和验证条件。

枚举范围：_____。

验证条件：_____。

2. 用流程图描述问题求解的算法，并编程实现。

3. 教学资源平台中提供了本题的两种不同的算法方案（算法A和算法B），任选一种与自己设计的算法进行对比分析，完成表2.4.1。思考其中哪种算法能更高效地解决问题，为什么？

表2.4.1　两种方案的算法比较分析

算法方案	枚举对象	枚举范围	验证条件
自己的算法			
对比的算法（□算法A　□算法B）			

中国地图出版社
人民教育出版社

算法是人们在解决问题中积累的一种“智慧”。通过大量实践，现在已经总结出许多行之有效的算法，除了本节中用到的解析算法和枚举算法外，还有许多经典算法，如排序、查找和分治等。

阅读拓展

递归方法

递归是用于设计算法的一种思想方法，也是计算机科学中一个十分重要的概念。它通常是把一个大型复杂的问题层层转化为一个与原问题相似的、规模较小的问题来求解，通过构建递归关系式，将解决小问题作为解决大问题的入口，由此，大问题也就“迎刃而解”。

具体来讲，一个问题能够适用递归方法解决，必须符合两个条件：

■ 一个规模较大的问题可以分解为若干性质相同的规模较小的问题，这些规模较小的问题仍然可以进一步分解，分解出的新问题的解法和原问题的解法相同，只是处理对象的规模不同。

■ 必须有一个明确的结束递归的条件，不得无限递归。

递归不仅仅是设计算法的一种思想方法，也是一种简化问题的思维方式。

2.4.3 算法与程序实现的综合应用

算法设计及其程序实现是用计算机解决问题的核心过程。在具体的问题解决中，需要综合应用不同的算法思想并编程实现。当程序运行结果不能完全满足问题求解要求时，还要对算法和程序进行完善和优化。

例3：查找文稿中高频词的问题。

问题：学校开展经典诵读活动，小明在阅读《三国演义》时，为了分析小说的写作特色，想把小说中出现次数最多的20个词查找出来。想一想小明如何通过编写程序来实现呢？

（1）分析问题

要解决《三国演义》小说中高频词的查找问题，使用手工方式逐一查找统计，不仅费时费力，而且任务完成难度也非常大。此时，我们可以对问题进行抽象处理，确定出能用计算机解决的任务，将其转化为可计算问题，通过编程实现问题的高效求解。

已知条件：文本文件《三国演义》；

求解目标：《三国演义》中的高频词（以出现次数最多的20个词为例）；

已知与未知的关系：统计《三国演义》文本中词频，找出出现次数最多的20个词。

（2）设计算法

在查找文稿中高频词的问题求解中，除了要完成读取文件和显示输出内容，还要重点实现分词、词频统计和排序等功能。因此，该问题可以分解为如图2.4.5所示的5个功能。

中国地图出版社
人民教育出版社

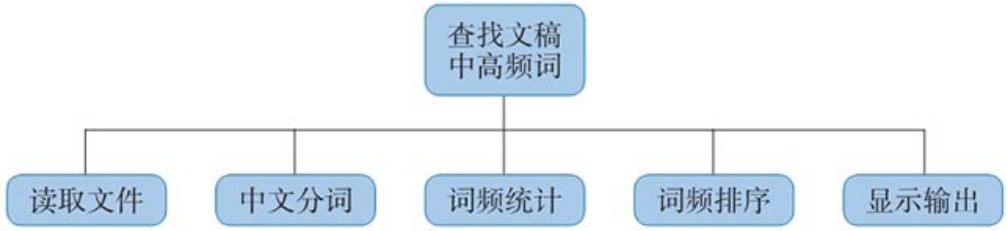


图2.4.5 “查找文稿中高频词的问题”功能分解示意图

1）读取文件

提前准备好《三国演义》文本文件，打开并读取文件内容。

2）中文分词

由于中文文本是由连续的字序列构成，没有明显的词语界限，因此分词处理的算法比较复杂。这里，我们可以直接使用现有的分词算法，如在Python语言中利用jieba分词功能，对读取到的《三国演义》字符串进行分词处理，如图2.4.6所示，将汉字字符序列切分成一个个单独的词，组合成一个“词汇表”，即词的序列。

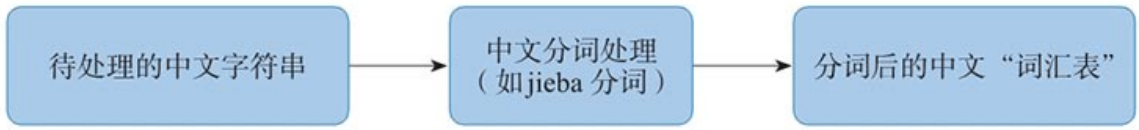


图2.4.6 中文分词过程示意图

3）词频统计

词频统计的过程主要应用了枚举算法，对于“词汇表”中的每一个词，依次计算出各自的出现次数，生成一个包含词和次数的“统计表”。需要注意的是，在算法设计时还要考虑问题求解的“高频词”应是具有明确指向意义的词语，不包括单个字的词。因此，《三国演义》小说中实现词频统计功能的算法流程图表示如图2.4.7所示。

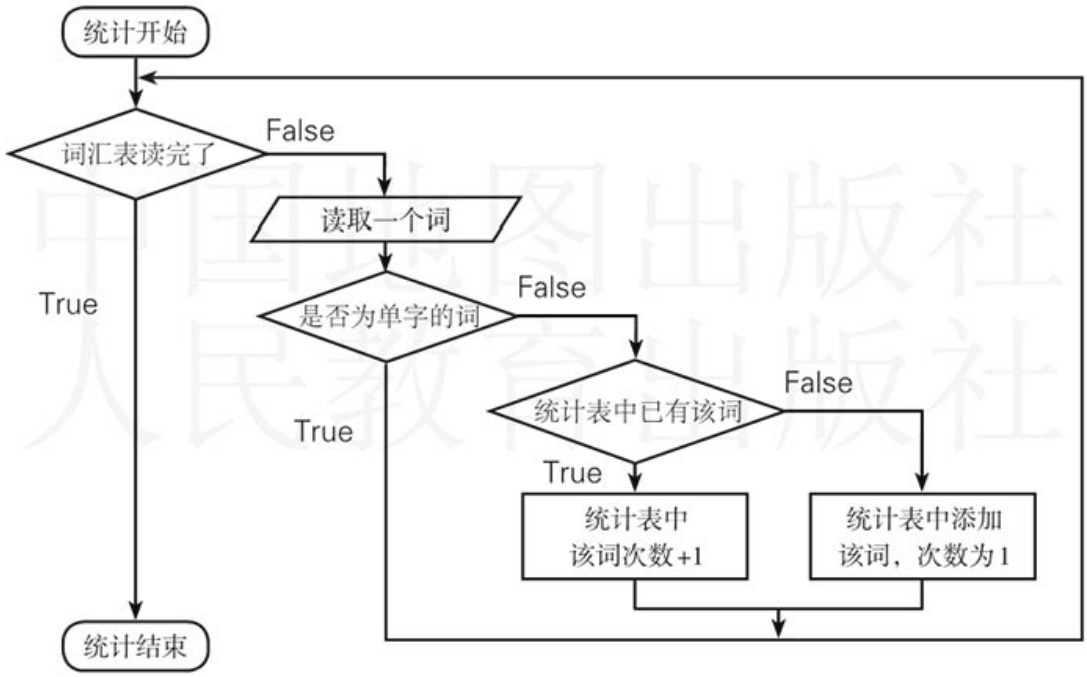


图2.4.7 《三国演义》小说中词语出现次数统计问题求解流程图

4）词频排序

要找到出现次数最多的20个词，需要对统计出来的词语按次数从大到小进行排序。排序的算法有很多，我们可以自己设计排序算法，也可以调用已有的排序算法功能，如在Python语言中直接调用内置的排序函数快速实现序列排序。

5）显示输出

降序排序后，序列中的前20个元素可认为是《三国演义》小说中出现次数最多的20个词。此时，只需要显示输出序列中这前20个元素的值（包括词和对应次数）即可。

（3）编程实现与调试

根据算法设计进行编程实现，程序示例如下：

```
# 打开"三国演义.txt"文件，读取文件内容
import jieba                                        # 导入jieba分词
f = open("三国演义.txt", "r", encoding="utf-8")      # 打开文件
txt = f.read()                                      # 读取文件内容

# 分词
words = jieba.lcut(txt)                         # 将中文文本拆分为词语保存到列表中

# 统计
counts = {}                                     # 建立空字典，用于存储词和出现次数
for word in words:
    if len(word) == 1:                          # 单字的词语忽略不计
        continue
    else:
        counts[word] = counts.get(word, 0) + 1 # 出现次数+1

# 排序
items = list(counts.items())                     # 字典转换为列表后才可以排序
items.sort(key=lambda x:x[1], reverse=True)      # 对列表items按"次数"降序排序

# 输出前20个元素的值
for i in range(20):
    print(items[i][0], items[i][1])             # 输出词和出现次数
```

编程解决问题不是一蹴而就的。通常，人们需要根据程序运行结果，围绕问题需求，不断修改方案、改进算法、优化程序、完善功能，以实现问题的最终有效解决。例如，分析《三国演义》小说高频词查找程序的运行结果，会发现其中“诸葛亮”“孔明曰”“孔明”在小说中其实指向同一人物，“关羽”“关公”“云长”是同一人物称谓，同理，“刘备”“玄德”“玄德曰”及“曹操”“孟德”“丞相”等也都是多个词语指向同一人物。此时，需要修改算法，利用选择结构将这些词语进行合并处理，其算法实现的关键程序代码为：

```
for word in words:
    if len(word) == 1:
        continue
```

```
elif word == "诸葛亮" or word == "孔明曰":
    rword = "孔明"
elif word == "关公" or word == "云长":
    rword = "关羽"
elif word == "玄德" or word == "玄德曰":
    rword = "刘备"
elif word == "孟德" or word == "丞相":
    rword = "曹操"
else:
    rword = word
counts[rword] = counts.get(rword, 0) + 1
```

（4）保存文件，调试运行程序

为了更有效地查找、修改程序中存在的错误，需要仔细阅读和分析程序语句，掌握必要的调试方法。

较简单的调试方法是通过函数print( ) 直接输出程序执行过程中的变量值。

断点调试是一种较为直观的程序调试方式，它通过设置断点跟踪变量的取值，观察运行结果，进行程序调试。断点调试的基本方法为：

- 进入调试状态；
- 设置断点；
- 检查运行状态下各个变量的值，确定错误的位置，并进行修改；
- 反复调试直至程序运行正确。

技术支持

Python语言中的字典类型及排序方法

1. 字典类型及相关操作

字典是Python语言中的一种数据类型，其中每个元素都包含两部分：键和值。字典中元素用花括号括起，用冒号分隔键和值。例如，表示教师姓名及其年龄信息的数据可用字典类型表示为：

teachers = {"张明":25, "王丽":32, "李强":56 , "孙玲":44}

字典与列表不同，它不是数据的序列，而是键值对的无序集合。字典可以通过键访问值，这种方法与平时我们查字典或电话簿类似，如teachers["张明"]。

Python语言中内置了一些函数和方法，可以对字典类型数据进行操作，完成特定功能。表2.4.2给出了Python语言中对字典类型数据进行操作的部分函数。

表2.4.2　Python语言中部分字典函数操作

函数	功能说明
dict.get(key, default=None)	返回字典dict中指定键的值，如果值不在字典中，则返回该默认值
del(dict[key])	从字典dict中删除指定键的元素
list(dict.items())	将字典dict转换为列表类型

2. 排序方法

Python语言中内置了多种排序方法，函数sort就是其中常用的一种。例如，list.sort( )可以对列表list进行升序排序，若要降序排列，需要设置参数：reverse = True，如list.sort(reverse = True)。

要对字典类型数据进行排序，通常先将其转换为列表类型，然后再进行排序操作。例如：

```
counts = {"Apple":25, "Orange":40, "Green tea":28, "Coffee":15}
L = list(counts.items())
L.sort(key=lambda x:x[1], reverse=True)  # 根据列表中的第2列进行降序排序
```

排序后列表L的值为：[("Orange", 40), ("Green tea", 28), ("Apple", 25), ("Coffee", 15)]。

这里，使用lambda定义了一个匿名函数，作为sort( )的参数。lambda定义匿名函数的格式为：lambda<参数列表>:<表达式>，函数的返回值是表达式的值。

当程序调试、运行正常并得到正确结果后，就实现了前期设计的算法，解决了相应的问题。当然，要想更深入地了解算法与程序设计，还需要进一步学习编程语言、算法设计及数据结构等知识，创造性地利用计算机解决问题。

中国地图出版社
人民教育出版社

实践活动

编程查找小说中的主要人物

修改前面编写的《三国演义》小说高频词查找程序，实现功能：找出小说中出场次数最多的10位人物。

1. 在前面已经查找到的20个高频词中，包含有“却说”“二人”“不能”等与人名无关的词语。想一想如何去除这些词语，只显示10位出场次数最多的主要人物的人名？试着修改程序。

2. 统计小说中的高频词，不仅可以帮助我们了解其中的主要人物，还可以用来分析人物的主要活动地点及人物间的关系等。试着编程分析自己感兴趣的一本小说，说一说你的发现。

项目实施

丰富“自助式人行过街红绿灯”的功能

一、项目活动

1. 收集学校、医院或居住地周边的马路通行状况（如马路宽度、高峰段、人流量和车流量等信息），思考为其设计“自助式人行过街红绿灯”的合理配时方案，进一步完善程序功能，以实现真实问题求解。

2. 思考如何实现更人性化、智能化的交通管理，设计一种在智能交通环境下，控制十字路口红绿灯时长变化的合理方案，并在小组间交流展示。

建议：结合物联网和大数据技术为智能交通带来的新可能，有效整合多维的、海量的数据，挖掘各种交通数据之间的联系，思考精准化、人性化设计或提升交通智能化管理的策略，尝试描述相应的解决方案。

二、项目检查

1. 结合实际情况，优化“自助式人行过街红绿灯”程序功能，调试运行程序，完成项目。

2. 科学设计和完善交通智能化管理的方案。

练习提升

1. 交通信号“绿波带”是根据车辆运行情况对各路口红绿灯进行协调，使车辆通过路口时能连续获得一路绿灯。某路段启用了“绿波带”技术，如图2.4.8所示，全长1.6 km，5个灯控路口，提示的“绿波速度”为60 km/h，假设5个灯控路口间距相等。在仅考虑一辆车通行的情况下，如果实现“绿波”交通，那么相邻两个路口间绿灯亮起的最大时间间隔应该设置为多少？思考该问题求解的算法并编程实现。

中国地图出版社
人民教育出版社

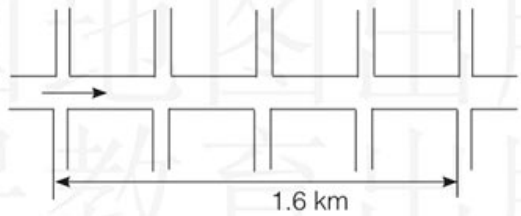


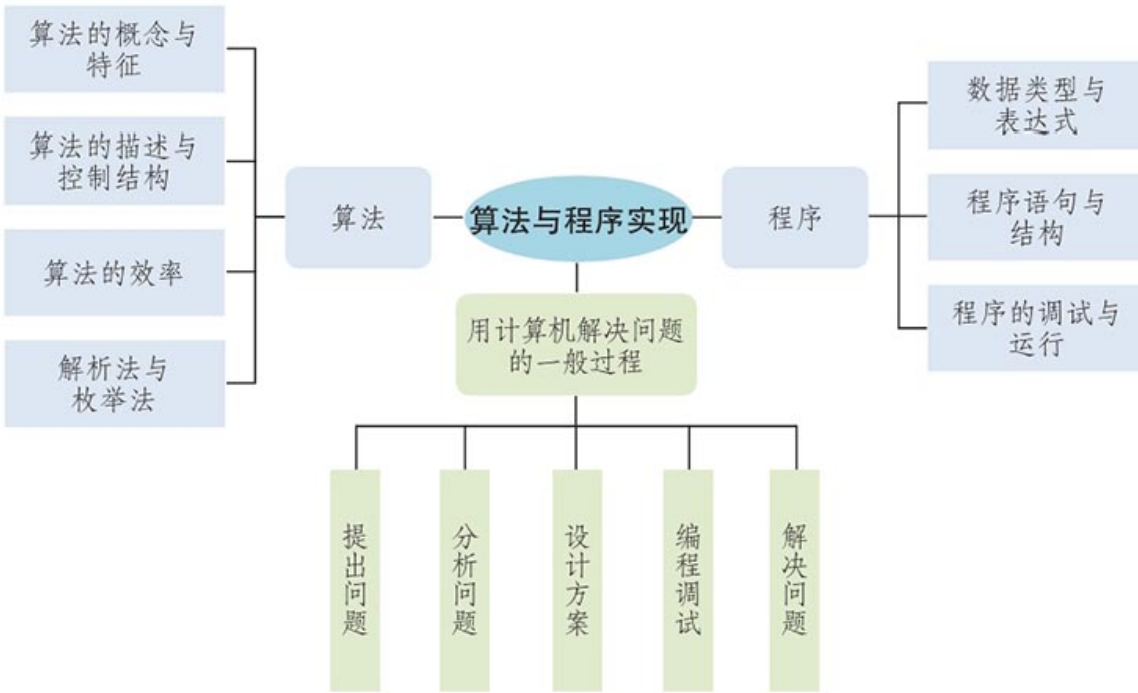
图2.4.8　路段示意图

2. 韩信是汉初著名军事家，传说他统计士兵数目有个独特的方法。例如，先令士兵排成5列纵队，结果余1人；接着，命令士兵排成6列纵队，结果余5人；再命令士兵排成7列纵队，结果余4人；最后，命令士兵排成11列纵队，结果余10人。这样他便知道士兵的总人数了。这种计数的方法被后人称为“韩信点兵”。试编写程序计算士兵的数目。

3. 编写程序，统计一段文本中分别有多少个汉字、英文字母、数字和其他字符。

总结与评价

1. 下图展示了本章的核心概念与关键能力，请同学们对照图中的内容进行总结。



2. 根据自己的掌握情况填写下表。

学习内容	掌握程度
用计算机解决问题的过程	□不了解 □了解 □理解
算法的概念与特征	□不了解 □了解 □理解
描述算法的方法	□不了解 □了解 □理解
算法的基本控制结构	□不了解 □了解 □理解
数据类型	□不了解 □了解 □理解
常量、变量及表达式	□不了解 □了解 □理解
程序语句与结构	□不了解 □了解 □理解
解析算法	□不了解 □了解 □理解
枚举算法	□不了解 □了解 □理解

中国地图出版社
人民教育出版社

第 3 章
数据处理与应用

南水北调工程是我国继三峡工程之后，又一个重大水利工程。南水北调工程的主要目的，是将长江流域的水资源科学合理地调配到我国华北和西北部分省市，解决当地水资源短缺问题。2014 年 12 月 12 日，南水北调中线工程正式通水，开始发挥其巨大的经济效益和社会效益。南水北调工程建设得益于我国强大的科技与工程实力，其中包括信息技术的大量应用。不仅每日采集大量数据，同时还要对庞杂的数据进行存储、筛选、加工和分析，以确保整个工程的安全与高效。可以说，没有巨量的科学数据和强大的科技实力做支撑，修建这样一个浩大的工程是难以想象的。

同学们，在信息社会中，大量数据的产生和积累，为人们提供了认识这个世界的新方式，但也对人们驾驭数据的能力提出了新挑战。如何利用信息技术有效处理数据，发现并利用其中的价值，已成为信息社会生存的一项基本能力。

在本章的学习中，我们将借助信息技术手段采集、分析和可视化数据，通过“用水分析助决策”项目活动，掌握数据处理和应用的基本方法。

3 主题学习项目：用水分析助决策

项目目标

南水北调工程是缓解我国北方地区水资源短缺局面，优化我国水资源配置，促进经济和社会可持续发展的重大战略性基础工程。本章围绕南水北调工程的实施情况开展项目学习。建议从供水、节水和水污染防治等方面确定项目研究主题，处理水资源数据，提交项目报告。

1. 通过项目活动，学会数据采集、分析和可视化的方法。
2. 能够从数据中提炼出有用的信息，提出合理的用水建议。
3. 合理选用数据处理工具，完成数据分析报告。

项目准备

为完成项目，需做以下工作。

- 3～5名同学组成一个小组，明确分工，制订项目规划。
- 查阅获取项目主题相关资料，结合地理知识，分析我国水资源特点。
- 安装数据处理软件，搭建数据分析和可视化环境。

项目过程

在学习本章内容的同时开展项目活动。为了保证本项目的顺利完成，要在以下各阶段检查项目的进度。

1 制订规划

小组讨论并制订项目规划，包括项目目标、实施方法和路径等。用思维导图记录规划结果。

P90

2 收集数据

使用网络搜索等方法收集、整理并存储相关水资源数据。

P103

3 分析与可视化

选用恰当方法分析水资源数据，进行可视化呈现。

P123

4 撰写报告

梳理并总结数据分析过程，撰写数据分析报告。

P129

项目总结

完成本章项目后，各小组要提交水资源数据分析报告，并进行交流分享。通过项目活动，感受数据在决策中的作用，体验数字化学习方式。

3.1
数据处理的一般过程

学习目标

- 认识数据处理，感受数据处理对日常生活的影响。
- 了解数据处理的一般过程，能通过数据处理获取有价值的信息。

体验探索

对比我国部分城市降水量和气温数据

我国幅员辽阔，南北气候差异较大。采集与分析气象数据，可以了解各地情况。图3.1.1是我国由北至南4个城市2016年降水量和气温数据。试从以下几方面对比图表中的数据：

1. 对比4个城市全年降水总量；
2. 对比哈尔滨和广州的气温变化曲线。

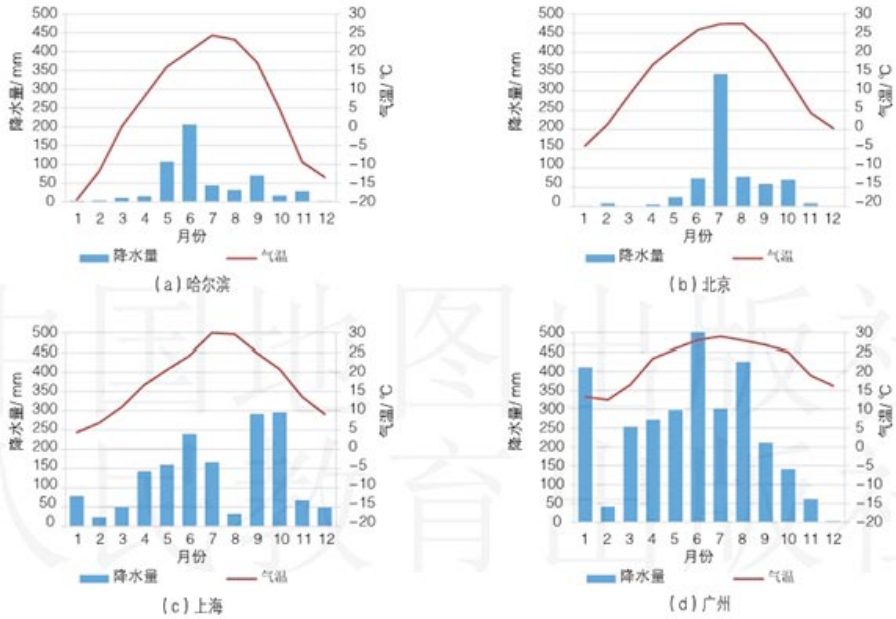


图3.1.1　2016年我国部分城市降水量和气温图

思考：

1. 通过对比，能从图中获取哪些信息，这些信息对制定用水政策有什么帮助？

2. 了解我国水资源分布的特点，思考实施南水北调工程的原因。

3.1.1 数据处理

日常生活和社会生产都有数据的产生与传播。这些数据经过处理后，就有可能发现其中包含的信息，挖掘其内在价值，更好地为我们服务。数据处理广泛应用于日常生活和社会生产的各个领域。例如，在身体锻炼的过程中，通过及时收集、分析生理和运动数据，合理调整运动计划，可提高运动质量。再如，在南水北调工程规划过程中，通过数据处理，全面掌握全国水资源分布情况，了解区域性资源供需矛盾，为南水北调工程的实施提供决策依据。

思考活动

南水北调工程中的数据处理

南水北调是一项跨世纪的重大工程，关系到子孙后代的利益，必须慎重研究，充分论证，科学决策。自20世纪50年代提出南水北调设想以来，有关部门进行了广泛深入的研究论证，确定了从长江上、中和下游干支流调水的南水北调西线、中线和东线工程的总体格局，制定了详细的工程总体规划。其中，中线工程规划中给出以下结论：实施中线工程，补充京、津和华北平原的水资源供应量，是缓解京津等地区水资源供需矛盾的最佳选择。

思考：

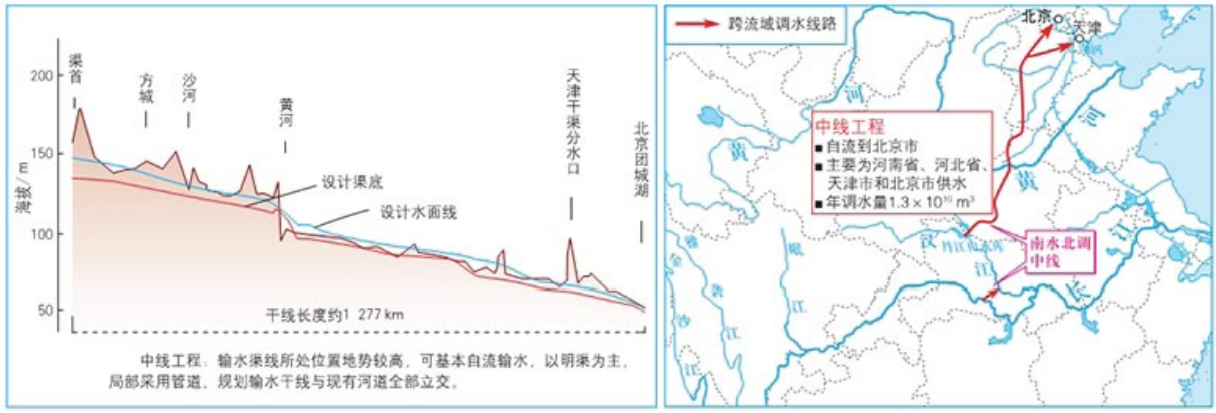
试查阅中线工程规划等相关资料，思考得到中线工程结论需要用到哪些数据，这些数据是如何被处理的？

数据处理一般指对数据进行采集、整理、分析和可视化表达的过程。

数据采集可为数据处理准备必要的数据。例如，为科学决策南水北调工程，论证我国北方地区是否缺水、缺多少水、水源地能调出多少水等问题，研究者就需要在采集多年的相关水资源数据的基础上，深入分析、严格论证，为分析问题、做出决策提供有力支撑。

用一定的方法对收集来的数据进行整理、分析，提取有用信息，形成结论，可以指导决策。例如，南水北调中线工程在计划加高调水源头的水库大坝的论证中，设计人员分析测算调水量、发电量、移民量等数据，经过反复比较论证、综合评估，最终确定了将水坝加高到176.6 m的方案。

借助数据可视化呈现方式能够更好地表达数据所蕴含的信息。例如，南水北调工程线路示意图以地图形式形象地展现了中线调水线路的走向和途经地区，为决策者和工程建设者们提供了帮助，如图3.1.2所示。



（a）中线工程输水干线纵断面示意图　　（b）南水北调中线线路走向示意图

图3.1.2　南水北调中线工程示意

可见，在数据处理过程中，从大量的、无序的、难以理解的数据中，提取有价值、有意义的数据，分析这些数据，获得有用的信息，有助于做出正确的判断和决策。例如，南水北调工程的规划需要处理众多数据，如受水区需调水量、水源投资、水库库容和工程效益等，涉及经济、社会和环境等各个方面。这些数据的处理为做好南水北调工程规划发挥了重要作用。

阅读拓展

数据分析助决策

数据的分析与应用，能够为制定决策提供依据和支持。例如，2016年底至2017年初，黄果树景区推出“全国免费活动周”。活动周期间，景区人流和车流出现了明显的滞留拥堵，景区游客变化趋势如图3.1.3所示。为了揭示滞留拥堵的原因，贵州交管局对2月10日—16日的景区旅游人流数据进行了分析，发现免费旅游活动措施虽然是针对全国的活动，但游客主要来自省内，外省游客相对较少，如图3.1.4所示。

图3.1.3　黄果树景区游客数量变化趋势

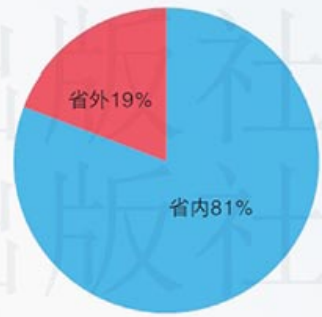


图3.1.4　黄果树景区游客来源分布（2月10日—16日）

从图3.1.3中可看出，2月10日—16日，景区游客人数多日接近甚至超过饱和接待值，图3.1.4则反映出这段时间内的游客绝大多数来源于省内。

据此，相关部门给出了有针对性的建议，如优化调整免费政策实施细则、采取免费活动周对省内游客按地市进行轮换免费等措施以缓解景区压力。

通过对黄果树景区游客数据的处理，还能够分析并预见安全隐患，为现场应急处置，特别是制订应急预案提供科学合理的建议，为政府相关部门提供了重要的决策依据。

3.1.2 数据处理的过程

面对纷繁复杂的数据，虽然分析的目标可能不同，对象也各异，但数据处理的流程基本一致，都要经历数据采集、数据整理、数据分析和数据呈现等环节。

例如，黄果树景区案例中对游客数据的处理经历了以下过程：首先采集并整理景区游客数据，如免费周期间每天省内省外的游客数量；然后计算出游客变化比例、园区饱和度等数据并进行分析，再采用图表等直观的方式呈现出来。对黄果树景区游客数据的处理，为旅游、交通等管理部门制订方案、调整政策，做出正确决策提供了支持。

1. 数据采集

数据有不同的表现形式，数据的来源也多种多样，因此数据采集的方法、采集的工具和采集后的存储方式也不尽相同。在“分析供水情况”的主题中，所需要采集的供水量数据，可以通过官方网站获取，也可以通过网络调查的方式获取。例如，要调查2016年我国部分地区水库蓄水情况，可通过国家统计和水利等部门的官方网站获取权威数据，表3.1.1就是从中华人民共和国水利部（以下简称水利部）网站获取到的黑龙江、北京、湖北和广东2016年水库蓄水量数据。

表3.1.1 我国部分地区水库蓄水量统计表（2016年）

序号	地区	汛初蓄水量 /（$\times 10^8$ m^3）	汛末蓄水量 /（$\times 10^8$ m^3）	汛初较常年同期增长百分比 /%	汛末较常年同期增长百分比 /%
1	黑龙江	19.9	82.3	19	−2
2	北京	13.7	18.3	−12	−2
3	湖北	495.6	630.9	19	12
4	广东	197.5	214.5	26	16

2. 数据整理

数据整理通常指对数据进行校验和标准化。由于采集到的数据可能有缺失、重复或错误，因而需要进行必要的整理。通常，数据整理是开展数据分析前的重要准备工作，整理后的数据按照一定的标准进行存储。例如，对采集到的黄河流域的水文资料数据进行整理、纠错，并将有效数据设置为两位小数的数据格式等。

3．数据分析

数据分析指运用适当的分析方法和工具，对整理后的数据加以详细研究和概括总结，从中提取有价值的信息，最终形成结论的过程。例如，对比黄河流域和长江流域近年来降水量、水位情况等数据，理解南水北调工程对我国水资源优化配置的意义。

4．数据呈现

数据呈现是将数据分析结果以恰当的方式呈现出来，以便于人们理解和使用。可视化表达具有直观、生动和易于理解的优势，被广泛应用于数据和数据分析结果的表示和呈现。例如，图3.1.5所示的可视化图表，以柱形的高度直观地对比出各地区汛初、汛末的蓄水量。

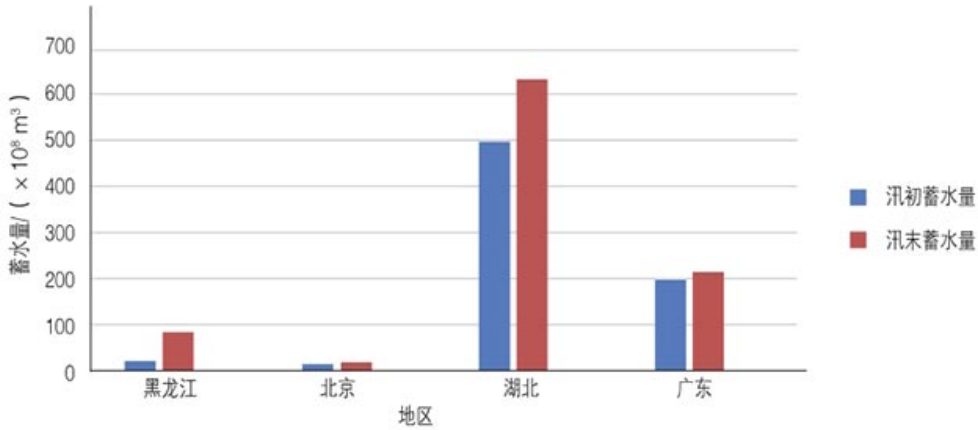


图3.1.5　2016年我国部分地区水库蓄水量情况

经过采集、整理、分析和呈现这几个环节，我们就有可能从数据中得到有价值的信息，进而利用这些信息做出正确的判断和决策。

实践活动

分析我国长江流域和黄河流域降水量情况

降水量的多少对水资源分布具有重要影响，分析各地降水量可以帮助我们从某个角度了解当地水资源情况。现有我国长江流域和黄河流域降水量数据，如表3.1.2所示。

表3.1.2　2006—2015年长江和黄河流域年平均降水量

单位：mm

流域	2006年	2007年	2008年	2009年	2010年	2011年	2012年	2013年	2014年	2015年
长江流域	974.5	1 011.1	1 072.4	998.7	1 160.4	931.3	1 159.3	1 029.6	1 100.6	1 134.4
黄河流域	407.2	484.1	433.1	440.4	449.2	489.1	490.1	481.6	487.4	409.5

试分析上述流域降水量数据，对比这些流域的水资源差异。

中国地图出版社
人民教育出版社

项目实施

制订项目规划

一、项目活动

围绕南水北调工程，从我国水资源数据入手，分析南水北调的可行条件。增强水资源忧患意识和水资源节约保护意识，形成节约用水、合理用水的良好风尚。

1. 确定主题

从供水、节水和水污染防治等方面，确定小组项目主题。

2. 制订项目规划

小组合作共同制订项目规划。规划内容主要包括项目研究目标、分析内容、成员分工、预期成果和时间进度等。

3. 绘制思维导图

用思维导图记录小组讨论和规划结果。

二、项目检查

确定主题，拟定项目规划，用思维导图呈现规划结果。

练习提升

1. 信息技术的发展，为同学们提供了丰富的数字化学习工具。试从数字化学习工具种类、使用广泛度、应用效果等方面设计一个数字化学习工具使用情况调查方案（方案中应当体现数据处理的一般过程），了解同学们使用数字化学习工具的现状。

2. 随着互联网学习资源的不断丰富，越来越多的人利用网络教学资源进行学习。互联网学习平台能够分析学习者在网络学习中的行为，可以掌握学习者的学习特征，发现学习中存在的困难，进而调整学习者的学习策略或方法。

思考：

① 为实现上述目标，需要处理哪些学习者网络学习行为方面的数据？

② 在处理这些数据时，怎样保护学习者的个人隐私不受侵犯？

中国地图出版社
人民教育出版社

3.2
数据采集与整理

学习目标

- 掌握数据采集和整理的基本方法，能根据需求选择适当的工具采集与整理数据。
- 应用恰当的方法保护数据，理解对数据进行保护的重要意义，增强数据安全意识。

体验探索

你能找到这些数据吗

参照黄河流域水资源数据汇总表（2014—2016年），见表3.2.1，收集并填写其他流域（如长江流域、珠江流域、松花江流域等）的水资源数据汇总表，见表3.2.2。

表3.2.1　黄河流域水资源数据汇总表

项目	2014年	2015年	2016年
水资源总量/（$\times 10^8$ m^3）	653.7	541.0	601.8
供水总量/（$\times 10^8$ m^3）	387.5	395.5	390.4
用水总量/（$\times 10^8$ m^3）	387.5	395.5	390.4
数据来源	水利部网站	水利部网站	水利部网站

表3.2.2　_____流域水资源数据汇总表

项目	2014年	2015年	2016年
水资源总量/（$\times 10^8$ m^3）			
供水总量/（$\times 10^8$ m^3）			
用水总量/（$\times 10^8$ m^3）			
数据来源			

思考：

获取上述数据，可采用哪些方法，怎样才能确保数据的准确性？

中国地图出版社
人民教育出版社

3.2.1 数据采集

数据采集，即根据需求采用适当的方法和工具获取所需要的数据。例如，采集交通出行数据用以预测交通情况，采集全国不同流域水资源数据用以分析水资源分布状况等。根据需求获取数据，为进行数据分析及获取有价值的信息奠定基础。

那么，在实际应用中，如何采集所需要的数据呢？通常，数据采集的主要环节包括明确数据需求，确定数据来源，选择采集方法和实施数据采集。通俗地说，就是找什么，去哪儿找，怎么找。

明确数据需求

为提高数据采集的针对性，采集数据前要明确数据需求。可以从需要分析的问题、研究的内容、期望达成的目标等方面来考虑，然后列出需要采集的数据目录。

思考活动

确定“研究我国水资源分布情况”的数据需求

黄淮海流域是我国水资源承载能力与经济社会发展矛盾最为突出的地区，为缓解该流域日益严重的水资源短缺问题，促进经济发展和社会进步，在深入研究和科学分析的基础上，政府做出实施南水北调工程的重大决策。其中，分析各地区水资源总量、人均水资源量等水资源数据，是掌握我国水资源分布情况必不可少的研究之一。

思考：

阅读以上材料，思考研究我国水资源分布情况需要获取哪些方面数据，并用思维导图呈现。

确定数据来源

在信息社会，数据来源途径广泛，社会调查、公众媒体、科学实验与实践活动等都可以提供大量数据。伴随着互联网的快速发展，政府机构的官方网络平台已成为获取权威数据的重要渠道。此外，物联网和社交网络中也拥有巨量数据。例如，在利用物联网实现的智能家居系统中，涉及温度、湿度、空气中挥发性有机化合物等多种数据。庞大的社交网络用户每天通过即时通信、微博、微信等社交应用，不断地产生出文字、图像、视频等各类数据。

通常，数据来源越可靠，收集渠道越多，就越容易收集到真实有效的数据。因此，在明确数据需求后，为了更好地应用数据解决问题，就需要在众多的数据来源中根据实际情况恰当地选择。例如，“用水分析助决策”项目活动所需要采集的数据，既可以通过查找图书、报刊等文献资料获取，也可以请教水资源方面的专家，还可以通过访问中华人民共

和国国家统计局（以下简称国家统计局）和中国南水北调工程等网站获取水资源的数据。

选择采集方法

确定数据需求和数据来源后，应选择恰当的方法采集数据。信息技术为高效精准地采集数据创造了便利的条件，应用传感器和网络等是当前数据采集的重要方法。

1. 传感器采集数据

传感器是一种检测装置，能感受到被测量的信息，并能将信息按一定规律转换成电信号或其他所需形式的信息输出。常用的传感器有温度传感器、压力传感器、红外传感器、距离传感器和声音传感器等，如图3.2.1所示。

（a）温度传感器　（b）压力传感器　（c）红外传感器

图3.2.1　常用传感器

当今，传感器和网络的结合创设出智能化生活环境。例如，楼宇中安装的智能水表就是利用传感器和无线网络自动采集用户用水量数据。图3.2.2展示的是一个远程智能抄表系统。抄表人员只需操作相应软件，即可了解用户的用水情况，实现远程检测。

图3.2.2　远程智能抄表系统

传感器不仅应用在日常生活中，而且在工业生产中也应用广泛。例如，自动化工厂中的设备借助各种传感器采集的数据，监测和控制生产中的各项参数，保障生产的正常运转。

传感器是物联网中的一种重要设备。在物联网中，通过传感设备，按约定的协议，将相关物体与网络连接，进行信息采集和通信，实现智能化识别、定位、跟踪、监控和管理。物联网是国家新兴战略性产业，已经在很多行业得到了广泛应用。

例如，南水北调工程中就大量应用了物联网技术，其监测数据越来越多地来自于水位计、流量计、水质监测设备、位移传感器和工程安全传感器等各种物联网传感器。在南水北调东线山东段94个闸泵站/监测点实时运行时，每隔一段时间采集一次水位、水情、水质、流量、闸门开度和基础运行状态的基础数据，每天数据量大概50万条左右、1 GB 容量左右的监测数据。这就要求其技术架构能够充分应对“庞大的数据实时收集，大规模存储，实现快速数据查询、历史数据挖掘和分析评估，为水量调度和综合会商提供充分的信息支持服务”等需求。

物联网传感器获取的监测数据通过自动化控制设备或直接连接网络实现定时的数据采集、信息转换和信息上传。

实践活动

体验手机传感器的数据采集

智能手机内置了多种传感器，通过手机应用软件可以读取相关传感器的数据，实现相应的功能。例如，重力传感器使手机能够自动在横竖屏间切换，温度传感器可用于监测手机发热情况，指纹传感器可用来识别身份等。图3.2.3呈现的即为手机中的传感器采集的数据。

图3.2.3　手机中传感器采集的数据

体验智能手机中的传感器应用，举例说出手机中还采用了哪些传感器来采集数据？采集到的数据帮助手机实现了哪些功能？

中国地图出版社
人民教育出版社

2. 网络获取数据

随着移动互联网的日益普及和数字化媒体的迅速崛起，网络成为人们快捷获取数据的重要渠道。目前，使用网络爬虫采集互联网上的数据成为获取网络信息的一种有效方式。网络爬虫，是按照一定的规则，自动抓取互联网内容的程序。网络爬虫的主要功能是自动采集其可以访问到的网页内容，这种技术已被广泛用于搜索引擎或其他类似网站。

搜索引擎是一种能为用户提供检索服务，并将检索结果呈现给用户的系统。网络爬虫是搜索引擎的重要组成部分。图3.2.4展示了搜索引擎的一般工作原理：网络爬虫通过网络链接进入网站，采集网站页面内容，采集到的内容经过处理后得到的索引结果被存放到搜索引擎的索引库中，当用户发起搜索请求时，搜索引擎检索索引库后，将结果返回给用户，从而实现快速检索。

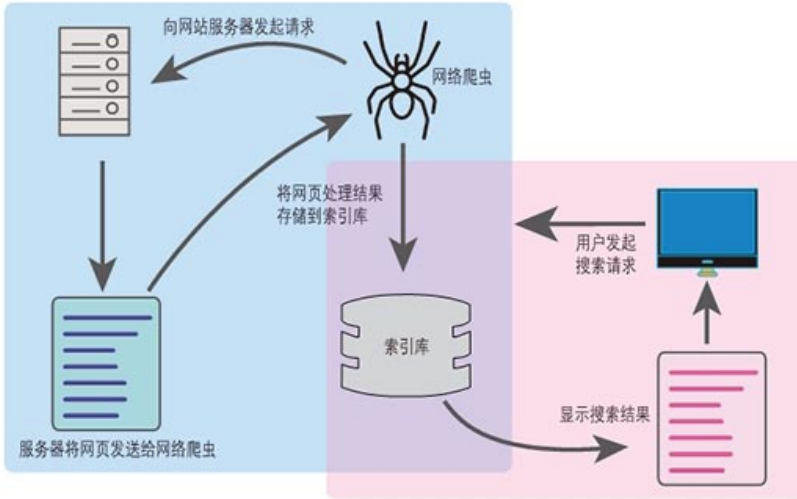


图3.2.4　搜索引擎的一般工作原理

海量数据蕴藏在成千上万的网页中。为便捷地获取这些数据，人们开发了用于采集数据的软件和在线平台。在现代社会中，熟练掌握通过网络获取工具采集数据的方法，是十分重要的。

在线问卷是通过网络采集数据的另一种方式。这种方式通过在线调查问卷网站完成问卷的设计、发放、回收和分析等工作。例如，通过“高中生信息安全问卷”调查，可以了解学生对信息安全知识的掌握情况；通过“家庭用水情况问卷”调查，可以了解一般家庭用水和节水情况等。

实践活动

通过网络获取水资源数据

1. 在前面的思考活动中，我们已经确定了“研究我国水资源分布情况”需要获取的数据，利用搜索引擎搜索我国不同地区水资源数据，并记录获取数据的网址。

分析提供水资源数据的网站可信度如何？网站发布的数据是否真实准确？试说明理由。

2. 为监测长江水资源生态环境，保障水资源质量，长江流域水资源保护局定期发布“长江流域水资源质量公报”，每份公报中通常包含5张表格，部分内容如图3.2.5所示。

中国地图出版社
人民教育出版社

长江流域2018年4月份水资源质量公报

时间：2018年07月04日

一、水功能区达标情况

1、总体达标情况

本月长江流域及西南诸河共评价1 241个水功能区，1061个达标，达标率为85.5%。水功能区总评价河长47 756.2 km，达标河长42 843.0 km，占总评价河长的89.7%；湖（库）总评价面积9 649.9 km²，达标面积6 289.8 km²，占总评价面积的65.2%。水功能区总体达标情况见表1，各水功能区的水质实测达标情况详见附表一。

表1 4月份水功能区达标情况统计表

水资源分区	水功能区类型		按水功能区个数统计			河流按代表河长统计			湖(库)按代表面积统计		
	一级水功能区	二级水功能区	评价个数	达标个数	达标率(%)	代表河长(km)	达标河长(km)	达标率(%)	代表面积(km²)	达标面积(km²)	达标率(%)
长江流域	保护区		133	112	84.2	5 501.9	5 124.7	93.1	7 205.9	4 839.8	67.2
	保留区		356	306	86.0	24 383.4	21 364.5	87.6	268.7	101.7	37.9
	缓冲区		87	72	82.8	2 612.9	2 198.8	84.2	42.0	0	0
	其中省界缓冲区		84	69	82.1	2 579.1	2 165.0	83.9	42.0	0	0
	小计		576	490	85.1	32 498.2	28 688.0	88.3	7 516.6	4 941.5	65.7
	开发利用区	饮用水源区	211	181	85.8	2 038.7	1 793.5	88.0	76.3	14.3	18.7
		工业用水区	196	178	90.8	2 476.2	2 189.3	88.4	161.4	91.4	56.6
		农业用水区	24	21	87.5	731.9	681.8	93.2			
		渔业用水区	10	10	100	97.7	97.7	100	445.6	445.6	100.0
		景观娱乐用水区	46	35	76.1	809.2	685.7	84.7	259.1	244.1	94.2
		过渡区	79	59	74.7	574.8	389.6	67.8	3.3	3.3	100
		排污控制区									
	小计		566	484	85.5	6 728.5	5 837.6	86.8	945.7	798.7	84.5

图3.2.5　长江流域2018年4月份水资源质量公报页面

试使用教学资源平台的网络爬虫程序ywrp_crawler.py，获取公报中的表格数据，记录爬取下来的文件数量和时间，并了解其功能特征及设计思路。

技术支持

使用网络爬虫采集数据

长江流域水资源保护局每月发布“长江流域水资源质量公报”，如图3.2.6所示。如果想用人工单击网络链接，逐一下载的方式获取多年来公报中提到的数据，将非常费时费力，而使用网络爬虫就能较轻松地采集。

图3.2.6　长江流域水资源质量公报链接页面

网络爬虫功能强大，可以满足多种数据采集需求。例如，使用网络爬虫获取高考选考科目数据，帮助学生选择修习课程；获取运动员比赛数据，帮助教练了解和分析运动员优势及问题；获取电商平台用户对商品的评价数据，帮助买家了解商品的口碑等。

中国科学技术出版社
人民教育出版社

根据需求编写爬虫程序可以帮助我们更灵活准确、方便快捷地获取各种数据。前面使用的网络爬虫程序ywrp_crawler.py，主要利用Python的Beautiful Soup库，实现从网页中提取数据。程序运行后，公报中的每一个数据表格被自动存储为一个网页文件，如图3.2.7所示。

图3.2.7　网络爬虫程序运行结果

网络爬虫已经成为数据采集的一种重要方式，得到日益广泛的应用及发展，但并不是所有数据都可以随意爬取和使用。《中华人民共和国网络安全法》及最新刑事司法解释规定“非法提供”和“非法获取”公民个人信息构成犯罪的都可以入刑。因此，在采集数据时，我们必须遵守相关法律法规，负责任地合理使用网络爬虫。

3.2.2　数据整理

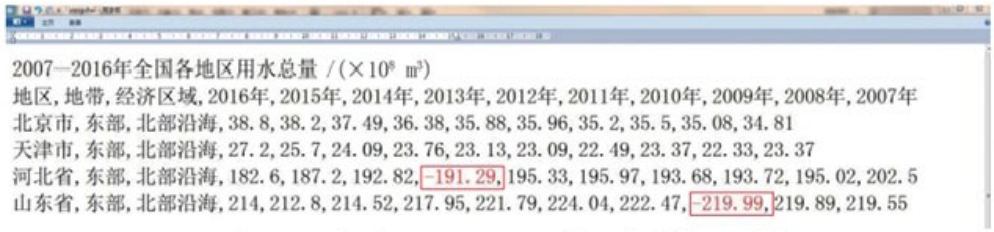
数据整理的目的是对数据进行校验和标准化。采集到的数据可能是非标准化的、不完整或重复的。通常，将这样的数据形象地称作“脏数据”。而用于分析的数据必须保证数据质量，即保证数据的完整性、统一性和准确性。其中，完整性指数据不能有缺失；统一性要求数据符合统一的标准；准确性要求数据不能有错误。

数据整理就是通过去重、补漏和勘误等方法，删除重复数据、补全缺失数据和校正错误数据，并对数据进行统一性和标准化处理，以确保数据是相关和准确的，从而将“脏数据”变为“清洁数据”。

采集到的数据可以保存为不同的文件，如CSV文件等。其中，CSV 文件将数据表格存储为纯文本，每一行代表一条数据，每条数据包含了一个或由逗号分隔的多个值。整理数据可以使用现成的软件或平台，也可以通过编写程序实现。Python语言丰富的标准模块和扩展库提供了许多高效灵活的函数，可以帮助我们较好地进行数据整理。

例：编写程序检测用水量为负值的数据。

问题：采集到的全国各地区2007—2016年用水量数据，存储为“yongshui.csv”文件，如图3.2.8所示。采集的数据由于误输入、传输错误等可能存在问题。例如，河北省2013年用水量中的-191.29和-219.99两个值，不属于正常用水量，需要检测出所有类似的异常数据，并提示用户对数据进行检查和修改。



2007—2016年全国各地区用水总量 /（×10^8 m^3）
地区，地带，经济区域，2016年，2015年，2014年，2013年，2012年，2011年，2010年，2009年，2008年，2007年
北京市，东部，北部沿海，38. 8，38. 2，37. 49，36. 38，35. 88，35. 96，35. 2，35. 5，35. 08，34. 81
天津市，东部，北部沿海，27. 2，25. 7，24. 09，23. 76，23. 13，23. 09，22. 49，23. 37，22. 33，23. 37
河北省，东部，北部沿海，182. 6，187. 2，192. 82，-191. 29，195. 33，195. 97，193. 68，193. 72，195. 02，202. 5
山东省，东部，北部沿海，214，212. 8，214. 52，217. 95，221. 79，224. 04，222. 47，-219. 99，219. 89，219. 55

图3.2.8 “全国各地区2007—2016年用水量数据”示例

（1）分析问题

已知条件：“yongshui.csv”文件中保存了各地区2007—2016年用水总量，如果其中的数据为负值，则认为数据异常；

求解目标：检测出异常数据，提醒用户核对；

已知与未知的关系：可通过检测每个用水量数据是否为负值，发现异常的数据。

（2）规划问题求解流程

首先读入CSV文件中的全部数据；经观察发现，从第3行开始，每行的第4列之后是各年用水量数据，因此从第4列开始逐行依次检查每个数据，判断其值是否为负值；如果为负值，则提示“数据异常，请核对！”。

（3）编程实现与调试

进行编程实现，程序示例如下：

```
# 导入pandas并设置别名为pd
import pandas as pd
# 用pandas中的函数read_csv打开数据文件，指定文件的文字编码方式，指定包含列标题
df = pd.read_csv("yongshui.csv", encoding="gbk", header=1)
# 按年份生成数据序列
for i in range(3, len(df.columns)):
    # 指定当前列
    current_col = df.columns[i]
    # 最终显示的列
    display_cols = ["地区", current_col]
    # 查找符合条件的行
    error_data = df.loc[df[current_col] < 0, display_cols]
    # 通过函数shape获得行数和列数，列数对本程序无用，使用 Python 约定的“_”变量忽略
    error_rows_count, _ = error_data.shape
    # 按年份选择符合条件的行，并显示
    if error_rows_count > 0:
        print(error_data, "数据异常，请核对！")
        print("\n")
```

（4）保存文件，调试运行程序

程序运行结果如图3.2.9所示。

图3.2.9 程序运行结果

实践活动

编写程序删除用水量数据中的重复值

登录教学资源平台，找到“yongshui.csv”文件。由于录入数据时操作失误，导致该文件中多个省市的用水数据被重复输入。

用Python语言编写程序找到这些重复的数据，并删除。

技术支持

Python数据分析库pandas

用Python语言进行数据分析，通常要用到数据分析库pandas。pandas属于Python的第三方扩展库，它提供了许多高效易用的数据分析工具和适合科学计算的数据结构，使得处理更加简洁，同时在处理速度上也做了不少优化。

pandas中的DataFrame是一种类似二维表格的数据结构，它由行和列组织数据，并具有行标签和列标签。如下程序展示了DataFrame的基本使用方法，运行该程序将创建一个DataFrame，并输出其全部数据及其指定行和列的数据。

```
import pandas as pd
# 初始化数据
values = [[37.49, 38.2, 38.8], [24.09, 25.7, 27.2], [192.82, 187.2, 182.6]]
# 初始化行标签
areas = ["北京", "天津", "河北"]
# 初始化列标签
years = ["2014", "2015", "2016"]
# 构建名为df的DataFrame
df = pd.DataFrame(values, areas, years)
# 输出df
print(df)
# 通过行列标签选择df中的一个值并输出
print("北京2015年用水量:", df.loc["北京", "2015"])
```

调用pandas库可实现数据查找、删除、排序和索引等功能，其部分函数/方法如表3.2.3所示。

表3.2.3　pandas库部分函数/方法

函数/方法	功能说明
pandas.read_csv	读取CSV文件
pandas.to_csv	导出数据到CSV文件
pandas.DataFrame.columns	描述DataFrame列
pandas.DataFrame.loc	通过行与列标签选择一个值
pandas.DataFrame.shape	返回DataFrame的行数和列数
pandas.DataFrame.drop_duplicates	删除重复的数据

为了更好地使用数据，整理后的数据可存储在本地或云存储空间。云存储是一种新兴的网络存储技术，它将网络中大量不同类型的存储设备通过应用软件集合起来协同工作，共同对外提供数据存储和业务访问功能。使用者可通过联网设备连接到云存储空间方便地存取数据。

3.2.3　数据安全

从健康数据到智慧交通，再到用水决策，进而到智能交互，数据正深刻地影响着大众生活及国家建设的方方面面。数据已经成为一种重要的资源，保护数据安全显得日益重要。

思考活动

我们的网上数据是否安全

网络的使用已经非常普遍，我们经常在网上学习、购物、交流、娱乐……不少互联网平台会记录用户使用其平台的情况，比如学习网站记录学生学习的课程、学习时长；购物平台记录用户购买的商品、消费的金额；聊天软件记录用户与他人聊天的频率，聊天的话题，等等。

思考：

1. 我们的网上行为数据可能存储在哪里？如果这些数据被泄露可能会带来哪些风险和问题？

2. 小组同学讨论，通过哪些措施可以使我们的网络行为数据更安全？

中国互联网络信息中心发布的第43次《中国互联网络发展状况统计报告》显示，2018年我国网民所遭遇的网络安全事件中，个人数据安全问题占了很大比重。据统计，

2018年，27.3%的网民遭遇过个人信息泄露事件，17.7%的网民遭遇过账号密码被盗事件。国家计算机网络应急技术处理协调中心发布的《2017年中国互联网网络安全报告》指出，2017年我国境内约有2万个网站的数据被篡改，严重影响了正常的工作秩序和社会秩序。

由于数据安全面临的风险有日益加剧的趋势，因此，我们要具备一定的数据保护能力，提高数据安全意识，做好数据安全保障。

数据安全的威胁

数据安全的威胁来自很多方面，既有人为因素，也有非人为因素。其中，主要的威胁包括计算机病毒、黑客攻击、数据存储介质损坏和个人失误等。

计算机病毒能影响计算机软件、硬件的正常运行，破坏数据的正确与完整，甚至导致系统崩溃，对数据安全的威胁很大。一个名称为“勒索”的计算机病毒于2017年席卷全球100多个国家，很多高校、企业和政府机构的计算机遭受严重攻击，很多计算机被感染，给众多的计算机用户造成了严重损失。

黑客攻击主要表现为入侵他人计算机系统、扰乱系统运行、盗窃系统保密信息和破坏目标系统等。2015年，某国知名大型医保企业信息系统被黑客攻破，超过8 000万名客户的详细信息被泄露，其中包括姓名、地址、就业信息、社保号码等敏感信息。

数据存储介质的损坏包括物理损坏、设备故障等。

个人失误也会给数据安全带来威胁。例如，口令设置过于简单易破解，将个人账号随意转借他人等行为都将给数据安全造成威胁。

除了上述威胁以外，非法数据交易也严重威胁着数据安全。例如，某电商员工因其涉嫌盗取并贩卖50多亿条个人信息而被捕。

移动互联网的数据安全问题同样很严重。图3.2.10是国家计算机网络应急技术处理协调中心发布的2007—2017年移动互联网恶意程序数量的走势。从图中可以看到，近年来恶意程序的数量激增，数据安全形势已非常严峻。

图3.2.10　2007—2017年移动互联网恶意程序数量走势

中国地图出版社
人民教育出版社

数据保护的方法

多方面的数据威胁给数据安全带来了极大风险。因此，学习和掌握保护数据的方法，能更好地保障数据的安全。数据安全有两方面的含义，即数据防护的安全和数据本身的安全。我们可以从这两方面入手对数据进行保护。数据防护的安全主要是采用存储手段保证数据的安全，如数据备份；数据本身的安全可采用加密的方式对数据进行保护。

1. 数据备份

数据备份是将需要备份的数据从应用主机的硬盘或磁盘阵列复制到其他的存储介质或不同位置存储空间的过程，其目的是在设备发生故障或发生其他威胁数据安全的灾害后，利用备份进行恢复，从而达到保护数据的目的。数据备份常见的方法有可移动存储设备备份和网络备份等。

云存储将数据存储在网络上的服务器中，是网络备份的一种重要方式，如图3.2.11所示。云存储具有成本低、管理方便和可扩展性高等优势，但也存在安全性不足等缺点。

图3.2.11　云存储示意图

2. 数据加密

数据加密是使用特定算法把敏感的明文数据变换成难以识别的密文数据。数据加密是保护数据传输安全的实用方法，也是保护数据存储安全的有效方法。实际应用中，为防止数据泄露，可以为数据文件设置密码，加密系统利用设定的密码将整个文件进行加密处理。这样没有正确的密码就无法打开文件查看内容。

此外，为保护数据安全，在使用各种信息设备时，应具有数据安全意识，注意以下几方面：

- 连接可信的无线网络；
- 合理配置智能终端的数据采集功能，保护个人隐私数据；
- 连接其他设备时，认真阅读操作提示和安全事项。

数据保护意义重大，不仅要保障个人数据安全，更应保障国家数据安全。国家大数据发展战略提出了切实保障国家数据安全和完善数据产权保护制度的要求，强调要加强关键信息基础设施安全保护，强化国家关键数据资源保护能力。

可以出版社
人民教育出版社

阅读拓展

量子保密通信

2017年世界首条量子保密通信干线——“京沪干线”正式开通，“京沪干线”与“墨子号”量子科学实验卫星的完美对接，打通了天地一体化广域量子通信的链路。利用“墨子号”在中国和奥地利之间实现洲际量子密钥分发，并利用共享密钥实现加密数据传输和视频通信，标志着“墨子号”已具备实现洲际量子保密通信的能力，如图3.2.12所示。

量子通信是迄今唯一被严格证明的一种无条件安全通信方式，可以有效解决信息安全方面的问题。我国将建立完整的量子通信产业链和下一代国家主权信息安全生态系统，最终构建基于量子通信安全保障的量子互联网。

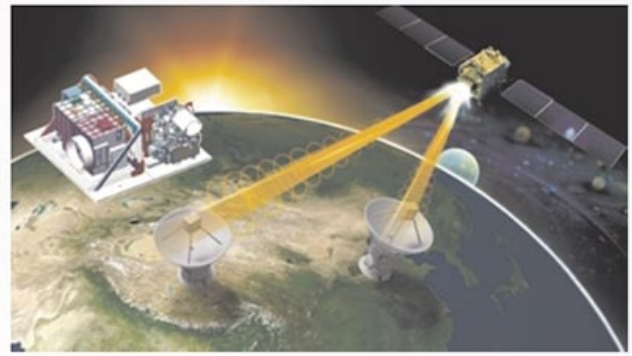


图3.2.12　量子通信业务应用示范

项目实施

采集并整理南水北调相关数据

一、项目活动

1. 根据调查“我国水资源分布情况”的数据需求确定的数据目录，选择合适的方法采集有关数据。

例如，采集北京市水资源数据，可访问北京市水务局网站，下载北京市各年水资源公报。

2. 将收集到的数据按照内容进行分类，填写数据采集记录表3.2.4。

表3.2.4　数据采集记录表

序号	数据内容	采集方法	数据来源
1	北京市水资源公报	网络获取	北京市水务局
2			
⋮			

3. 整理并保存数据。

各组整理采集到的数据，如处理异常数据、删除重复数据等，并将数据保存为CSV格式文件。

二、项目检查

采集整理水资源相关数据，保存为CSV文件。

中国地图出版社
人民教育出版社

练习提升

1. 搜索引擎通常提供高级搜索指令，帮助使用者更有效地进行检索。上网学习搜索引擎使用技巧，了解在指定的网站内进行检索的指令，并使用该指令检索数据。例如，搜索有关南水北调中线工程的内容。

2. 小明同学非常喜欢打篮球，经常登录体育网站了解球队、球员及赛事情况。学校研究性学习课程布置了研究任务，小明选择中国男篮作为研究课题，需要收集中国男篮相关数据，如球员的身体素质情况、球员特长、场上位置及表现等数据。试使用一款网络爬虫软件或某个平台帮助小明采集球员数据，整理并保存数据。

3. 在信息技术课上，学生通过签到系统记录出勤情况，但是经常有学生重复签到，导致签到记录中的学生数据出现重复，学生签到记录如表3.2.5所示。试使用 Python 语言编写程序删除签到记录中的重复数据。

表3.2.5　学生签到记录表

学号	机器号	性别
23	A01	女
12	B05	男
12	B05	男
39	A06	女
3	D06	男
18	D02	男
15	A07	男
42	B01	女
17	C08	男
36	A03	女
22	C05	男
22	C05	男
35	D07	女

4. 小明同学准备参加一门网络课程的学习，但是在注册账号时，系统提示密码安全强度低。试分析安全强度低的原因，并说明小明应如何重新设计密码，写出设计方案和依据。

5. 某些压缩软件不仅具有压缩文件的功能，还可以用密码对文件加密。查找并下载一款带加密功能的压缩软件，试用文件加密功能，并了解其加密原理，写出使用心得。

中国地图出版社
人民教育出版社

3.3
数据分析与可视化

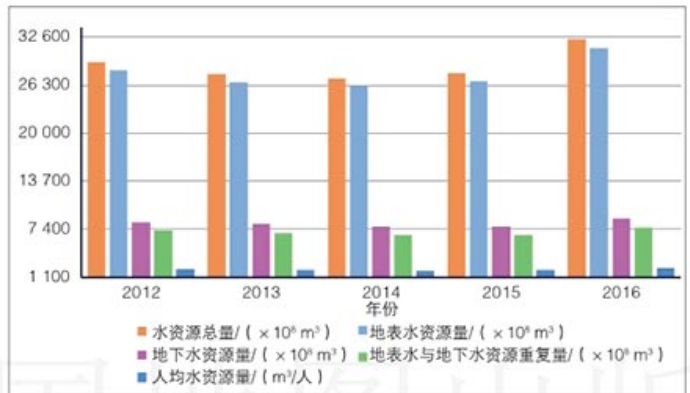
学习目标 ▶▶▶

- 掌握常用的数据分析方法，能根据需要选择恰当的方法进行数据分析。
- 了解数据可视化的形式，能选用合适的工具进行数据可视化表达与呈现。
- 通过项目研究，感受数据分析和可视化表达对日常生活的影响。

体验探索

你能读懂这些数据吗

为了解近年来我国水资源状况，访问国家统计局政府网站获得如图3.3.1所示的数据和图表。



指标	2016年	2015年	2014年	2013年	2012年
水资源总量/(×10⁸ m³)	32 466.40	27 962.60	27 266.90	27 957.86	29 526.88
地表水资源量/(×10⁸ m³)	31 273.90	26 900.80	26 263.91	26 839.47	28 371.35
地下水资源量/(×10⁸ m³)	8 854.80	7 797.00	7 745.03	8 081.11	8 416.12
地表水与地下水资源重复量/(×10⁸ m³)	7 662.30	6 735.20	6 742.04	6 962.75	7 260.64
人均水资源量/(m³/人)	2 354.92	2 039.25	1 998.64	2 059.69	2 186.05

图3.3.1　我国水资源部分数据指标呈现

思考：

1. 利用该网站数据可视化功能，调整水资源数据指标和数据呈现方式，进一步分析不同类型图表所表达信息的异同。

2. 根据自己的需求，浏览网站栏目，了解相应的数据，并分析其中蕴含的信息。

中国图书出版社
人民教育出版社

3.3.1 数据分析

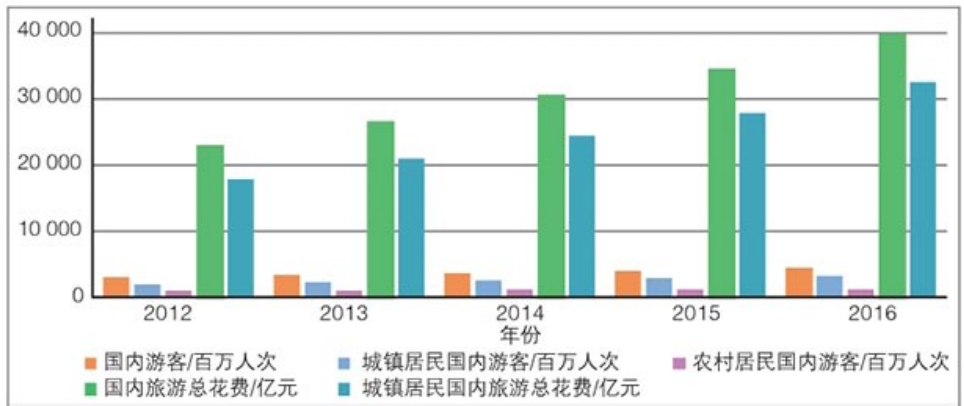
采集了所需要的数据，并对数据进行了组织和整理后，为了获得有价值的信息，还需要对数据做进一步的分析。数据分析指用适当的计算方法与工具对收集来的数据进行处理，提取有用信息，形成结论从而支持决策。

数据分析的作用

数据分析广泛应用于社会各个领域，发挥着日益重要的作用。数据分析可以让我们了解事物的现状、剖析其发展历程和预测未来走向。

1. 了解事物的现状

通过数据分析，可以帮助我们更好地认识所调研事物的目前状况。例如，通过分析游客人数变化、旅游花费等数据，如图3.3.2所示，能够了解旅游业当前整体的发展状况，帮助旅游业经营者有针对性的调整经营策略，制订发展方案。



指标	2016年	2015年	2014年	2013年	2012年
国内游客/百万人次	4 440	4 000	3 611	3 262	2 957
城镇居民国内游客/百万人次	3 195	2 802	2 483	2 186	1 933
农村居民国内游客/百万人次	1 240	1 188	1 128	1 076	1 024
国内旅游总花费/亿元	39 390. 0	34 195. 1	30 311. 9	26 276. 1	22 706. 2
城镇居民国内旅游总花费/亿元	32 241. 3	27 610. 9	24 219. 8	20 692. 6	17 678. 0

图3.3.2　我国国内旅游部分数据呈现（资料来源：国家统计局）

2. 剖析事物的发展历程

通过数据分析，可以剖析事物的发展历程，了解事物的过去，揭示某些现象产生的原因。例如，分析全国27个主要城市近50年来月平均降水量和降水天数，再对比相关城市航班正常率的数据，可以发现航班正常率的走向与月降水量和天数呈相反方向变化的趋势。对航空公司、空中交通管理措施和天气等各方面大量历史数据的分析研究，可以帮助管理者找到影响航班正常运行的干扰因素，从而选择最佳的应对措施。

3. 预测事物的未来走向

通过数据分析，可以预测事物未来可能会发生什么，推断未来的发展趋势，并为制

订相应的目标及策略提供依据。目前利用数据分析进行预测已经应用于各个领域，包括体育赛事预测、用户行为预测、能源消耗预测、交通行为预测和人体健康预测等。例如，通过各个家庭安装的智能电表，对电力的使用状况进行监控，检测出用电模式，从而对用电需求做出预测。

数据分析的基本方法

从统计应用上讲，数据分析可以分为描述性数据分析、探索性数据分析和验证性数据分析。其中，描述性数据分析侧重于对数据的各种特征进行分析，对变量之间的关系进行估计和描述；探索性数据分析主要用于在数据之中发现新的特征；验证性数据分析侧重于验证已有假设的真伪。常用的数据分析方法有对比分析法、平均分析法和结构分析法等。

1. 对比分析法

对比分析法也叫比较分析法，是将两个或两个以上的数据进行比较，分析它们的差异，揭示出这些数据所反映的事物规律的方法，是一种常用的分析方法。

对比分析法通常用于从数量上展示和说明研究对象规模大小、水平高低、速度快慢及各种关系是否协调等。例如，在商品生产中，已完成商品数量与目标数量的差距对比，本月生产的数量与上月生产数量的对比，同公司各部门间的对比等都是对比分析法的典型应用。

对比分析法分为横向和纵向对比两大类。任务完成量与目标量的对比，部门之间、地区之间的对比都是横向对比，而不同时期的比较则为典型的纵向对比。例如，图3.3.3中的全国水资源开发利用综合对比属于横向对比，图3.3.4中的2016年中国创新能力指数数据与上年数据的对比即是纵向对比。

图3.3.3　全国水资源开发利用综合对比

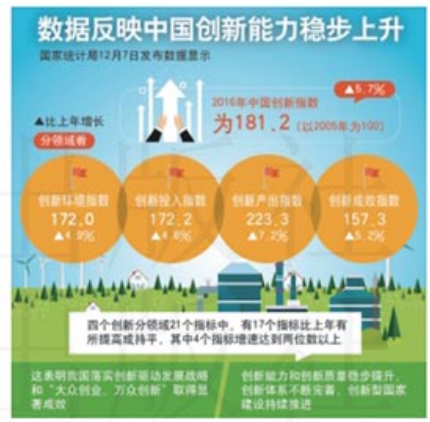


图3.3.4　中国创新能力指数数据

2. 平均分析法

平均分析法是运用计算平均数的方法来反映总体在一定时间、地点等条件下某一数量特征的一般水平。平均指标中最常用的是算术平均数，如平均身高、平均工资和平均降水量等。图3.3.5展示了我国2012—2016年国内旅游人均消费情况。

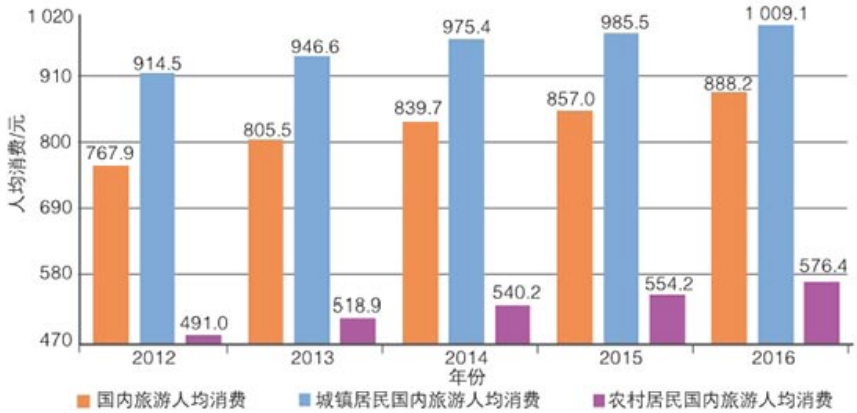


图3.3.5　我国2012—2016年国内旅游人均消费情况（资料来源：国家统计局）

平均分析法多用于比较同类现象在不同地区、不同行业、不同类型单位等之间的差异程度，分析现象之间的依存关系，进行数量上的推算；也可以对某一现象在不同时间的水平进行比较，以说明现象的发展规律及趋势。

3. 结构分析法

结构分析法也称构成分析法，是将各个部分与总体进行对比，是分析事物内部的结构和部分与整体之间关系的方法。结构分析法的基本表现形式就是计算结构指标。

结构指标即各个部分相对于总体所占的百分比，因此总体中各结构指标的总和等于100%。图3.3.6为某网站的用户使用不同无线网络访问该网站的次数统计情况。可以看出使用4G网络访问该网站的占63.19%，使用Wi-Fi的占28.08%，使用3G网络的占8.73%。图中清楚地反映出各部分在总体中所占的百分比。

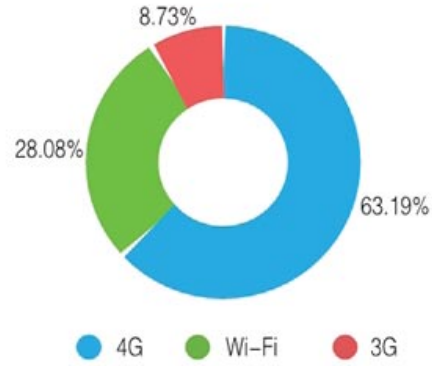


图3.3.6　使用不同无线网络访问某网站的百分比

上述几种基本的数据分析方法分别适用不同的情况，在实际应用中，需要根据具体问题合理选取并综合运用。

体验数据分析

确定了数据分析方法后，我们还需要借助合适的工具才能更快捷地进行数据分析。计算机普及前，数据分析主要依靠人工方式完成。随着计算机的广泛应用，涌现出很多功能强大的数据分析工具，例如电子表格软件、在线数据分析平台和程序设计语言编写的程序等。

1. 用电子表格软件进行数据分析

电子表格软件可对数据进行格式编辑、统计计算和图表处理等。在电子表格软件中，通常用若干工作表存储数据。图3.3.7就是使用电子表格软件处理数据的示意图。

	A	B	C	D	E	F	G	H
1	水资源情况(2006–2016年)　　单位：(×10⁸ m³)							
2	年份	水资源总量	水资源总量变化量	地表水资源量	地下水资源量	地表水与地下水资源重复量	人均水资源量	人均水资源量变化量
3	2006	25 330.10	-9.71%	24 358.10	7 642.90	6 670.80	1 932.1	-10.21%
4	2007	25 255.20	-0.30%	24 242.50	7 617.20	6 604.50	1 916.3	-0.82%
5	2008	27 434.30	8.63%	26 377.00	8 122.00	7 064.70	2 071.1	8.08%
6	2009	24 180.20	-11.86%	23 125.20	7 267.00	6 212.10	1 816.2	-12.31%
7	2010	30 906.40	27.82%	29 797.60	8 417.00	7 308.20	2 310.4	27.21%
8	2011	23 256.70	-24.75%	22 213.60	7 214.50	6 171.40	1 730.2	-25.11%
9	2012	29 526.90	26.96%	28 371.40	8 416.10	7 260.60	2 186.1	26.35%
10	2013	27 957.90	-5.31%	26 839.50	8 081.10	6 962.70	2 059.7	-5.78%
11	2014	27 266.90	-2.47%	26 263.90	7 745.00	6 742.00	1 998.6	-2.97%
12	2015	27 962.60	2.55%	26 900.80	7 797.00	6 735.20	2 039.2	2.03%
13	2016	32 466.40	16.10%	31 273.90	8 854.80	7 662.30		

图3.3.7　电子表格软件处理数据的示意图

2. 用在线数据分析平台进行数据分析

随着数据规模越来越大，应用范围越来越广，分析数据和挖掘数据价值的需求也越来越多，许多在线数据分析平台应运而生。这些平台集成了多种分析功能，提供了简单易用的交互界面，使得多种数据分析需求得以实现。

实践活动

使用在线数据分析平台分析全国供水情况

登录国家统计局网站，查找全国供水数据，体验在线数据分析平台分析数据的过程和方法。

1. 注册账号并登录，熟悉国家统计局网站在线数据分析平台的使用方式；

2. 查找全国供水数据，根据平台的提示，尝试完成在线数据分析，查看并保存分析结果；

3. 分析该平台的数据分析功能，有哪些优缺点。

数据分析除了可以对数值型数据进行统计和描述以外，还可以对文本、图像等数据进行分析。常用的文本分析有字词频率统计、语句分词、句法分析、分类分析和情感分析等。例如，对一段描述南水北调工程的文字进行文本分析后，得到的结果如图3.3.8所示。

图3.3.8　文本分析结果呈现

阅读拓展

政府数据网站

在互联网浪潮的推动下，政府机构向信息化、网络化和智能化转变已成为一种趋势。中央和各级政府机构高度重视信息公开和数据发布，并在其官方网络平台为大众提供数据服务。例如，中华人民共和国国家统计局专门设立了国家数据网站，用于发布我国各级各类统计数据，如图3.3.9所示。国家数据网站不仅提供数据的下载，还可以对数据进行在线处理，如编辑、计算和生成图表等。

图3.3.9　国家数据网站页面

3. 使用编程工具进行数据分析

除了使用现成的工具分析数据，也可以根据需要编写程序进行数据分析。例如，可以使用Python语言编写程序对公共交通运营数据进行分析，找出不同线路的发展特点，针对线路的历史成本，分析线路发展的基本规律，从而找出有用的信息；还可以编写程序分析海量用户行为数据，从而了解用户的真实需求，为用户建立独有的行为档案，提供个性化服务。

例1：编写程序统计分析南水北调中线工程各地区水资源情况。

问题：分析南水北调中线工程各地区水资源数据，有助于了解我国水资源的分布情况。将采集并整理后的中线工程各地区水资源数据保存为“07-16nsbd.csv”文件，如图3.3.10所示。编写程序，计算中线工程各地区2007—2016年水资源总量的平均值，并将结果保存到CSV文件中。

指标：水资源总量/($\times 10^8$ m^3)
地区, 2007年, 2008年, 2009年, 2010年, 2011年, 2012年, 2013年, 2014年, 2015年, 2016年
北京市, 23. 81, 34. 2, 21. 84, 23. 1, 26. 81, 39. 5, 24. 81, 20. 3, 26. 8, 35. 1
天津市, 11. 31, 18. 3, 15. 24, 9. 2, 15. 39, 32. 94, 14. 64, 11. 37, 12. 8, 18. 9
河北省, 119. 79, 161, 141. 16, 138. 9, 157. 15, 235. 53, 175. 86, 106. 16, 135. 1, 208. 3

图3.3.10　整理后的2007—2016年中线工程各地区水资源数据

（1）分析问题

已知条件：已有中线工程各地区水资源数据文件“07-16nsbd.csv”，其中包含了2007—2016年相关地区水资源数据。

求解目标：统计计算中线工程各地区2007—2016年水资源总量的平均值，将统计结果保存为CSV文件“nsbd_avg.csv”，并显示中线工程各地区十年来的平均水资源总量。

已知与未知的关系：可以根据中线工程各地区水资源数据文件“07-16nsbd.csv”中的2007—2016年水资源数据，计算出各地区这十年水资源总量的平均值。

（2）规划问题求解流程

首先导入用到的库，读入“07-16nsbd.csv”文件中的全部数据，然后统计各地区十年水资源总量的平均值，生成一个数据表，并将统计结果写入CSV文件。

（3）编程实现与调试

进行编程实现，程序示例如下：

```
# 导入所需库并设置别名
import pandas as pd
df = pd.read_csv("07-16nsbd.csv", encoding="gbk", header=1)
# 为DataFrame增加"平均水资源总量"列,并计算各地区2007—2016年水资源量的平均值填入该列
df["平均水资源总量"] = df.mean(axis=1)
# 显示各地区十年的平均水资源总量
print(df[["地区", "平均水资源总量"]])
df.to_csv("nsbd_avg.csv", encoding="gbk", float_format="%.3f")
```

（4）保存文件，调试运行程序

程序运行结果如图3.3.11所示。

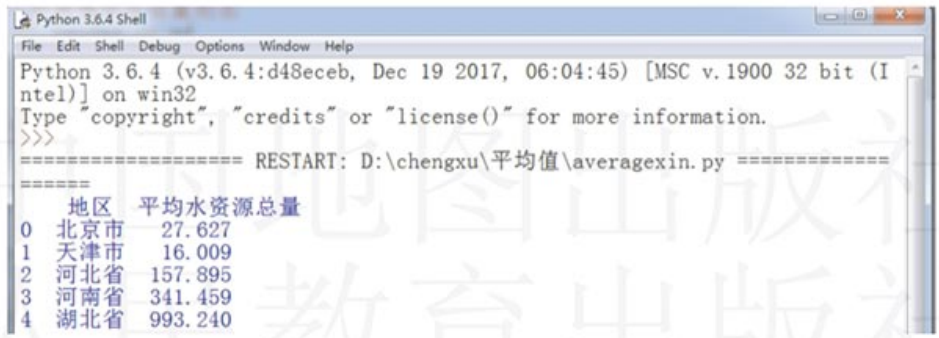


图3.3.11 程序运行结果图

通过以上统计分析，用数据展现了我国部分地区水资源的实际情况：南方地区湖北省水资源较为丰富，而北京市、河北省等北方地区水资源相对匮乏。因此，政府实施南水北调工程意义重大，有助于优化我国水资源配置，缓解北方地区水资源短缺局面，促进经济和社会的可持续发展。

实践活动

编写程序计算京津地区人均水资源量

中国水资源短缺，且时空分布不均，南方水多，北方水少。例如，黄淮海流域水资源严重短缺，人均水资源量约为全国平均水平的21%。请利用前面收集到的全国各地区水资源数据，计算京津地区人均水资源量是全国平均水平的百分之几。

1. 分析问题，写出已知条件和求解目标。
2. 设计算法并编程实现。
3. 依据水资源数据，从某个角度分析实施南水北调工程的必要性。

技术支持

DataFrame的方法

DataFrame具有一组常用的汇总和统计方法，包括计数、求和、求平均值等相关计算。

1. pandas.DataFrame.count

格式：pandas.DataFrame.count( )

用于返回DataFrame的个数。默认按列统计非空数据的个数，axis=1表示按行统计非空数据的个数。

2. pandas.DataFrame.sum

格式：pandas.DataFrame.sum( )

用于返回所有列的和。axis=1表示对各行的数据计算和。

3. pandas.DataFrame.mean

格式：pandas.DataFrame.mean( )

用于返回所有列的平均值。axis=1表示对各行的数据计算平均值。

上述方法的具体应用举例如下：

```
import pandas as pd
values = [[38.8, 38.2], [27.2, 25.7]]
areas = ["北京", "天津"]
years = ["2016", "2015"]
df = pd.DataFrame(values, areas, years)
print ("DataFrame中的数据为:", "\n", df, "\n")
df1 = df.sum()
print("按列求和结果为: ", "\n", df1)
```

该程序将计算DataFrame各列的和，运行结果如图3.3.12所示。

中国铁道出版社
人民教育出版社

```
Python 3.6.4 Shell
File Edit Shell Debug Options Window Help
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:04:45) [MSC v.1900 32 bit (Intel)]
 on win32
Type "copyright", "credits" or "license()" for more information.
>>>
======================== RESTART: D:\chengxu\dfsum.py ========================
DataFrame中的数据为:
      2016  2015
北京  38.8  38.2
天津  27.2  25.7

按列求和结果为:
2016    66.0
2015    63.9
```

图3.3.12　DataFrame按列求和

若添加axis=1，将程序倒数第二行修改为df1 = df.sum(axis = 1)，最后一行中的“按列求和”修改为“按行求和”，程序将计算DataFrame各行的和，运行结果如图3.3.13所示。

```
Python 3.6.4 Shell
File Edit Shell Debug Options Window Help
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:04:45) [MSC v.1900 32 bit (Intel)]
 on win32
Type "copyright", "credits" or "license()" for more information.
>>>
======================== RESTART: D:\chengxu\dfsum.py ========================
DataFrame中的数据为:
      2016  2015
北京  38.8  38.2
天津  27.2  25.7

按行求和结果为:
北京    77.0
天津    52.9
```

图3.3.13　DataFrame按行求和

3.3.2　数据可视化

数据可视化是以图形、图像和动画等方式直观生动地呈现数据及数据分析结果，揭示数据之间的关系、趋势和规律等，便于人们更好地理解数据。可视化的作用体现在多个方面，如观察事物变化的趋势、揭示想法和关系、总结或汇聚数据、形成论点或意见等。例如，国家统计局网站中以可视化的形式呈现的全国交通运输状况，如图3.3.14所示。其中，各柱形的高低呈现了各年的货物运输总量情况，折线点的变化反映出运输量增长速度在几年内的变化趋势。

图3.3.14　全国交通运输状况

数据可视化的形式

数据可视化的形式丰富多样，常见的有图表、词云等。例如，图3.3.15形象地展示了某公交线路的路线及客流情况。图中的蓝色线条显示了公交车行进的路线，右侧的柱形图显示了各站点乘车的人数，折线图显示出客流随时间的变化情况。

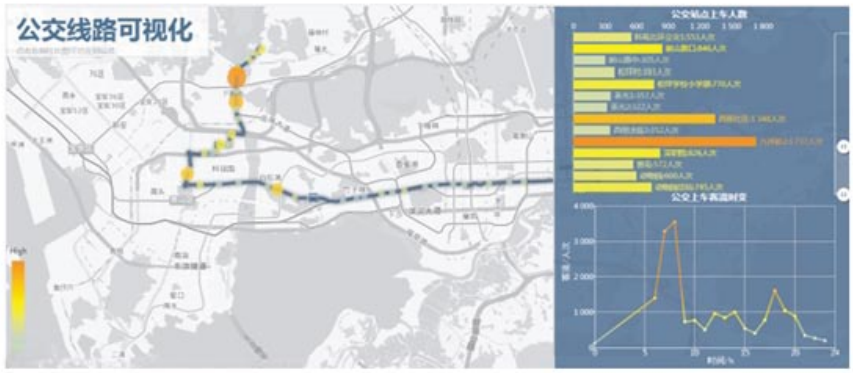


图3.3.15　公交线路及客流情况的可视化呈现

实践活动

了解数据可视化

了解各种数据可视化形式，体验数据可视化效果和魅力。

1. 访问国家统计局网站，了解数据可视化在国家数据统计分析领域的应用（如图3.3.16所示），体验数据可视化效果。

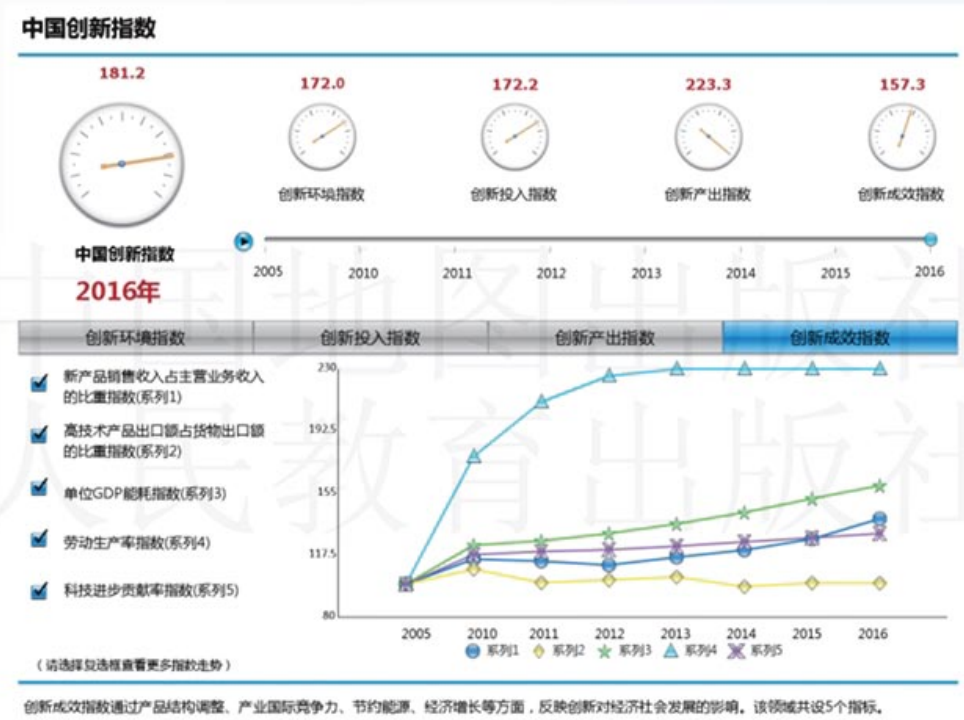


图3.3.16　可视化示例

2. 上网查找数据可视化的网站，熟悉各种数据可视化形式。

3. 举例说明数据可视化效果，填入表3.3.1中。

表3.3.1　数据可视化效果

可视化形式	示例	包含的信息
图表		
词云		
⋮		

4. 分析上述数据可视化效果可以使用什么工具来实现。

数据可视化的实现

数据可视化帮助人们形象地获取其中的信息。我们可以使用多种形式对数据进行可视化。

1. 用图表将数据可视化

图表是数据处理中常用的数据可视化形式之一。常用的图表包括折线图、柱形图、饼图、散点图和雷达图等。各种图表呈现数据的形式各有特点，应该根据实际需要选择不同的图表进行数据呈现。

■ 折线图：折线图可显示随时间而变化的连续数据，常用于分析相等时间间隔下数据的发展趋势。图 3.3.17 呈现了 2017 年2月—2018年1月鲜果类居民消费价格的变化趋势。

图3.3.17　折线图示例

中国地图出版社
人民教育出版社

■ 柱形图：通常用于显示一定范围内数据的变化情况或用于各项数据的比较。例如，图3.3.18可以对比2012—2016年全国研究生招生人数。

图3.3.18　柱形图示例

■ 饼图：用于显示各部分数据在总数据中的大小和比例关系。例如，图3.3.19呈现了2016年全国各类用水总量在用水总量中所占的比例。

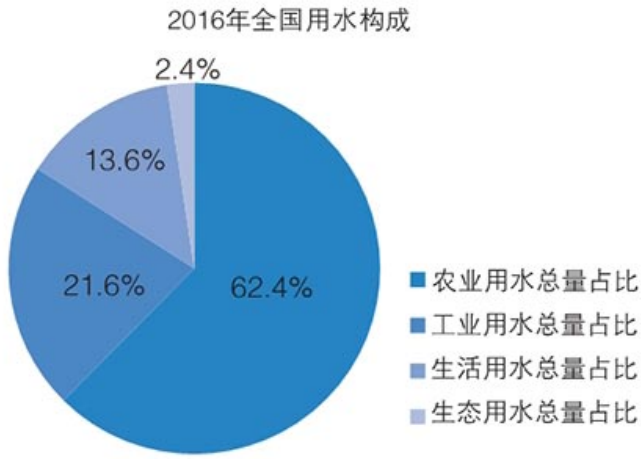


图3.3.19　饼图示例

■ 散点图：用于表示若干数据系列中各数值之间的关系，以便判断两个变量之间是否存在某种关联。适用于三维数据集，但其中只有两维需要比较。图3.3.20展现了4种产品每月销售量数据之间的关系。

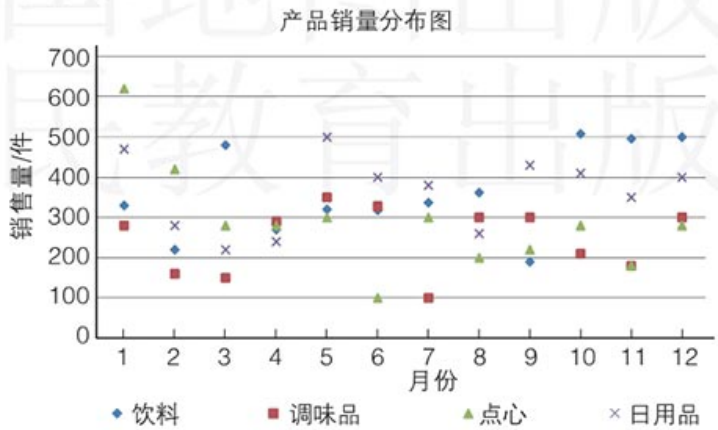


图3.3.20　散点图示例

中国地图出版社
人民教育出版社

■ 雷达图：用来比较每个数据相对中心的数值变化，适用于多维数据的呈现，且每个维度可以排序。例如，图3.3.21显示了用户1和用户2在购买频率、购买商品种类等5个维度上的情况。

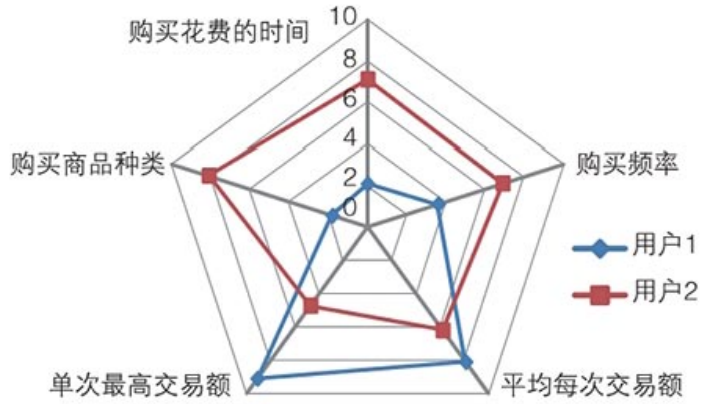


图3.3.21　雷达图示例

使用Python语言可以灵活地制作出以上图表，绘制时可以使用Matplotlib库。该库是提供数据绘图功能的第三方扩展库，其pyplot子库主要用于实现各种数据图表的绘制。绘图时可使用import matplotlib.pyplot as plt语句导入该库，在后续的代码中，则用plt代替pyplot。

例如，使用pyplot子库中的画图函数plot(x,y)可以绘制折线图，程序如下：

```
import matplotlib.pyplot as plt
x=[1, 2.2, 3, 4, 5, 6, 7]
y=[1, 7, 25, 11, 8, 10, 16]
# 绘制折线图
plt.plot(x, y)
# 显示图表
plt.show()
```

该程序运行的效果如图3.3.22所示。

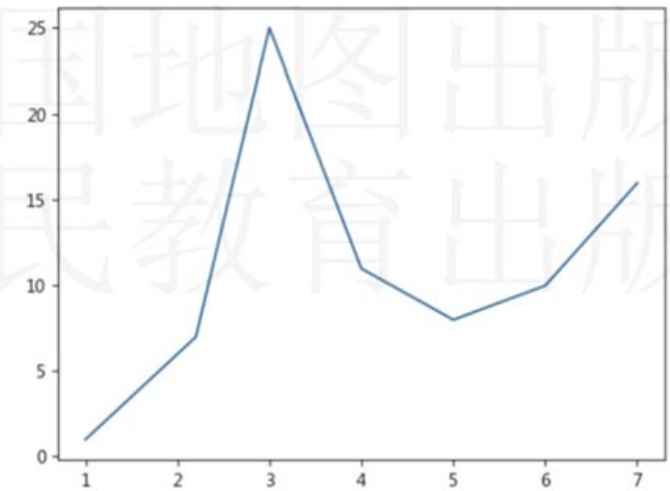


图3.3.22　用函数plot()绘制的折线图

中国地图出版社
人民教育出版社

若将plot(x, y)修改为plot(x, y, "o")，则可绘制出散点图（函数中的"o"表示的是点形标记符），程序运行效果如图3.3.23所示。

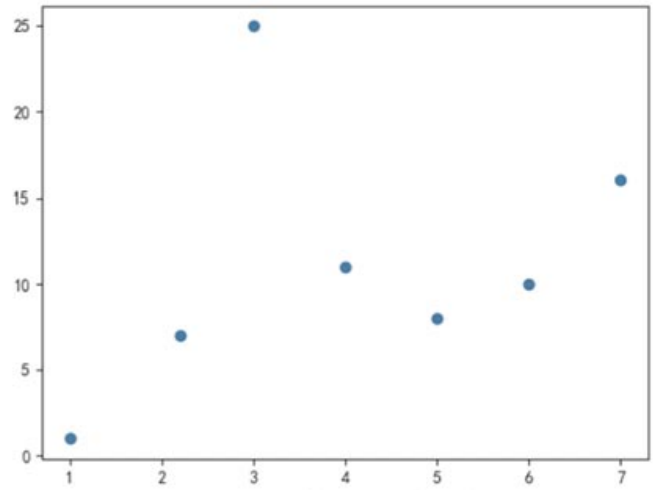


图3.3.23　用函数plot()绘制的散点图

使用pylot子库我们还可以绘制其他图表，也可以为图表添加需要的文字，比如标题、横纵坐标轴的文字说明等。

例2：编写程序用图表可视化呈现南水北调中线工程各地区水资源总量数据。

问题：前面，我们已经计算出中线工程各地区水资源总量的10年平均值，为便于直观地对比分析中线工程各地区水资源情况，需编写程序绘制适当的图表进行呈现。

（1）分析问题

已有需要进行可视化的水资源数据，即供水区湖北省数据，受水区河南省、河北省、北京市和天津市数据，需制作图表呈现数据。为了便于对比各区域数据，选择用柱形图呈现。

（2）规划问题求解流程

首先导入绘制图表需要用到的Matplotlib库，然后从“nsbd_avg.csv”文件中读取数据，再绘制图表，设置图表标题和横、纵坐标轴，最后显示出柱形图。

（3）编程实现与调试

进行编程实现，程序示例如下：

```
# 导入需要的库
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.rcParams["font.sans-serif"] = ["SimHei"]   # 指定Matplotlib库使用的中文字体

# 读取数据文件
# 用pandas库中的函数read_csv打开数据文件,指定文件的文字编码方式,指定不包含列标题
```

中国地图出版社
人民教育出版社

```
df = pd.read_csv("nsbd_avg.csv", encoding="gbk", header=0)
water_resource = df[df.columns[12]] # 取各地区的平均水资源量做柱形图的数据序列
# 绘制图表
ind = np.arange(5)                  # 指定柱形图列数
width = 0.35                        # 指定柱形图的宽度
plt.bar(ind, water_resource, width, color="#d62728") # 按顺序绘制每一列的数据
plt.ylabel("/亿立方米")              # 设置纵轴标签
plt.title("2007—2016年南水北调中线各地区平均水资源总量对比")  # 设置图表标题
# 设置横轴的刻度
plt.xticks(ind, ("北京市", "天津市", "河北省", "河南省", "湖北省"))

# 显示图表
plt.show()
```

（4）保存文件，调试运行程序

程序运行结果如图 3.3.24 所示。

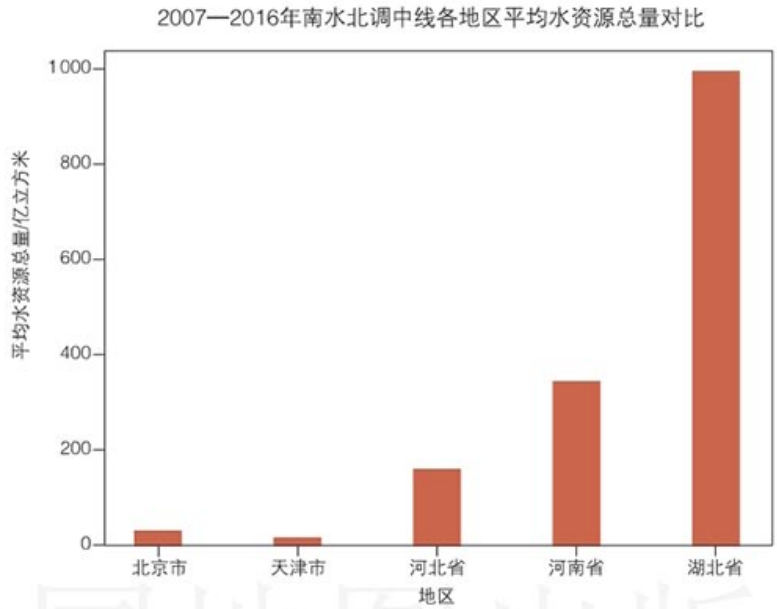


图 3.3.24　程序运行结果图

通过对比图 3.3.24 中的柱形图，能够直观地看到供水区湖北省水资源远远多于受水区各省市。使用类似的方法，我们可以选择适当的图表可视化南水北调工程的各种数据。

技术支持

pyplot 子库中的函数

利用 pyplot 子库中的函数可以创建图表、绘图区域、画线和添加标签等。表 3.3.2 列出了 pyplot 子库中绘制图表常用的部分函数。

中国地图出版社
人民教育出版社

表3.3.2　pyplot子库中部分函数

函数	功能说明
pyplot.title	设置图表标题
pyplot.bar	绘制柱形图
pyplot.xlabel, pyplot.ylabel	设置横轴和纵轴标签
pyplot.xticks, pyplot.yticks	设置横、纵坐标轴的刻度
pyplot.show	弹出窗口显示图表

阅读拓展

地理信息可视化

我们生活在三维空间中，来自现实世界的数据通常包含位置信息。地理空间特指真实的人类生活的空间，地理空间数据一直是可视化研究和应用的重要对象，地理信息可视化在现实生活中应用广泛。

人类长期以来通过对地球和自然环境的观测来研究和了解自然空间，科学家们也通过建立数学模型来模拟环境的变化。这些观测和模拟得到的数据通常包含了地理空间中的位置信息，最常见的是与气象相关的数据。例如，图3.3.25是某一时刻全球大气环流和风速的可视化在线动态展示，能够清晰地反映全球的大气环流状况。当有台风形成时，在地图上能清楚地看到台风的位置。

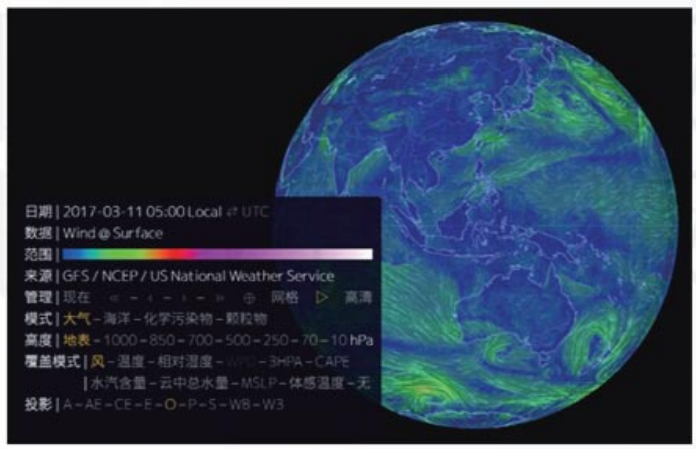


图3.3.25　地球大气环流可视化地图

2. 用词云将数据可视化

日常生活中需要处理大量的文本数据，如新闻、邮件、报告等。面对日益加快的工作节奏，人们需要更高效的文本阅读和分析方法。文本可视化通过丰富的图形或图像，以易于理解和接受的方式揭示文本中的信息，因而得到广泛应用。

词云是目前常用的关键词可视化形式，它能直接抽取文本中的关键词，并将其按照一定顺序和规律整齐美观地呈现在屏幕上。关键词是从文本的文字描述中提取的语义单元，可反映出文本内容的重点。用词云可视化文本数据可以帮助人们快速地了解文本的内容和特征等信息。

图3.3.26　词云示例

词云通常使用字体的大小和颜色表示关键词的重要程度或出现频次。图3.3.26是为50首唐诗制作的词云，字越大表示该关键词使用频率越高。从图中可以看出，50首唐诗中使用频率比较高的词有长安、门前、子夜等。

例3：编写程序制作词云。

问题：《2016年中国水资源公报》中从水资源量、蓄水动态、水资源开发利用和水资源质量4个方面全面描述了我国2016年水资源的整体情况。编写程序对《2016年中国水资源公报》中的文字进行词频分析，并生成词云。

（1）分析问题

已知条件：已有水资源公报文本文件“2016年中国水资源公报.txt”，制作词云需要用到SciPy、wordcloud、jieba和Matplotlib等第三方扩展库；

求解目标：为“2016年中国水资源公报.txt”文件制作词云；

已知与未知的关系：利用第三方扩展库可以对已有的文本文件进行分词，生成并绘制词云。

（2）规划问题求解流程

首先导入用于制作词云的第三方扩展库，再读入报告的文本文件，然后进行分词，随后配置制作词云的各项参数，如字体、背景色和词云形状等，最后生成词云图片。

（3）编程实现与调试

进行编程实现，程序示例如下：

中国地图出版社
人民教育出版社

```
# 导入需要用到的库
from scipy.misc import imread
import matplotlib.pyplot as plt
from wordcloud import WordCloud
import jieba
text = open("2016年中国水资源公报.txt", "rb").read()      # 读入 txt 文件
text_jieba = " ".join(jieba.cut(text))                    # 使用 jieba 分词
bg_pic = imread("1.png")                                  # 读入图片
# 配置词云参数
wc = WordCloud(
    font_path = "msyhbd.ttf",                             # 设置字体
    background_color = "white",                           # 设置背景色
    max_words = 200,                                      # 允许最大词汇
    mask = bg_pic,                                        # 词云形状
    max_font_size = 100                                   # 最大号字体
)
wc.generate(text_jieba)                                   # 生成词云
# 生成图片并显示
plt.figure()
plt.imshow(wc)
plt.axis("off")
plt.show()
wc.to_file("水资源.jpg")                                  # 保存图片
```

（4）保存文件，调试运行程序

制作完成的词云效果如图3.3.27所示。

图3.3.27 《2016年中国水资源公报》词云

中国地图出版社
人民教育出版社

实践活动

编写程序为《新一代人工智能发展规划》制作词云

人工智能的迅速发展将深刻改变人类的社会生活、改变世界。为把握住人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国，2017年我国政府发布了《新一代人工智能发展规划》。为快速了解规划内容和重点，请编写程序为其制作词云并对词云的形状和颜色进行创意设计。

1. 分析问题，写出已知条件和求解目标。
2. 设计算法并编程实现。

技术支持　制作词云使用的库和函数

1. 制作词云用到的Python扩展库

SciPy：科学计算核心库之一。SciPy的misc模块用于对图像进行操作，函数imread()读入用于生成词云的图像文件。

jieba：中文分词工具，使用cut方法进行分词。

wordcloud：词云生成工具，使用generate方法读取文本生成词云。

导入上述库的语句如下：

```
from scipy.misc import imread
from wordcloud import WordCloud
import jieba
```

2. 涉及的pyplot子库函数

使用pyplot子库的函数显示词云图片，各函数功能参见表3.3.3。

表3.3.3　pyplot子库中显示词云的函数

函数	功能说明
pyplot.figure	创建一个绘图区
pyplot.imshow	显示图片
pyplot.axis	设置坐标轴，axis("off")为不显示坐标轴

项目实施　完成项目数据分析及可视化

一、项目活动

1. 根据各组确定的任务主题，确定数据分析的方法和可视化的形式及工具，填写表3.3.4。

表3.3.4　数据分析方法和可视化形式及工具

要分析的问题	采用的数据分析方法、工具	采用的可视化形式、工具
供水区和受水区10年平均水资源总量对比	对比分析法、平均分析法	柱形图、编程语言
⋮		

2. 各组对收集到的数据进行分析和可视化处理，为制作数据分析报告做好准备。

二、项目检查

各组完成数据分析和可视化任务，梳理数据分析过程，为制作数据分析报告准备好资料。

人民教育出版社

练习提升

1. 现有5名篮球队员的比赛数据，如表3.3.5所示。试分析比较他们在各项能力上的差异。思考：哪种图表形式适合呈现队员的数据，比较他们在各个能力维度上的强弱。

表3.3.5　篮球队员比赛数据

球员	得分/分	篮板/个	助攻/个	抢断/个	盖帽/个
方　欣	6.2	5.7	1.1	0.7	0.5
李　立	22.5	9.1	1.5	1.3	1.2
张松涛	4.6	1.7	2.4	0.7	0.1
王博文	3.0	1.6	0.2	0.3	0.1
张　凯	2.5	1.7	0.6	0.4	0.1

2. 案例分析。

中国天气网对天气大数据进行分析，研究我国多云天气对太阳能发电量的影响。报告部分内容如图3.3.28所示。该报告通过分析云量、日照时长等数据，发现北京等9地阴雨天气导致空中云量明显增加，造成太阳能光伏发电量大幅下降。登录中国天气网查阅相关报告，思考以下问题并填写表3.3.6。

① 相关报告分析的数据有哪些，使用了哪种数据分析方法，采用该方法的原因；

② 采用了哪些可视化形式呈现分析结果，使用该形式的原因。

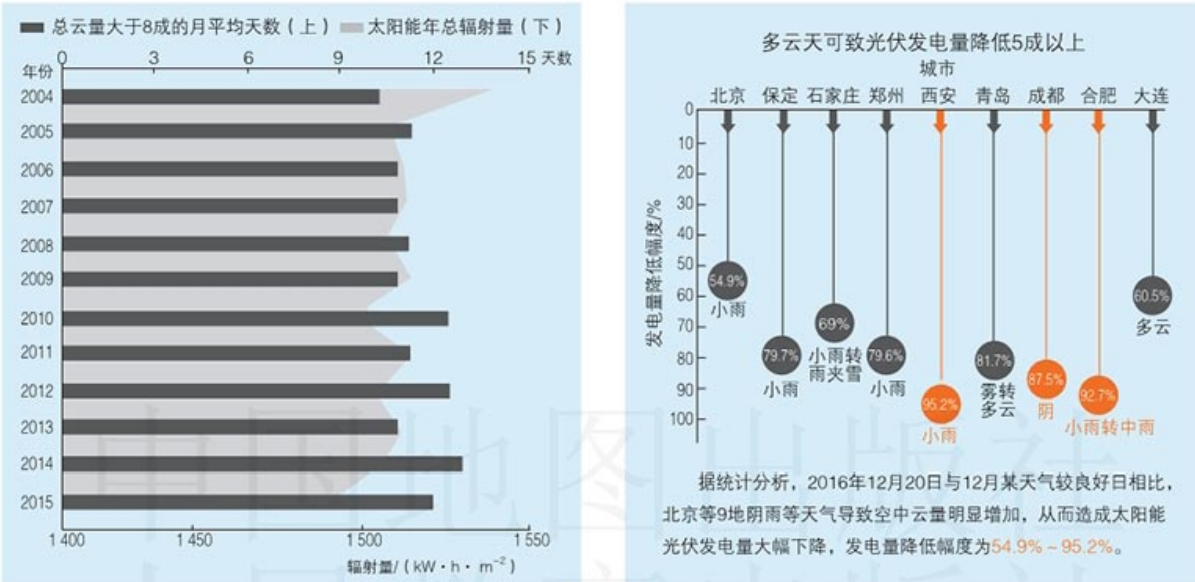


图3.3.28　中国天气网数据报告部分内容

表3.3.6　案例分析表

分析的数据	数据分析方法	使用该方法的原因	可视化形式	使用该形式的原因

3.4
数据分析报告与应用

学习目标

- 掌握数据分析报告的撰写方法，完成数据分析报告。
- 提高处理与应用数据的能力，体会数据应用的重要作用。

体验探索

撰写数据分析报告

访问国家统计局和水利部等网站（图3.4.1）。查阅网站中数据报告的体例和内容，分析这些报告撰写的特点与方法。

图3.4.1　水利部网站统计公报页面

思考：

1. 研究这些报告，你认为数据分析报告应该包含哪些内容？如何呈现数据分析结果？

2. 有哪些数字化工具可以辅助撰写数据分析报告？

中国地图出版社
人民教育出版社

3.4.1 数据分析报告

通过数据采集、整理、分析和可视化等工作，我们完成了对数据的处理，得到的数据分析结果需要以数据分析报告的形式发布出来。数据分析报告不仅是对整个数据处理过程的总结与展示，更能为决策提供参考。通过数据分析报告，可以将分析结果、可行性建议及其他有价值的信息传递给使用者。

数据分析报告的基本结构

数据分析报告内容的呈现形式可以各不相同，但数据分析报告结构形式基本相似，一般都包括开篇、正文和结尾三部分。通常开篇部分包括标题、目录和前言，主要内容是分析背景、目的和思路等；正文部分主要包括具体的分析过程和结果；结尾部分包括结论、建议及附录，如图3.4.2所示。

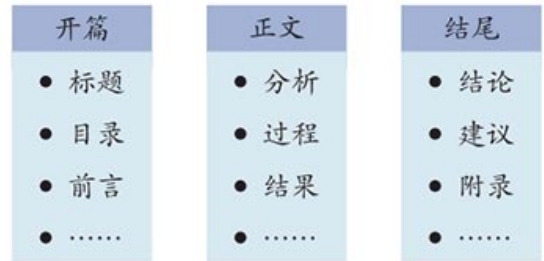


图3.4.2　数据分析报告基本结构

数据分析报告的基本组成部分

1. 标题页

标题页包含报告基本信息：标题、作者、报告日期。标题应短小精悍，体现数据分析的核心内容。

2. 前言

前言是对报告的说明和概述，主要包括以下几部分内容：

分析背景，包括开展此次分析的缘由及意义等；

分析目的，也就是此次分析要解决的问题与要达到的目标；

分析思路，包括数据分析的过程、方法与工具等。

3. 正文

正文是报告的主体部分，包含所有数据分析事实和观点，要系统地展示数据分析的过程和结果。正文各部分之间应具有清晰的逻辑关系，条理清楚，可通过数据图表和相关的文字结合进行分析，帮助使用者理解。

4. 结论和建议

报告结尾部分应根据数据分析结果进行总结和提炼，给出明确的结论、具体建议或解决方案。这部分是整个报告的重点，能够帮助使用者深入思考。

5. 附录

附录是对全文的补充，可以将不便于在正文中体现的内容放在这里。包括名词解释、数据收集方式、调查问卷、分析方法说明和参考资料等。

中国地图出版社
人民教育出版社

一份优秀的数据分析报告，应结构清晰、主次分明、图文并茂，有助于使用者更形象、直观地理解问题和结论。

阅读拓展

撰写数据分析报告应注意的问题

1. 考虑报告使用者

撰写报告时，需要考虑使用者的知识水平、理解能力和他们可能在什么环境下阅读，以及他们会如何使用报告。

2. 行文流畅，通俗易懂，简明扼要

报告中的内容之间要有逻辑性，要注意衔接，要帮助使用者读懂报告各部分内容的内在联系。

3. 内容客观，分析准确

分析报告要用事实说话。说明分析问题和阐述结论要客观准确，合理利用图表等形式使表达的内容清晰、直观。

4. 提出建议与对策

建议是提出的措施、方案或行动策略，应当积极且具有建设性。

5. 引用资料要加注释

分析报告中若引用他人资料，应详细说明资料的来源。这既是对他人成果的尊重，也可供使用者查证。

数据分析成果的展示与交流

数据分析报告完成后，如需要展示交流，建议从以下几方面进行陈述：

项目活动背景，包括确定本项目的原因、本项目的意义和价值等；

数据处理过程，包括数据采集的过程，采集、分析数据所使用的方法和工具，呈现数据使用的可视化形式和工具等；

数据分析成果，包括数据分析的结果、结论和建议等；

项目活动体会，包括项目活动中的感悟和体验，对数据处理重要意义的理解和认识等。

3.4.2 数据处理与问题解决

前面涉及的体质数据、交通数据和水资源数据等的处理，其实都源于实际应用的需求。在日常生活中，我们要善于发现数据，善于利用数据解决问题，让数据成为学习、工作和生活的好帮手。

思考活动

应用数据助力生活

假期小明邀请朋友来北京旅游。为了做好接待，小明除了要提前安排朋友的食宿出行、游览观光等活动，还需要了解各方面的信息，制订出行计划。通过北京市政务数据资源网，他找到了许多有用的数据，做出了详细的计划。图3.4.3是小明访问的部分北京市政务数据资源网页面。

图3.4.3　北京市政务数据资源网页面

思考：

浏览北京市政务数据资源网，仔细观察网站上包含哪些数据，要做好接待方案，需要用到哪些数据，为什么使用这些数据？将思考结果填入表3.4.1。

表3.4.1　数据应用记录

接待方案内容	应用的数据	应用目的
住宿	饭店房价、服务质量、位置等	选择合适的住宿
⋮		

中国地图出版社
北京教育出版社

上述活动中应用了北京市政务数据资源网中的部分数据，该网站还为公众提供了很多数据。比如，“医疗健康数据”中提供了北京市医疗机构、紧急救助和医药保健等数据，可以为人们看病就医提供帮助；“生活服务数据”中提供了家庭生活、日常生活和车辆驾驶等数据，为人们的生活提供了便利。

数据不仅对个人生活影响深刻，更对国家发展发挥着重要的作用。例如，国家统计局提供的大量数据，涵盖了国家各个主要行业和部门，应用这些数据，可以对国民经济、社会发展、科技进步和资源环境等情况进行分析、预测和监督，从而为国家有关部门提供决策依据和咨询建议。国家组织开展的全国住户调查，通过收集居民家庭的收入支出、

生产就业、教育医疗、社会保障和居住条件等数据，为促进城乡居民增收、制定社会保障标准提供了数据参考。

项目实施

完成数据分析报告

一、项目活动

1. 各组梳理总结数据分析结果，根据项目研究主题，撰写本组的数据分析报告。

2. 各组将数据分析报告制作成演示文稿或其他形式的作品在全班展示交流。

二、项目检查

各小组完成数据分析报告，并进行成果展示。

练习提升

1. 记录自己每天的作息时间，如表3.4.2所示。连续记录一个月，月末时对表中的数据进行统计，并生成可视化的作息时间统计图。通过图表找出自己的作息规律，掌握个人时间利用的情况，为更加合理安排生活和学习，提高效率提供帮助和指导。

表3.4.2　作息时间统计表

序号	日期	开始时间	结束时间	内容
1	3月1日	6:30	7:00	起床、洗漱
2				
⋮				

2. 参考表3.4.3的样式收集同学数据，从性别、身高、体重等方面进行体质健康分析，思考在应用数据时应注意哪些事项。

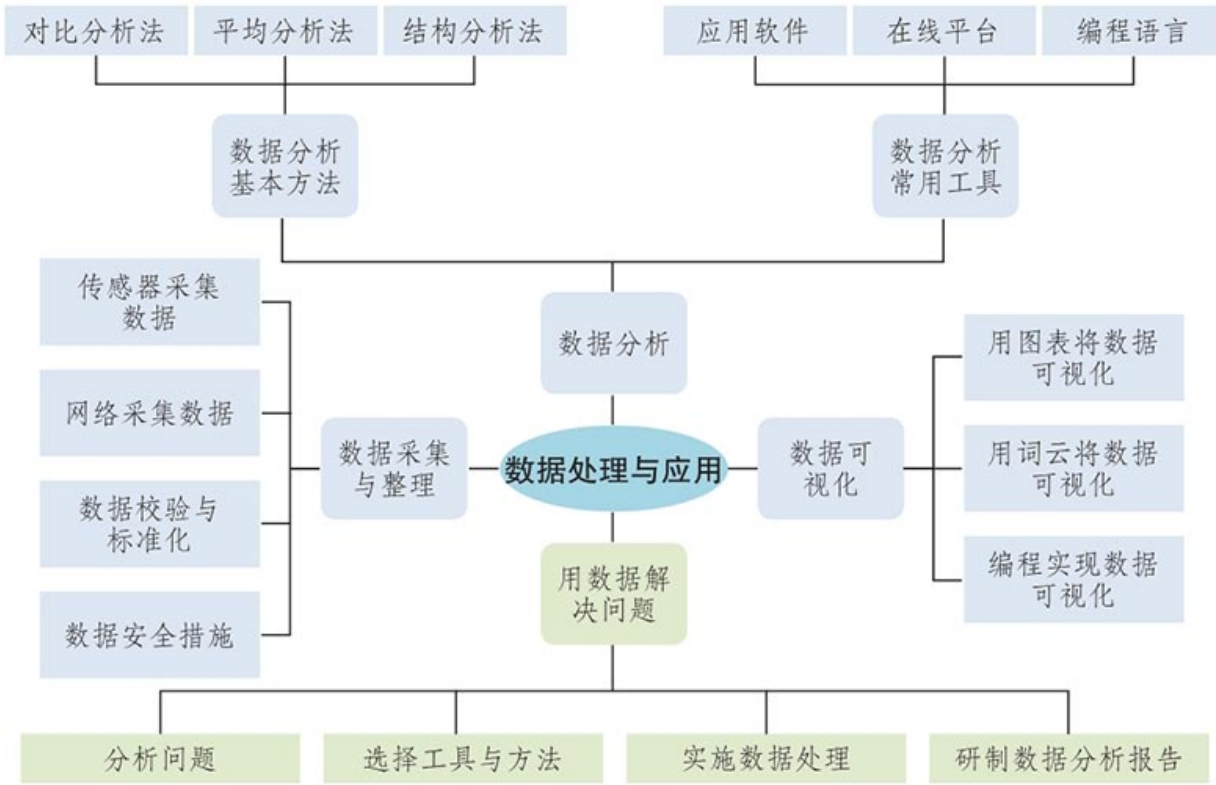
表3.4.3　数据样式表

学号	性别	年龄	身高/m	体重/kg
1	女	16	1.70	60.5
2	男	17	1.82	73.0
⋮				

中国地图出版社
人民教育出版社

总结与评价

1. 下图展示了本章的核心概念与关键能力，请同学们对照图中的内容进行总结。



2. 根据自己的掌握情况填写下表。

学习内容	掌握程度
传感器采集数据	□不了解 □了解 □理解
网络获取数据	□不了解 □了解 □理解
数据整理的方法	□不了解 □了解 □理解
数据备份的方法	□不了解 □了解 □理解
设置密码的规则	□不了解 □了解 □理解
数据分析的方法	□不了解 □了解 □理解
数据可视化方法	□不了解 □了解 □理解
数据分析报告	□不了解 □了解 □理解

中国地图出版社
人民教育出版社

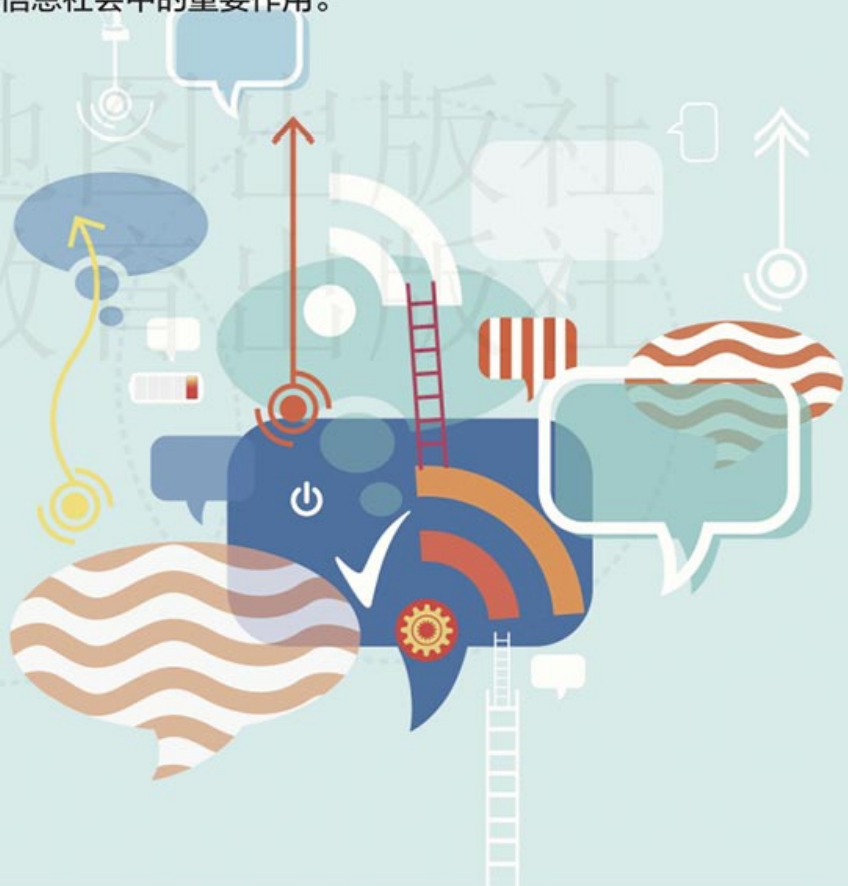
第 4 章
走进智能时代

在 2018 年的上海科技博览会上，有一位声音甜美、行动自如的机器人“小 i”，受到了人们的关注。“你会唱歌吗？”“可不可以给我讲个笑话？”“明天的天气怎么样？”人们争抢着和这个可爱的机器人对话。小 i 不仅逐一回答了它所“听”到的每一个问题，还不时地用一个个笑话、萌劲十足的表情，以及语气逗乐了现场的观众。其实，它是我国一家智能科技公司开发的一款具有自我学习能力的实体机器人。该科技公司在自然语言处理、语义分析和理解、知识工程和智能大数据等方面走在了世界前列，它研发的相关智能技术已经在我国的金融、医疗、交通等领域得到广泛应用。

同学们，想想还有哪些人工智能技术就在我们的生活中？它有哪些奇妙之处？又如何影响着我们的生活？现在，就让我们一起走进这神奇的人工智能（artificial intelligence，简称 AI）世界吧！

本章我们将以“智能交互益拓展”为主题开展项目活动，体验人工智能对日常生活的影响，了解人工智能的关键技术，认识人工智能在信息社会中的重要作用。

中国地图出版社
人民教育出版社



4 主题学习项目：智能交互益拓展

项目目标

本章将以“智能交互益拓展”为主题开展项目学习，感受人工智能的魅力，学习应用智能工具拓展人们的感知与解决问题的能力。为微信公众号设置智能对话机器人，使其成为一个“智能班级交互系统”，通过对交互数据的分析，拓展学习和交流的方法。最后，提交作品展示交流。

1. 与智能机器人进行对话，体验人机对话的过程，了解人工智能的产生与发展，感受其魅力。

2. 学习人工智能平台中智能工具的使用，尝试利用智能工具解决实际问题。

3. 分析“智能班级交互系统”中的数据，撰写支持学习改进的方案。

项目准备

为完成项目，需做如下准备。

- 全班分成若干小组，建议每组3～5人，明确目标与分工，制订项目的实施方案。
- 注册微信公众号和智能平台账号，搭建数字化学习环境。

项目过程

在学习本章内容的同时开展项目活动。为了保证本项目的顺利完成，要在以下各阶段检查项目的进度。

1 体验人机对话

在带有智能机器人的微信公众号中与机器人对话，体验人机对话过程。

P139

2 开发交互系统

利用智能平台上的智能工具开发“智能班级交互系统”。

P145

3 撰写拓展方案

分析“智能班级交互系统”中的数据，撰写支持学习改进方案。

P150

项目总结

完成“智能班级交互系统”的开发和学习拓展方案后，交流体验应用感受；探讨人工智能在信息社会中的作用和影响，提高智能时代的社会责任感。

4.1
认识人工智能

学习目标

- 了解人工智能的产生与发展，体会人工智能对社会发展的影响。
- 通过典型实例理解人工智能技术，感受人工智能魅力。

体验探索

人机博弈

人机博弈一直是人工智能研究的经典领域。人机博弈的发展历程可以用“三盘棋”来概括。

第一盘棋：1956年计算机技术的先驱萨缪尔（Arthur Lee Samuel），在计算机上编写出了世界上第一款国际跳棋程序。

第二盘棋：1997年人工智能系统——深蓝首次在正式比赛中战胜国际象棋世界冠军卡斯帕罗夫（Garry Kasparov），成为人工智能发展史上的一个里程碑。

第三盘棋：2016年阿尔法围棋程序（AlphaGo）在与围棋世界冠军李世石进行的围棋人机博弈中以4：1的总比分获胜，这是人工智能发展史上又一个新的里程碑。2017年，它又进化为阿尔法元（AlphaGo Zero），通过“自学成才”，仅用3天就成了围棋界的顶尖高手。

思考：

1. 查找资料，从智能水平、不同类型棋的难度等方面分析以上三次人机博弈的区别。

2. 体验人与机器对弈的过程，讨论计算机是如何进行“思考”的，如图4.1.1所示。

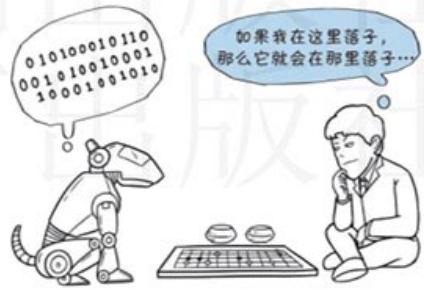


图4.1.1　人机博弈

中国地图出版社
人民教育出版社

4.1.1 人工智能的产生与发展

说到人工智能，很容易联想到机器人，其实人工智能有很多表现形式，如智能网络、智能设备和智能软件等。随着技术的发展，人们对人工智能的认识也在不断深入。“三盘对弈”掀起了人工智能发展的三次高潮：第一款国际跳棋程序中使用的“机器学习”概念，至今仍然对人工智能程序设计有着重要的影响；深蓝将知识库结合搜索等机器学习算法引入人工智能，让机器从数据中进行学习，获得知识；阿尔法围棋推动了深度学习作为人工智能继续前进的核心技术。

思考活动

图灵测试与人工智能

英国科学家图灵（Alan Mathison Turing）提出测试机器智能的方法：测试者与被测试者在隔开的情况下，通过一些装置（如键盘）向被测试者任意提问。经过5分钟问答后，如果测试者不会有多于70%的机会做出正确的区分，那么这台机器就通过了测试，这就是著名的图灵测试。图4.1.2是图灵测试示意图。

图4.1.2 图灵测试示意图

思考：

你认为通过图灵测试能否判断出机器是否具有智能，试说明理由。机器是如何获得智能的？

中国地图出版社
人民教育出版社

人工智能这一术语是1956年在美国达特茅斯学院举行的一次学术研讨会上被提出的。经过多年的演进，特别是在移动互联网、大数据、超级计算、脑科学等新理论、新技术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，呈现出深度学习、跨界融合、人机协同和自主操控等新特征。中国电子技术标准化研究院组织编写的《人工智能

标准化白皮书（2018版）》中将人工智能定义为：利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

经过多年的持续发展，我国在人工智能领域也取得了重要进展，其中吴文俊院士提出了几何定理证明的“吴氏方法”，对人工智能的理论研究做出了巨大的贡献。此外，我国在语音识别技术、视觉识别技术等方面处于世界先进水平，在自适应自主学习、直觉感知、综合推理、混合智能和群体智能等方面初步具备跨越发展的能力。目前，中文信息处理、智能监控、生物特征识别、工业机器人、服务机器人、无人驾驶逐步进入实际应用，在国际上获得广泛关注和认可。

如今，人工智能已成为未来经济发展的新引擎和新一轮产业变革的核心驱动力，作为国家未来的建设者和接班人，我们应当努力学习新知识和新技术，提高自己的信息意识，培养创新的能力，在信息社会中担当起我们的社会责任。

阅读拓展

吴文俊与“吴氏方法”

吴文俊（1919—2017），我国著名数学家（图4.1.3）。1940年毕业于上海交通大学，1949年获法国斯特拉斯堡大学博士学位。他在拓扑学、自动推理、机器证明、代数几何、中国数学史和对策论等研究领域均有杰出的贡献，在国内外享有盛誉。2001年，获首届国家最高科学技术奖。

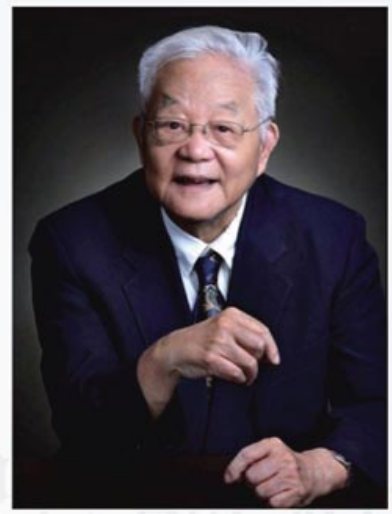


图4.1.3　著名数学家吴文俊

在人工智能的机器证明方面，他从初等几何着手，在计算机上证明了一类高难度的定理，同时也发现了一些新定理，进一步探讨了微分几何的定理证明，提出了利用机器发现与证明几何定理的新方法，这就是著名的“吴氏方法”。这项研究在数学领域产生了深远的影响，为数学研究和人工智能的机器证明开辟了新的领域，有重要的应用价值。

“吴文俊人工智能科学技术奖”是为了奖励我国人工智能领域有成就和创新的个人或项目而设立的，设有科学技术成就奖、科学技术创新奖和科学技术进步奖等奖项，从2011年开始，每年评奖一次。

4.1.2 感受人工智能的魅力

人工智能离我们并不遥远，人脸识别、手写识别、语音识别、机器翻译和智能推荐等人工智能技术已经进入我们的生活，逐渐改变着我们的生活方式。人工智能的研究领域非常广泛，我们在此仅从人机交互、图像识别与生物特征识别、自然语言处理、机器学习和深度学习等几个方面来感受人工智能的魅力。

人机交互

人机交互技术主要是研究人与计算机之间的信息交换，包括人到计算机和计算机到人的信息交换两部分。人们可以借助键盘、鼠标、话筒、操纵杆、压力笔、数据手套或服装、眼动跟踪器、位置跟踪器等设备，用手、脚、声音、姿势或动作、眼睛甚至脑电波等向计算机传递信息；同时，计算机通过打印机、绘图仪、显示器、头盔式显示器和音箱等输出设备给人提供信息。人机交互技术除了传统的基本交互（键盘、鼠标等）和图形交互外，还包括语音交互、情感交互、体感交互及脑机交互等技术，其中语音交互是一种高效的交互方式，是人以自然语音或机器合成语音同计算机进行交互的综合性技术。语音交互过程包括四部分：语音采集、语音识别、语义理解和语音合成。

语音采集完成音频的获取、采样及编码；语音识别完成语音信息到机器可识别的文本信息的转换；语义理解根据语音识别转换后的文本字符或命令完成相应的操作；语音合成完成文本信息到音频信息的转换。作为人类沟通和获取信息最自然、便捷的手段，语音交互比其他交互方式具备更多优势，能为人机交互带来根本性变革。它在智能电视、智能导航、智能家居等方面都有着广泛的应用。

实践活动

体验“语音识别”

很多导航软件都有语音交互功能，试使用语音控制的导航软件，通过体验语音导航查询路况和周边信息等功能，了解语音交互过程。

试用家乡的方言体验导航软件识别方言的效果，查阅资料并讨论导航软件提高方言识别率的方法。

图像识别与生物特征识别

图像识别是利用计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技术。图像识别技术是以图像的主要特征为基础的。每个图像都有其特征，例如，字母A有个尖，P有个圈，而Y的中心有个锐角等。对图像识别时眼动研究表明，视线通常是集中在图像的主要特征上，也就是集中在图像轮廓曲度最大或轮廓方向突然改变的地方，这些地方的信息量最大。图像识别技术是人工智能技术的一个重要领域。生物特征识别技术是在图像识别基础上发展起来的，是利用生物特征（如指纹、虹膜和人脸等）进行识别，目前已经在生活中得到了广泛的应用。图4.1.4所示为指纹签到。

图4.1.4　指纹签到

生物特征识别技术是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，主要是对生物特征进行取样，从中提取其唯一特征，并转换成相应的代码，并进一步将这些代码组合成具有代表性的特征模板。在利用识别系统进行身份识别时，识别系统将获取的生物特征与具有代表性特征的模板进行比对，以确定是否匹配，从而给出判断。生物特征识别技术涉及的内容十分广泛，包括指纹、掌纹、人脸、虹膜、脉搏、步态等多种生物特征，其识别过程涉及图像处理、计算机视觉、语音识别、机器学习等多项技术。

实践活动

体验“图像识别”

生活中有许多事物，我们可能不认识或不能准确地叫出它们的名称，借助具有人工智能技术的智能软件就可以帮助我们解决这个问题。使用识别软件，体验实物识别、图片查询和相关知识搜索等功能，如图4.1.5所示。实际体验或上网查询图像识别工具，并与同学探讨图像识别在生活中有哪些应用。

图4.1.5　图像识别

中国地图出版社

自然语言处理

自然语言处理技术是计算机科学和人工智能领域中的一个重要方向，主要研究实现人与计算机之间用自然语言进行有效通信、交流的各种理论和方法。它涉及的领域很多，包括机器翻译、语义理解（机器阅读理解）和问答系统等。机器翻译利用计算机技术实现从一种自然语言到另外一种自然语言的翻译，机器翻译的一般过程包括源语言文字的输入、识别与分析、生成与综合和目标语言的输出。语义理解（机器阅读理解）则是利用计算机技术实现对文本篇章的理解，并且回答与篇章相关的问题。语义理解更注重对上下文的理解及对答案精准程度的把控。问答系统技术指让计算机像人类一样用自然语言与人交流的技术，人们可以向问答系统提交用自然语言表达的问题，系统会返回关联性较高的答案。

实践活动

体验“机器翻译”技术

在生活和学习中，我们会经常用到机器翻译工具，如图4.1.6所示。尝试使用一款机器翻译软件对几段文字进行中英文互译。讨论机器翻译相比人工翻译所具有的优势和不足。

图4.1.6　机器翻译

机器学习

人工智能在生活中应用广泛，也要归功于机器学习技术。语音识别、图像识别都有它的身影。机器学习研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能，是人工智能技术的核心。基于数据的机器学习是现代智能技术中的重要方法之一，研究从观测数据出发寻找规律，利用这些规律对未来数据或无法观测的数据进行预测。

机器学习的实现是通过在计算机中存储历史数据，应用机器学习算法对这些数据进行处理，这个过程在机器学习中称为“训练”，处理的结果可以被用来对新的数据进行预测，这个结果一般称之为“模型”。对新数据的特性进行推测的过程在机器学习中称为“预测”。“训练”与“预测”是机器学习的两个过程，“训练”产生“模型”，“模型”指导“预测”。图4.1.7是机器学习的示意图。

图4.1.7　机器学习示意图

深度学习其实是“深度神经网络的学习”，是机器学习的一个重要领域。它通过能逐步发现原始数据中的特征与待求解问题之间对应和关联的算法模型，把原始数据转变成更高层次的、更加抽象的表达。深度学习的实质是通过构建具有较多隐层的网状学习模型和海量的训练数据来学习更有用、更高阶的特征，从而提升分类和预测的准确性。例如，很多智能手机中的图像快速分类功能（能从上万张照片中选出包含某个人的照片）就应用了深度学习技术。目前，深度学习在自然语言处理、图像识别等领域都取得了很大的突破。

中国地图出版社
人民教育出版社

项目实施

制订项目方案，体验与智能机器人对话

一、项目活动

关注带有“智能对话机器人”的微信公众号，体验人机对话的过程，理解人机对话所涉及的人工智能技术。

1. 按照项目目标从项目任务、项目过程和项目检查等方面制订项目方案。

2. 体验手写识别和语音识别：通过手写或语音输入问题，观察“智能对话机器人”的识别效果。

3. 体验与“智能对话机器人”的交互：输入学习中的问题，观察机器人的回复，体验“智能对话机器人”帮助我们学习的过程，如图4.1.8所示。

图4.1.8　与智能机器人对话

二、项目检查

小组通过手写或语音输入信息的方式，体验与“智能对话机器人”对话的过程。

中国地图出版社
人民教育出版社

练习提升

1. 目前很多网络平台都设计有“客服”功能。找到一个网络平台的“客服”，请设计一个提问方案，并实施该方案，判断该客服是人还是人工智能程序，说明原因。

2. 试观察一些应用程序中的“智能客服”，讨论它与项目活动中的“智能对话机器人”在对话效果上有什么区别。

4.2
利用智能工具解决问题

学习目标

- 了解人工智能平台中的智能工具，体会人工智能对我们生活的影响。
- 经历使用智能工具解决问题的过程，掌握利用智能工具解决问题的一般方法。

体验探索

北斗卫星导航系统

北斗卫星导航系统（如图4.2.1）是我国着眼于国家安全和经济社会发展需要，自主建设、独立运行的卫星导航系统，是为全球用户提供全天候、全天时、高精度的定位、导航和授时服务的国家重要空间基础设施。北斗卫星导航系统是实现安全可信、高效便捷的国家综合体系的核心，在人工智能等影响未来产业战略格局的前沿领域发挥着巨大的作用。

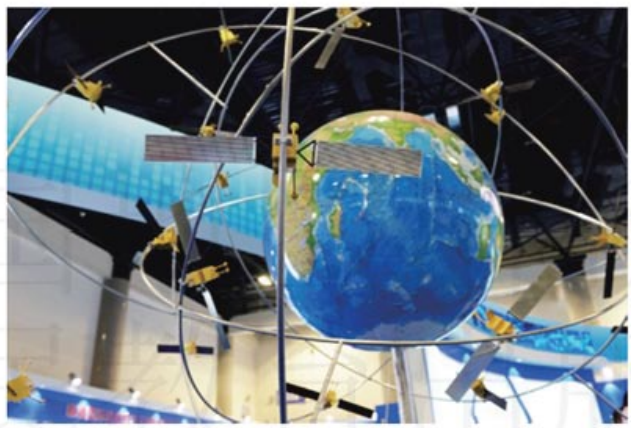


图4.2.1　北斗卫星导航系统模型

思考：

查阅相关资料，讨论北斗卫星导航系统与人工智能技术结合是如何改变我们的生活的。

中国科学技术出版社
人民教育出版社

4.2.1 人工智能平台中的智能工具

随着人工智能技术的发展，智能工具层出不穷，使用起来也越来越方便。2017年，我国发布的《新一代人工智能发展规划》，明确了构建开放协同的人工智能科技创新体系，建设安全便捷的智能社会，构建安全高效的智能化基础设施体系，加强人工智能领域的军民融合，培育高端高效的智能经济，前瞻布局新一代人工智能重大科技项目六大任务。

此后，2018年国家公布了新一代人工智能开放创新平台，包括自动驾驶、城市大脑、医疗影像、智能语音和智能视觉五大方面。这些平台选择大数据智能、跨媒体智能、群体智能、混合增强智能和自主智能系统等主力方向，从基础理论、支撑体系、关键技术、创新应用等层面进行系统性、前瞻性任务布局。这些开放创新平台已经发布了一些研发成果，登录这些人工智能平台，使用其中的智能工具能够帮助我们解决很多问题，如图4.2.2所示。

图4.2.2 利用人工智能平台实现“刷脸签到”的设置页面

中国地图出版社
人民教育出版社

实践活动

利用人工智能平台体验人脸识别

一些人工智能平台提供了开放的智能工具服务功能，例如：文字识别、语音识别和人脸识别等。尝试搜索这些平台，利用其中的“人脸会场签到”功能解决实际问题。在组织班会活动中，同学们可体验一下“人脸会场签到”。

1. 会议准备：注册人工智能平台账号，使用“人脸会场签到”工具，收集参会人员的相关信息。

2. 会议现场：参会人员通过数字设备进行人脸签到。

3. 会后信息处理：会后将参会信息进行相应的数据分析，并将信息保存。

观察用照片签到能否成功，讨论人脸签到相比其他签到方式的优势与不足。此外，说出人脸识别还有哪些应用场景。

4.2.2 编程调用平台中的智能工具

用智能工具解决问题时，除了使用智能平台上已经设计好的智能工具以外，还可以编程自主开发，这样可以提高智能工具的灵活性。

目前可以开发人工智能的高级语言比较多，我们前面学习过的Python 语言也可以用于开发人工智能。使用Python语言开发人工智能有以下几个优势：

- 多平台运行，可以在mac OS、Windows、Linux等操作系统上运行；
- 强大的数据处理库，包括NumPy、SciPy、pandas、Matplotlib等；
- 丰富的第三方库，包括Web开发、爬虫、数据库、机器学习、科学计算框架、图形界面开发等。

通过Python语言调用人工智能工具库可以较方便地开发智能工具。但是，人工智能工具在开发与使用过程中，通常需要进行较长时间的、大量数据的前期训练，这给学习体验智能工具带来了一定困难。在此，我们利用Python语言调用智能平台上的智能工具来体验编程开发智能工具的方法。

例：编程开发“信息值守机器人”。

在使用个人微信时，有时我们不能及时回复朋友的信息，“信息值守机器人”可以帮助我们解决这个问题。使用Python语言就可以开发这样的机器人，如图4.2.3所示。

图4.2.3 开发“信息值守机器人”过程示意图

（1）引入Python模块

该程序需要调用sinomaps模块（本教材专用模块）和itchat模块（itchat 模块是连接微信的一个开源模块，使用它可以进行微信中的操作），用于与网络平台的服务器进行交互和与微信交互。

（2）向网络服务发送请求

通过向一个网络服务的网址发送请求，网络服务提供商一般会提供调用相应封装好

的服务。本程序中使用的是sinomaps当中的函数get_response（该函数用以返回数据）调用服务。

（3）关联机器人平台账号

关联机器人平台账号首先要注册账号获取KEY。KEY起到的作用是与平台机器人服务进行认证和对接。具体关联方法可参考教学资源平台。

（4）运行程序

```
import itchat                                          # 引入微信回复的模块
import sinomaps                                        # 引用网络服务请求模块
KEY = "4d1273276b3448ea89c32121ee4cadd8"               # 关联KEY
@itchat.msg_register(itchat.content.TEXT)              # 注册
def tuling_reply(msg):  # 为了保证KEY出现问题仍旧可以回复，设置一个默认回复
    # 若没有找到回复，defaultReply为提问内容
    defaultReply = "我收到：" + msg["Text"]
    reply = sinomaps.get_response(msg["Text "], KEY) # reply为找到的回复内容
    # 返回找的内容或提问的内容
    if reply:
        return reply
    else:
        return defaultReply
itchat.auto_login(hotReload=True)    # 为了让修改程序不用多次扫码，使用热启动
itchat.run()
```

运行此程序后，会出现一个二维码图片。用微信扫描该二维码，并确认登录，即可将自己的微信变为一个具有简单对话功能的信息值守机器人。

实践活动

与“信息值守机器人”对话

邀请好友与自己通过微信进行对话，观察当你不回答时机器人是怎样回复的。

1. 思考机器人的回答与自己的回答有何不同，它能取代你自己吗？

2. 找一位不知情的好友，让他与机器人对话。一段对话后，询问好友，看他是否知道是机器人在和他对话。

技术支持

Python语言第三方库的功能

Python语言所具有的快速、可移植、开源、可扩展等优点，深受人们的欢迎，特别是其众多的人工智能库，使得编写人工智能工具变得非常方便。目前，Python已成为人工智能程序设计的重要编程语言。了解Python语言的第三方库的功能有助于我们进一步学习。

中国地图出版社
人民教育出版社

1. Web开发

Django：是一个高效的Web开发框架。使用Django能够快速、便捷地构建和维护高质量的Web应用。

Tornado：是一种Web服务器软件的开源版本。

Flask：是一个使用Python编写的轻量级Web应用框架。

2. 爬虫

Requests：是一个很实用的客户端库，编写网络爬虫和测试服务器响应数据时经常会用到它。

Scrapy：是一个为了爬取网站数据，提取结构性数据而编写的应用框架。

Selenium：是一个用于Web应用程序测试的工具，运行在浏览器中，可模拟人的动作自动进行浏览操作。

3. 数据库

PyMongo：用于连接MongoDB数据库。

PyMySQL：用于连接MySQL数据库。

4. 机器学习

NLTK：是一个比较优秀的自然语言处理工具包，是聊天机器人需要的重要工具。

Keras：是一个高层神经网络的应用程序编程接口。

scikit-learn：是一个机器学习库，其中包含大量机器学习算法和数据集，是方便的数据挖掘工具。

5. 图形界面开发

PyQt：是一个创建图形用户界面（GUI）的工具包。

中国地图出版社
人民教育出版社

项目实施

开发“智能班级交互系统”

一、项目活动

借助智能平台上的智能工具和班级微信公众号，开发“智能班级交互系统”，利用这个系统答疑解惑。

1. 在开放的智能平台上注册账号。

2. 对接微信公众号和智能平台（如图4.2.4所示），开发“智能班级交互系统”。

图4.2.4　对接微信公众号和智能平台

3. 使用“智能班级交互系统”，体验人机对话的过程。

二、项目检查

完成开放智能平台上账号的注册，实现微信公众号和智能平台对接，利用“智能班级交互系统”实现人机对话。

练习提升

1. 诗歌是人类文学皇冠上的明珠，深度学习技术的发展让机器自动生成诗歌成为可能。查找相关智能应用程序，体验机器作诗，分析机器是如何做到的。思考机器能否替代人通过诗歌抒发情怀。

2. 随着移动设备的普及，越来越多的移动应用程序也使用了人工智能技术。分享一个使用了人工智能技术的移动应用程序，说明其特点。

中国插图出版社
人民教育出版社

4.3
人工智能的应用与影响

学习目标 ▶▶▶

- 了解人工智能技术在不同领域的应用。
- 感受人工智能技术对人类的影响，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。

体验探索

诚信点亮人生

人工智能技术可以用于基础研究、技术研发、产业发展和行业应用。通过采集用户的日常行为信息，结合大数据分析，利用人工智能技术进行用户数据画像、信用评估并产生信用报告（图4.3.1）。信用报告在找工作、买房、买车、办信用卡等过程中都起着重要的作用。我国正逐步建立和完善社会信用体系，个人诚信信息也会记录到个人信用体系中。自觉遵守信息社会的行为规范和道德准则，是每个公民需要承担的信息社会责任。“诚信点亮人生”，诚信是幸福生活的基础！

中国地图出版社
人民教育出版社

图4.3.1　个人信用报告

思考：

查阅相关资料，从数据收集、数据分析及信息应用等方面讨论人工智能技术在形成个人信用报告中所起的作用。

4.3.1 人工智能的应用

人工智能不仅影响着个人生活，对社会其他领域也有重要影响。人工智能与行业领域的深度融合改变和重塑了传统行业，其应用主要包括工业、农业、医疗及安全保卫等行业。

人工智能在工业中的应用

人工智能引发了工业的巨大变革。如今，工业机器人在工业生产中已经可以代替人做某些单调、重复的长时间作业，或从事危险、恶劣环境下的作业。通过整合来自物联网传感器、维护日志和其他外部资源的数据，使用图像识别技术进行目视检查和故障检测，人工智能可以更好地预测和避免机器故障，从而提高生产力，降低废料率和测试成本。例如，上海洋山深水港是我国运用物联网、自动化、大数据、云计算和人工智能等技术自主建造的规模大、技术先进的智慧型港口。全自动化码头智能生产管理控制系统是洋山深水港的“大脑”，它与上海港各大数据信息平台衔接，提供智能的生产计划模块、实时作业调度系统及自动监控调整等功能。智能控制系统是洋山深水港的“神经”，它把自动导引车变成一个个智能车队，把一台台场桥起重机变成一个个智能化的堆场，把一座座岸桥起重机变成能自主作业的巨型机器人。操作员只需坐在计算机前，用智能控制系统协调这些智能化设备和系统，自动高效地完成码头操作系统下达的装卸任务，实现了码头装卸、水平运输、堆场装卸环节的全过程智能化。其中，自动导引车借助无线通信、自动调度和精密定位系统，在地面敷设的6万多个磁钉的引导下，能准确无误地将集装箱运送到指定位置，如图4.3.2所示。

图4.3.2　洋山港无人驾驶的自动导引车

人工智能在农业中的应用

智能化农业指利用信息技术结合设施、设备来指导农业预测、实施生产。基于物联网的智能化农业可用于大中型农业种植基地、设施园艺、畜禽水产养殖和农产品物流。智能化农业机器人能够自动识别农作物，确定幼苗间距是否合适，杂草是否该除，什么时候除等，以帮助农民减少购买农药的支出和人力的投入。例如，基于机器视觉的病虫害

自动监测平台，可以帮助农民科学判断病虫害的状况。实际使用时，农民为农作物拍照，将照片上传到病虫害自动监测与预警系统，系统通过图像识别，识别出破坏农作物的病虫害，找出对应的措施和方法，帮助农民进行相应的处理，如图4.3.3所示。

图4.3.3　农作物有害生物监测信息系统

人工智能在医疗卫生中的应用

人工智能的快速发展为医疗卫生领域向更高的智能化方向发展提供了技术条件。智能医疗在辅助诊疗、疾病预测、医疗影像辅助诊断、药物开发等方面发挥了重要作用。在辅助诊疗方面，通过人工智能可以有效提高医护人员工作效率，提升一线全科医生的诊断治疗水平。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。在医疗影像辅助诊断方面，通过智能化的影像判读系统对特征进行提取和分析，为患者预前和预后的诊断和治疗提供评估方法和精准诊疗决策，节约了人力成本。

一位经过训练的医疗专家每天诊治患者的数量是有限的。如今，利用人工智能技术可以辅助医疗专家提高对患者的诊治效率。例如，智能机器人使用数量庞大的患者诊疗电子档案数据，可以在极短的时间内找出患者的病理依据，辅助医疗专家迅速、准确地对患者的病情作出判断，如图4.3.4所示。

图4.3.4　智能医疗

人工智能在安全保卫行业中的应用

随着智能安防系统在全国各地普及，“天网”已初具规模。2017年，英国广播公司的一名记者在贵阳经过当地警方的许可做了一次模拟测试。警方用手机拍下他的面部照片，

将其录入“嫌疑人”名单。随后，这名记者开始了他的“潜逃之旅”。然而仅用了几分钟，他就被警察“抓获”了。图4.3.5是人工智能在警务工作中的应用。

图4.3.5　智能警务

人工智能的发展给社会建设带来了新的机遇，在教育、养老、环境保护、城市运行等领域的广泛应用，极大地提高了公共服务精准化水平，全面提升了人们的生活品质。

4.3.2　人工智能的影响

人工智能技术给人类社会带来的影响是全方位的。人工智能机器可以替代人类做一些危险、繁重的工作；人工智能机器可以持续工作，大大地提升工作效率，节约大量的社会成本；人工智能机器可以代替人类做家务，帮助人类学习，实时监护人类的健康，并能为人做精准诊断和治疗；人工智能技术还可以拓展人类的感知，帮助人类探索新的未知领域，让人们的生活越来越美好。

但是，任何事物都具有两面性，人工智能技术也不例外。它在隐私、安全和伦理方面给人类带来了新挑战，值得我们深思。例如，国外一个“打击犯罪的机器人”曾因误判而导致一名16岁的少年受伤。

早在1950年，就有人提出了机器人三定律。

第一定律：机器人不得伤害人类个体，或者目睹人类个体将遭受危险而袖手旁观。

第二定律：机器人必须服从人类给予它的命令，当该命令与第一定律冲突时例外。

第三定律：机器人在不违反第一、第二定律的情况下要尽可能保护自己的生存。

我国《人工智能标准化白皮书（2018版）》提出了应当遵循的原则：一是人类利益原则，即人工智能技术应以实现人类利益为终极目标；二是责任原则，即在技术开发和应用两方面都建立明确的责任体系，以便在技术层面可以对人工智能技术开发人员或部门问责，在应用层面可以建立合理的责任和赔偿体系。

人工智能技术的不断发展和完善，为人类社会的发展提供了新的工具和动力。如何面对人工智能对人类社会的影响，如何更好地利用它为人类社会服务，是我们所有人应该思考的问题。

项目实施

“智能班级交互系统”的应用拓展

一、项目活动

在与“智能班级交互系统”对话的过程中产生了大量数据，这些数据被保存在智能平台上，处理这些数据，可以帮助我们拓展“智能班级交互系统”的应用。

1. 在使用“智能班级交互系统”对话过程中会产生诸如对话时间、学生问题、机器人回复解答等数据，对这些数据进行归类。

2. 对“智能班级交互系统”数据进行分析。利用前面章节学习的知识，找出同学们最关心的问题。

3. 对“智能班级交互系统”进行个性化设置。观察：哪些问题有答案，哪些问题无答案，正确答案是什么等，然后根据问题结果不断充实完善知识库。

4. 完善项目成果，并进行交流。

二、项目检查

导出“智能班级交互系统”的数据并丰富其知识库，撰写支持学习改进、提升的方案。

练习提升

1. 查阅资料，了解人工智能在不同领域应用的典型案例，完成表4.3.1。

表4.3.1　人工智能的应用领域及典型案例

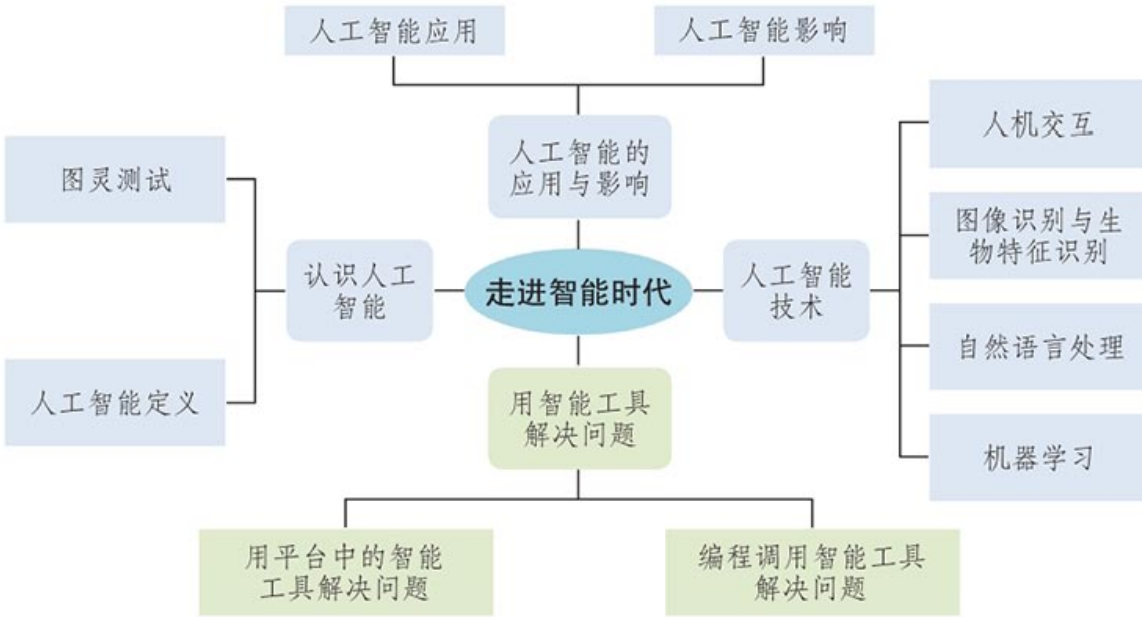
序号	应用领域	典型案例
1	智能工业	
2	智能农业	
3	智能医疗	
4	智能教育	
5	智能安防	
6	智能交通	
7	智能家居	
8	智能服务	

2. 讨论人与人工智能的关系，思考人工智能发展的未来。作为一名信息社会的公民，应如何信守社会道德与伦理准则？应担负起怎样的社会责任？

中国地图出版社
人民教育出版社

总结与评价

1. 下图展示了本章的核心概念与关键能力，请同学们对照图中的内容进行总结。



2. 根据自己的掌握情况填写下表。

学习内容	掌握程度
人工智能概念	□不了解 □了解 □理解
人工智能技术	□不了解 □了解 □理解
人工智能平台	□不了解 □了解 □理解
人工智能工具	□不了解 □了解 □理解
人工智能应用	□不了解 □了解 □理解

中国地图出版社
人民教育出版社

项目 评价

在完成项目活动后，请各组对项目完成情况进行评价。评价实施围绕项目主题、实施过程、分工合作、项目成果和展示交流五方面进行。根据项目评价中的评分参考，结合项目实际完成情况，确定各项评分结果，给出评分理由。同时，对项目活动进行全面梳理，指出需要进一步改进的地方。将评价内容如实填写到项目评价表中。

评价项	评分参考	评分（1～5分）	评分理由	待改进之处
项目主题	项目主题能反映出学科核心素养的要求（信息意识、计算思维、数字化学习与创新、信息社会责任）；主题任务与学习目标保持一致			
实施过程	项目研究计划详细，准备充分；实施过程完整，过程记录翔实，资料丰富；研究数据来源渠道多，出处明确，收集方式多样，质量高；研究方法得当，技术手段适宜			
分工合作	小组成员分工明确，态度积极，参与度高；善于提出问题，分析问题，解决问题能力强；踊跃分享观点，交流充分；能在完成自己任务的前提下，乐意帮助他组完成任务			
项目成果	项目活动成果丰富，内容具体，符合项目目标要求；研究结论清晰准确，有价值，有创新，具有指导及建设意义；项目报告或作品内容完整，论述充分，表述清楚，整齐美观			
展示交流	项目展示形式新颖，综合运用多种技术呈现成果，表现力强；语言表达清晰准确，逻辑性好			
项目总分				

中国出版社
人民教育出版社

后　记

本册教科书是中国地图出版社与人民教育出版社依据教育部《普通高中信息技术课程标准（2017 年版）》，由双方共同组织团队联合编写的，经国家教材委员会 2019 年审查通过。

本册教科书的编写，集中反映了我国十余年来普通高中课程改革的成果，吸取了 2004 年版《普通高中课程标准实验教科书 信息技术》的编写经验，凝聚了参与课改实验的教育专家、学科专家、教材编写专家、教研人员和一线教师，以及教材装帧设计专家的集体智慧。本册教科书的编写人员还有兰大鹏、王南迪、张怡、李维钧、马宏波、沈万君，审校人员有吴劲松、胡志刚、叶慧琳、相远红。为本册教科书进行装帧设计的有吕旻、李媛，绘制插图的有北京大方四象，摄影或提供照片的有新华社记者等。

我们感谢所有对教科书的编写、出版、试教等提供过帮助与支持的同仁和社会各界朋友。同时，我们还要感谢 2004 年版《普通高中课程标准实验教科书 信息技术》的编写人员。

本册教科书出版之前，我们通过多种渠道与教科书选用作品（包括照片、画作）的作者进行了联系，得到了他们的大力支持。对此，我们表示衷心的感谢！恳请未联系到的作者与我们联系，以便及时支付稿酬。

我们真诚地希望广大教师、学生及家长在使用本册教科书的过程中提出宝贵意见。我们将集思广益，不断修订，使教科书趋于完善。

联系方式

电　　话：010-83543863　　010-58758866

电子邮箱：sinomaps@yeah.net　　jcfk@pep.com.cn

中国地图出版社教材出版分社

人民教育出版社课程教材研究所信息技术课程教材研究开发中心

2019 年 4 月

PUTONG GAOZHONG JIAOKESHU

XINXI JISHU

绿色印刷产品

ISBN 978-7-107-33845-8

9 787107 338458 >

定价： 元