

21世纪高等教育计算机规划教材

COMPUTER

计算思维 与算法设计

Computational Thinking and
Algorithm Design

麻新旗 王春红 主编

张世民 李颖 程欣 副主编

注重思维方法的训练

以实际案例驱动教学

结合新技术拓展思维



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

21世纪高等教育计算机规划教材

COMPUTER

计算思维 与算法设计

Computational Thinking and
Algorithm Design

■ 麻新旗 王春红 主编

■ 张世民 李颖 程欣 副主编



人民邮电出版社

北京

图书在版编目(CIP)数据

计算思维与算法设计 / 麻新旗, 王春红主编. -- 北京: 人民邮电出版社, 2015.9(2016.9重印)
21世纪高等教育计算机规划教材
ISBN 978-7-115-39828-4

I. ①计… II. ①麻… ②王… III. ①计算方法—思维方法—高等学校—教材②电子计算机—算法设计—高等学校—教材 IV. ①0241②TP301.6

中国版本图书馆CIP数据核字(2015)第211785号

内 容 提 要

本书以计算思维能力培养为主线,以算法设计为依托,以介绍计算机基础和算法设计为主要目标,主要包括计算与计算模型、计算机基础知识、算法设计、算法评测与分析、软件测试基础等。通过学习本书,学生可以了解与现代计算机相关的计算理论基础,了解算法设计与问题求解的关系,体会计算在现代生活中的重要性和普适性,进而为今后利用计算机解决专业问题打下良好基础。

本书可作为高等院校计算机基础课程相关的教材,也可以作为计算机基础知识及算法设计培训、自学的参考用书。

-
- ◆ 主 编 麻新旗 王春红
 - 副 主 编 张世民 李 颖 程 欣
 - 责任编辑 许金霞
 - 责任印制 沈 蓉 彭志环
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 大厂聚鑫印刷有限责任公司印刷
 - ◆ 开本: 787×1092 1/16
 - 印张: 9.75 2015年9月第1版
 - 字数: 254千字 2016年9月河北第3次印刷
-

定价: 25.00 元

读者服务热线: (010) 81055256 印装质量热线: (010) 81055316
反盗版热线: (010) 81055315

前 言

教育部高等学校大学计算机课程教学指导委员会对大学非计算机专业计算机基础教学提出的以计算思维为核心的基本要求,是本书的编写依据。编写本书的目的是为了加强学生对计算、计算模型、计算思维及算法等与计算学科相关的知识的理解、应用和掌握,提高学生利用计算机进行问题求解的能力,进而培养学生的信息素养和计算思维能力。

随着移动通信、物联网、云计算、大数据等新概念和新技术的出现,信息技术深刻地改变着人类的思维、生产、生活、学习方式,与之相关的计算思维成为人们认识 and 解决问题的能力之一。我们认识到,计算思维不光是计算机专业学生应该具备的能力,更应该是所有大学生都要具备的能力。并非每一个学生都要成为计算机科学家,但是我们期望他们能够掌握正确的计算思维的基本方式,这种思维方式对学生的事业和学习都是有益的。

对于当代大学生而言,大学学习,除了掌握本专业的基础知识外,科学精神的培养、思维方法的训练、严谨踏实的品质,以及分析问题、解决问题的能力,都是日后工作的基础。本书在内容编排、课后习题选择等方面做了精心设计,并力求按提出问题、分析问题、解决问题的思路和方法,采用“案例驱动”循序渐进地展开教学。本书的重点放在初学者对计算机科学的核心理论和经典问题的理解和掌握上,使初学者感悟:当遇到一个实际问题时能否用计算机实现,使用计算机如何实现及如何设计解决问题的方法和步骤,并进一步设计相应问题求解的算法,分析比较同一问题的不同算法的有效性和复杂性,拓展思维,建立起利用计算机技术解决问题的思路。

本书共 6 章。第 1 章介绍计算及计算思维的相关知识,由王春红编写。第 2 章介绍现代计算机的基础知识,由张世民编写。第 3 章和第 4 章重点介绍算法基础及常用算法,由麻新旗编写。第 5 章介绍算法的评测与分析,由李颖编写。第 6 章介绍软件工程基础,由程欣编写。

在本书的编写过程中,作者参考了大量的书刊和文献资料,征求了许多任课老师的宝贵意见,也得到了学院领导的大力支持,在此一并表示感谢。

编 者

2015 年 8 月

目 录

第1章 计算与计算模型..... 1

1.1 计算..... 1
1.1.1 什么是计算..... 1
1.1.2 什么是计算的本质..... 2
1.1.3 计算与算法..... 2
1.2 图灵机模型..... 2
1.3 停机问题..... 3
1.4 计算思维..... 4
1.4.1 科学思维与计算思维..... 4
1.4.2 计算思维的概念及内涵..... 5
1.4.3 计算思维的本质..... 7
1.4.4 计算思维能力..... 7
1.5 计算在其他学科的应用..... 7
1.5.1 计算社会学..... 7
1.5.2 计算生物学..... 8
思考题..... 9

第2章 计算机基础知识..... 10

2.1 计算机组成与工作过程..... 10
2.1.1 计算机简介..... 10
2.1.2 计算机的组成..... 10
2.1.3 计算机系统结构..... 11
2.1.4 计算机工作过程..... 13
2.2 信息在计算机中的表示..... 14
2.2.1 进位计数制及数制转换..... 14
2.2.2 带符号数在计算机中的表示..... 18
2.2.3 定点数与浮点数..... 18
2.2.4 逻辑运算..... 20
2.2.5 二进制编码..... 20
2.2.6 多媒体信息在计算机中的表示..... 23
2.3 计算机应用技术..... 25
2.3.1 普适计算..... 25
2.3.2 网格计算..... 25
2.3.3 云计算..... 25

2.3.4 物联网..... 26

2.3.5 大数据..... 26

思考题..... 26

第3章 基本算法设计方法..... 28

3.1 算法基础知识..... 28
3.1.1 算法的概念..... 29
3.1.2 算法的描述..... 30
3.1.3 简单算法举例..... 32
3.2 Raptor 流程图编程..... 37
3.2.1 Raptor 简介..... 38
3.2.2 输入语句..... 43
3.2.3 处理语句..... 44
3.2.4 输出语句..... 44
3.2.5 Raptor 应用基础..... 45
3.3 枚举法..... 56
3.4 迭代与递推算法..... 60
3.5 递归方法..... 65
3.6 数组在算法中的应用..... 69
3.6.1 数值数组的使用..... 70
3.6.2 字符串处理..... 77
3.7 优化算法的基本技巧..... 80
3.7.1 选择高效的数学模型..... 81
3.7.2 巧妙利用算术运算功能..... 82
3.7.3 设置标志量..... 83
3.7.4 将非数值信息数字化..... 87
思考题..... 89

第4章 复杂算法设计方法简介..... 94

4.1 分治法..... 94
4.2 回溯法..... 98
4.3 贪心法..... 105
4.4 动态规划法..... 108
4.5 算法总结..... 116
4.5.1 算法策略小结..... 117

4.5.2 算法策略之间的关系及 侧重解决的问题类型.....	119	6.1.2 软件工程	136
思考题.....	120	6.1.3 软件生存周期.....	136
第 5 章 算法的评价与分析.....	122	6.1.4 软件工程方法.....	137
5.1 算法的评价	122	6.1.5 结构化程序设计方法与 程序设计风格.....	137
5.1.1 算法的正确性	122	6.2 软件测试的基本概念	138
5.1.2 算法的时间复杂度.....	123	6.3 软件测试方法	139
5.1.3 算法的空间复杂度.....	126	6.3.1 软件测试方法分类	139
5.2 算法的分析	127	6.3.2 软件测试的步骤	139
5.2.1 最优算法	127	6.4 软件测试用例设计	140
5.2.2 算法的实现	130	6.4.1 白盒测试的测试用例设计	140
思考题.....	132	6.4.2 黑盒测试的测试用例设计	145
第 6 章 软件工程与软件测试 基础.....	135	6.5 软件排错.....	146
6.1 软件工程概述	135	6.5.1 软件排错的原则	146
6.1.1 软件危机	135	6.5.2 软件排错的主要方法	146
		思考题	148
		参考文献.....	150

第 1 章

计算与计算模型

本章主要从计算的角度阐述计算、计算模型、计算思维的基本概念，并以计算思维为切入点，以案例为导向，引导学生感悟、认知、体会现实中的计算与计算思维。

1.1 计算

计算可以说无处不在，从日常生活费、购房费及装修成本的规划和结算到我们熟知的卫星云图、天气预报的运用都涉及计算。如今学科繁多，涉及面广，每个学科都需要进行大量的计算，与古代结绳计数不同的是，当今的计算是指使用计算机来进行大量的统计、处理、转换操作：天文学研究组织需要计算机来分析太空脉冲（pulse）及星位移动；生物学家需要计算机来模拟蛋白质的折叠（Protein Folding）过程，以发现基因组的奥秘；药物学家想要研制抵抗癌症或各类细菌与病毒的药物、医学家想找出防止衰老的新办法；数学家想计算最大的质数和圆周率的更精确值；经济学家要用计算机分析计算在几万种因素考虑下某个企业、城市、国家的发展方向，从而给出正确的宏观调控建议；工业界需要准确计算生产过程中的材料和能源的耗用量，加工与时间配置的最佳方案。由此可见，人类未来的科学，时时刻刻离不开计算。

1.1.1 什么是计算

早在 2500 年前，古希腊数学家、哲学家毕达哥拉斯（Pythagoras，约公元前 572—公元前 501）就说过“凡物皆数”，意思是万物的本源是数、数的规律统治万物。很早以前我国学者认为，对于一个数学问题只有当确定了其可用算盘解算它的规则时，这个问题才算可解。这也是古代中国的算法化思想。它蕴含着中国古代学者对计算的根本问题——能行性问题的理解。这种理解对现代计算学科的研究仍有重要的意义。

由此看来，计算的概念由来已久，但在大众的脑海中“计算”是一个数学概念，人类很早就学会了加、减、乘、除等运算，但直到 20 世纪 30 年代，由于哥德尔、邱奇和图灵等人的工作，人们才对计算的本质有了清楚的理解，进而形成了一个专门的数学分支，即递归论和可计算理论，并因此导致计算机科学的诞生。

简单地说，计算就是依据一定的法则对有关符号串进行变换的过程。进一步理解：从一个已知符号开始，按照一定的规则，一步一步地改变符号串，经过有限步骤，最后得到一个满足预先规定的字符串，这种变换过程就是计算。例如， $1+1=2$ ，就是一个数值计算；两字母和汉字均用二进制数来表示，这种运算称为非数值计算。以汉字“保”为例，通常汉字的转换过程是“输入

码(bao)-国标码(3123H)-机内码(B1A3H)-字形码(保的字形编码)。首先键盘输入拼音“bao”，然后经过变换会得到预先规定的汉字“保”，这就是一个非数值计算。类似地，文字识别、图形图像处理等也都是计算，因为它们都是一种符号变换过程。

1.1.2 什么是计算的本质

根据图灵的研究，直观地说，计算就是计算者（人或机器）一步步改变一条两端可无限延长的纸带上的一串0和1的执行指令，在经过有限步骤后，得到一个满足预先规定的符号串的变换过程。图灵用形式化的方法成功地表述了计算的本质。

抽象地说，计算的本质就是递归。数学家们已经证明，凡是可以从某些初始符号串开始在有限步骤内得到计算结果的函数都是一般递归函数。或者说，凡是可计算的函数都是一般递归函数（计算机科学中，递归函数是一类从自然数到自然数的函数，其在某种直觉意义上是“可计算的”）。至此，人们才弄清楚计算的本质，以及什么是可计算的、什么是不可计算的等根本性问题。例如，若 m 和 n 是两个正整数，并且 $m \geq n$ ，则求 m 和 n 的最大公因子的欧几里德算法，可表示为如下步骤。

- ① 求余数。以 m 除 n 得余数 r 。
- ② 判断余数 r 是否为0。若 $r=0$ ，计算结束， n 即为答案；否则转到步骤3。
- ③ 互换（递推）。把 n 的值变为 m ， r 的值变为 n ，重复上述步骤。

依照以上三个步骤，可计算出任何两个正整数的最大公因子。这时可以把计算过程看成执行这些步骤的序列。可以看出，计算过程是有穷的，而且计算的每一步都是能够机械实现的（机械性）。因此，可以认定：任何两个正整数的最大公因子是可通过计算获得的。

1.1.3 计算与算法

在进行可计算问题研究时，需要用到一个与计算紧密联系的概念，那就是算法。算法也称为能行方法或能行过程，是求解某类问题的方法和步骤，由一组定义明确且能机械执行的规则组成。计算的目的是由算法实现，算法的执行由计算完成。也就是说，计算过程就是执行算法的过程，而算法的过程正好是可以在计算机上执行的过程。从算法的角度讲，一个问题是不是可计算的，与该问题是不是具有相应的算法的答案是完全一致的。

总之，计算或算法的观念在当今已经渗透到宇宙学、物理学、生物学、经济学和社会学等诸多领域。计算已经不仅是人们认识自然、生命、思维和社会的一种普适的观念和方法，还是一种新的世界观。

1.2 图灵机模型

首先了解一下图灵。艾伦·麦席森·图灵（Turing Alan Mathison），1912年生于英国伦敦，1954年逝于英国的曼彻斯特，被誉为计算机科学之父、人工智能之父，如图1-1所示。为纪念图灵在计算机领域的卓越贡献，美国计算机协会（Association of Computing Machinery, ACM）自1966年起设置了“图灵奖”。该奖项用以表彰在计算机科学中做出突出贡献的人。图灵对现代计算机的贡献主要是：建立了图灵机的理论模型，发展了可计算性理论，并提出了定义机器智能的图灵测试。

下面通过图灵机的组成及其基本思想,认识、理解图灵机模型与现代计算机的关系。

1936年,图灵在可计算性理论的研究中,提出了一种抽象的计算机模型——图灵机(Turing Machine)。该机器由以下几部分组成。

1. 一条无限长的纸带

纸带自左至右被划分为一个连一个的格子,每个格子都有相应的编号,并且纸带的右端可根据需要无限延伸,而纸带上的格子可以用于书写符号和运算。

2. 一个读写头

读写头能够读取纸带上某一方格内的信息,并能够在当前格子上书写、修改或擦除数据。

3. 一套控制规则

根据当前读写头所指的格子上的符号和机器的当前状态来确定读写头下一步的动作,并改变状态寄存器的值,令机器进入一个新的状态。

4. 一个状态寄存器

用来保存图灵机当前所处的状态。图灵机所有可能的状态的数目是有限的,其中停机状态是一种特殊状态。

图灵机不是具体的机器,而是一种思想模型,可制造十分简单但运算能力极强的计算装置。该装置可用来计算所有能想象得到的可计算函数。图灵的基本思想是用机器来模拟人们用纸笔进行数学运算的过程。

图灵机被公认为现代计算机的原型,可以读入一系列的0和1。这些数字代表了解决某一问题所需要的步骤,按步骤走下去,就可以解决某一特定的问题。图灵机只用保留一些最简单的指令,而困难的是,如何确定最简单的指令集,怎么样的指令集才是最少的,而且又能顶用,还有一个难点是如何将复杂问题分解为简单指令。

可用一个图灵机来计算其值的函数是可计算函数,找不到图灵机来计算其值的函数是不可计算函数。可以证明,存在一个可以模拟任何其他图灵机的图灵机U,这样的图灵机U称为通用图灵机。在给出通用图灵机的同时,图灵就指出,通用图灵机在计算时,其“机械性的复杂性”是有临界限度的,超过这一限度,就要靠增加程序的长度和存贮量来解决。这种思想开启了后来计算机科学中计算复杂性理论的先河。

总之,图灵机反映的是一种具有能行性的、用数学方法精确定义的计算模型。该计算模型的目标就是要建立一台可以计算的机器,也就是将计算自动化。而现代计算机正是这种模型的具体体现。



图 1-1 艾伦·麦席森·图灵

1.3 停机问题

停机问题(Halting Problem)是目前逻辑学的焦点和第三次数学危机的解决方案,其本质问题是:若给定一个图灵机T和一个任意语言集合S,那么T是否会最终停机于每一个s,其中 $s \in S$ 。该问题的意义相同于可确定语言。显然任意有限S是可判定性的,可列的S也是可停机的。

停机问题的“现实”意义可直观理解为程序不是无所不能的。不存在这样的程序:该程序能够对程序的任何输入都能够正常结束运行(停机)。

通俗地说,停机问题就是判断任意一个程序是否会在有限的时间之内结束运行的问题。如果某一个问题可以在有限的时间之内解决,可以有一个程序判断其本身是否会停机并做出相反的行为,那不管停机问题的结果是什么都不会符合要求,这是一个不可解的问题。

为什么图灵提出停机问题?原因是:根据哥德尔完备性定理,任何可满足的一阶逻辑公式都可以形式推导出来。于是,图灵站了出来,发明了图灵机,并给出了图灵机在一阶逻辑上的定义,将一阶逻辑上的满足性可判定问题成功转化成了图灵机上的停机问题。通过证明停机问题本身就是半可判定的,证明了不存在一个算法能判定任何一阶逻辑公式的可满足性。

“停机问题”是用反证法证明出来的。如同证明 $\sqrt{2}$ 是无理数的判定类似:假定“ $\sqrt{2}$ 是有理数”,然后“得出矛盾”,再反推得出结论。停机问题与理发师问题类似,相关知识可自行查阅了解。

图灵在1936年证明图灵机的停机问题是不可判定的,即不存在一个图灵机能够判定任意图灵机对于任意输入是否停机。图灵机的停机问题是半可判定的。由图灵机的停机问题,可以推出计算机科学、数学、逻辑学中的许多问题是不可判定的。因此,了解图灵机的停机问题,是很有必要的。

1.4 计算思维

随着移动通信、物联网、云计算、大数据等新概念和新技术的出现,信息技术深刻改变着人类的思维、生产、生活、学习方式,且伴随这一进程的全面深入,使得无处不在的计算思维成为人们认识 and 解决问题的能力之一。一个人若不具备计算思维的能力,将在从业竞争中处于劣势;一个国家若不培养其公民的计算思维,则其将在竞争激烈的国际环境中处于落后地位。计算思维,是计算机专业学生应该具备的能力,也是所有大学生应该具备的能力。每一个学生并非都要成为计算机科学家,但是能够正确掌握计算思维的基本方式对于以后的事业发展和学习都是有益的。

1.4.1 科学思维与计算思维

科学思维(Scientific Thinking)是指理性认识及其实现过程,即通过整理和改造将感性阶段获取的大量材料转化为概念、判断和推理以便反映事物的本质和规律的理论体系。进一步理解,科学思维是大脑对科学信息的加工活动。

如果从人类认识世界和改造世界的思维方式出发,科学思维又可分为实证思维、逻辑思维和计算思维三种。

实证思维(Positive Thinking)又称经验思维,是通过观察和实验获取自然规律法则的一种思维方法。它以实证和实验来检验结论正确性为特征,以物理学科为代表。与逻辑思维不同,实证思维需要借助于某种特定的设备来获取客观世界的的数据以便进行分析。例如,洗手液的生产前,需要用不同比例酸碱度试剂借助特定设备进行大量的实验,最终找出合理的减少皮肤刺激配方后,才开始生产。这就是实证思维。

逻辑思维(Logical Thinking)又称理论思维,其通过抽象概括,建立描述事物本质的概念,并应用逻辑的方法探寻概念之间联系。它是以推理和演绎为特征,以数学学科为代表。逻辑源于人类最早的思维活动,而逻辑思维支撑着所有的学科领域。例如,歌德巴赫猜想就是一个逻

辑思维。

计算思维 (Computational Thinking) 又称构造思维, 其从具体的算法设计规范入手, 是通过算法过程的构造与实施来解决给定问题的一种思维方法。它是以设计和构造为特征, 以计算机科学为代表。例如, 今天的物联网、大数据、电子商务等都蕴含着计算思维。

计算思维的本质是抽象和自动化, 特点是形式化、程序化和机械化, 在问题求解、系统设计和人类行为理解方面具有重要的作用。实证思维、逻辑思维和计算思维的一般过程都是对客观世界的现象进行分析和概括而得到认识论意义上的结论。根据分析与概括方式的不同, 上述一般过程可以是推理和演绎、观察和归纳, 也可以是设计和构造。计算思维与实证思维、逻辑思维的关系是相互补充、相互促进的。计算思维相对于实证思维和逻辑思维, 在工程技术领域尤其具有独特的意义。

计算思维和实证思维、逻辑思维一样, 是人类目前为止认识世界和改造世界的三种基本科学思维方式。

1.4.2 计算思维的概念及内涵

2006 年卡内基梅隆大学教授周以真 (Jeannette M. Wing), 如图 1-2 所示。在《美国计算机学会通讯》上发表的《计算思维》(Computational Thinking) 一文对计算思维的定义是: 计算思维是运用计算机科学的基础概念进行问题求解、系统设计以及人类行为理解等涵盖计算机科学之广度的一系列思维活动。

计算思维虽然具有计算机科学的许多特征, 但是其本身并不是计算机科学的专属。即使没有计算机, 计算思维也会逐步发展。但是, 正是由于计算机的出现, 给计算思维的研究和发展带来了根本性的变化, 让计算思维的概念、结构、格式等变得越来越明确, 相关内容也得到不断地丰富和发展。计算机的出现丰富了人类改造世界的手段, 同时也强化了原本存在于人类思维中的计算思维的意义和作用。从思维的角度, 计算机科学主要研究计算思维的概念、方法和内容, 并发展成为解决问题的一种思维方式。

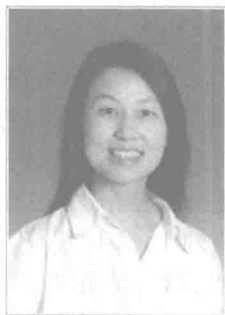


图 1-2 周以真

学生可通过计算机科学基本知识和应用能力的学习来理解和掌握计算思维。计算科学的核心概念可以分为计算、通信、协作、记忆、自动化、评估和设计。所以, 针对上述定义, 计算思维的学习包括三要点: 求解问题中的计算思维、设计系统中的计算思维及理解人类行为中的计算思维, 下面并分别予以阐述。

1. 求解问题中的计算思维

利用计算手段求解问题的过程是: 首先把实际的应用问题转换成数学问题, 然后建立模型、设计算法并编程实现, 最后在实际的计算机上运行求解。

归根结底, 就是在求解问题时, 首先想到的是将该问题的求解过程转换为利用计算机实现的过程。今天的大数据可以体现求解问题中的计算思维。例如, 相当一部分农民的生产种植还停留在面朝黄土背朝天的阶段, 农业种植受环境、技术资源等因素的限制, 没有得到充分发展, 属于粗放型。但若把大数据引入现代农业, 对田间实行实时监测, 为每一种作物、每一块土地建立数据模型、种植模型, 让种植者严格按照环境和作物的生长规律进行种植, 那么, 此时的农业就属于精准型的农业, 这就是计算思维在农业种植上的体现。

2. 设计系统中的计算思维

任何自然系统和社会系统都可视为一个动态演化系统，而演化伴随着物质、能量和信息的交换，这种交换可以映射为符号变换。当动态演化系统抽象为离散符号系统后，就可以采用形式化的规范来描述，通过建立模型、设计算法和开发软件来表达、模拟、控制系统的演化，揭示演化的规律。

例如，全国的航模比赛中，由于空气湿度是不断变化的，那么，设计该系统时首先把空气湿度划分成不同的区间，然后对不同区间的湿度设定特定的符号或数字来代表。这样就把动态演化系统抽象为离散的符号系统，接着就可以通过试验，确定飞行速度，找准目标位置。也就是，首先建立模型，然后根据该模型设计问题求解算法，最后通过计算机软件来实现模型系统的功能。由此看来，今天我们做的很多事情都潜移默化地利用了计算思维。

3. 理解人类行为中的计算思维

理解人类行为中的计算思维是利用计算手段来研究人类的行为，通过设计、实施和评估人与环境之间的交互，来研究人们之间的交互方式、社会群体的形态及演化规律等问题。

例如，今天你上网购买或查阅了某类产品，明天打开网页时会自动弹出与之相关的网页超链接，这便是当今人们在理解人类行为中的计算思维的体现。由此我们会联想到“网络爬虫”这一网络用语。

网络爬虫的产生是有其相应背景的。随着网络应用迅猛发展，万维网成为大量信息的载体，如何有效地提取并利用这些信息成为一个巨大的挑战。搜索引擎，如传统的通用搜索引擎 Yahoo 和 Google 等，作为一个辅助人们检索信息的工具成为用户访问万维网的入口和指南。但是，这些通用性搜索引擎也存在着一定的局限性，主要体现在以下几个方面。

(1) 不同领域、不同背景的用户往往具有不同的检索目的和需求，通用搜索引擎所返回的结果包含大量用户不关心的网页。

(2) 通用搜索引擎的目标是尽可能大的网络覆盖率，有限的搜索引擎服务器资源与无限的网络数据资源之间的矛盾将进一步加深。

(3) 万维网数据形式的丰富和网络技术的不断发展，图片、数据库、音频、视频多媒体等不同数据大量出现，使通用搜索引擎往往对这些信息含量密集且具有一定结构的数据无能为力，不能很好地发现和获取。

(4) 通用搜索引擎大多提供基于关键字的检索，难以支持根据语义信息提出的查询。

为了解决上述问题，定向抓取相关网页资源的聚焦爬虫应运而生。聚焦爬虫将目标定为抓取与某一特定主题内容相关的网页，所有被爬虫抓取的网页将会被系统存储，进行一定的分析、过滤，并建立索引，以便之后的查询和检索。这就是为什么当我们利用网络进行信息搜索、网上购物之后，再次上网时，与之相关的网页超链接会自动弹出的原因。这也是理解人类行为中的计算思维的例证。

在维基百科 (Wikipedia) 中，计算思维被解释为一种新的计算机科学技术广泛使用的问题求解方法，可利用算法高效率地求解大规模复杂问题。

美国科学家斯蒂芬·沃尔夫勒姆 (Stephen Wolfram) 在他的科学巨著《一种新科学》(A New Kind of Science) 中指出，传统的科学建立在数学基础上，新的科学建立在计算机程序上。

孙家广院士在《计算机科学的变革》一文中指出，计算机科学界最具有基础性和长期性的思想是计算思维。

1.4.3 计算思维的本质

计算思维的本质是抽象和自动化（程序、算法）。计算思维可以用“抽象”“算法”来概括，也可用“合理抽象”“高效算法”来概括。

1. 抽象

对物理世界进行建模和模拟，把物理世界的变化解释成某种计算过程，并以形式化的方式表达出来。

2. 合理抽象

懂得计算的能力和极限，知道哪些问题可以计算，哪些问题不可以计算。同时，把待解决的问题抽象成有效的计算过程，建立有效的计算模型。

3. 算法

算法是指解题方案准确而完整的描述，是一系列解决问题的清晰指令或步骤。

4. 高效算法

解决同一个问题有不同的算法，如何设计计算过程在最短时间内正确可靠地完成。

计算思维中的抽象完全超越物理的时空观，并完全用符号来表示。与数学和物理科学相比，计算思维中的抽象显得更为丰富，也更为复杂。在计算思维中，所谓抽象就是要求能够将问题抽象并进行、形式化表达（这些是计算机的本质），使设计的问题求解过程达到精确、可行，并通过程序（软件）作为方法和手段对求解过程予以“精确”地实现。也就是说，抽象的最终结果是能够机械地一步步自动执行。

1.4.4 计算思维能力

计算思维能力是指建立起利用计算机技术解决问题的思路，并理解问题的可求解性。

计算思维能力的核心是求解问题的能力，包括发现问题、寻求解决问题的思路、分析比较不同的方案、验证方案。

计算思维能力的关键是问题抽象、模型建立、算法设计和实现。

计算思维能力的培养是指深入掌握计算机解决问题的思路，更好地用好计算机，把计算机处理问题的方法用于各个领域，推动在各个领域中运用计算思维，更好地与信息技术相结合。

中国科学院计算技术研究所研究员徐志伟总工程师认为，计算思维是一种本质的、所有人都必须具备的思维方式，就像识字、做算术一样；在2050年以前，让地球上每一个公民都应具备“计算思维”的能力。

在普适计算的今天，人们的计算思维能力必将得到快速发展和提高，但真正的能力是“练”出来的，而不是“教”出来的，因此，计算思维能力需要在不断实践中提高。

1.5 计算在其他学科的应用

1.5.1 计算社会学

2009年2月，美国哈佛大学大卫·拉泽（David Lazer）等15位美国学者在《科学》（*Science*）杂志上联合发表了一篇具有里程碑意义的文章《计算社会学》（*Computational Social Science*）。该

文指出,“计算社会学”这一学科正在兴起,人们将在前所未有的深度和广度上收集和利用数据,为社会科学研究服务。

计算社会学是社会学的一门分支,其使用密集演算的方法来分析与模拟社会现象。计算社会学通过计算机模拟、人工智能、复杂的统计方法以及社会性网络分析等新的途径,由下而上地塑造社会互动的模型,是发展与测试复杂社会过程的理论。

计算社会学主要研究网络、群体。总有一天,计算社会学不仅可以得到人与人之间的规律,而且可以预测整个社会未来的发展趋势。这些趋势包括经济规模、社会制度等。

从我国“一带一路”发展战略体会计算社会学。“一带一路”(One Belt and One Road, OBAOR 或 OBOR)是指“丝绸之路经济带”和“21 世纪海上丝绸之路”。“一带一路”不是一个实体和机制,而是合作发展的理念和倡议,是充分依靠中国与有关国家的双多边机制,借助既有的、行之有效的区域合作平台(可以理解为社会网络),旨在借用古代“丝绸之路”的历史符号,高举和平发展的旗帜,积极主动地发展与沿线国家的经济合作伙伴关系,共同打造政治互信、经济融合、文化包容的利益共同体、命运共同体和责任共同体(可以理解为:塑造社会互动的模型)。“一带一路”规划,被认为是“中国版马歇尔计划”的战略载体,实质是,在通路、通航的基础上通商,主要影响铁路、航空、航海、农业、商贸流通、油气进口等行业。

1.5.2 计算生物学

计算生物学(Computational Biology)是生物学的一个分支。根据美国国家卫生研究院(National Institutes of Health, NIH)的定义,它是指开发和应用数据分析及理论的方法、数学建模和计算机仿真技术,用于生物学、行为学和社会群体系统的研究的一门学科。

当前,生物学数据量和复杂性不断增长,每 14 个月基因研究产生的数据就会翻一番,单单依靠观察和实验已难以应付。因此,必须依靠大规模计算模拟技术,以从海量信息中提取最有用的数据。

计算生物学的研究内容主要包括以下几个方面。

(1) 生物序列的片段拼接。

(2) 序列对接。

(3) 基因识别。人类长达 30 亿的 DNA 序列中只有 3%~5% 是基因,阐明人体中全部基因的位置、结构、功能、表达等。与之相关的一个重要应用就是模拟基因表达数据集。

(4) 种族树的建构。

(5) 蛋白质结构预测。任意给一段蛋白质序列,生物学家虽然可以用传统的生物学方法求出其结构,但这不但成本高而且费时,而计算生物学的蛋白质结构预测工具通过序列分析可以直接得出其结构,如 CYTO 为人类 T 细胞中的因果蛋白质信号网络。

(6) 生物数据库。生物学数据量不断增长,每 14 个月基因研究产生的数据就会翻一番,观察和实验已不能满足需求。这时,生物数据库显示了强大的威力。例如, CATH 蛋白结构分类数据库、果蝇交互数据库。随着科学技术的发展,计算生物学的应用也越来越广泛,如对生物等效性的研究、皮肤的电阻、骨关节炎的治疗、哺乳动物的睡眠等。

以上,仅仅介绍了计算社会学、计算生物学,而类似这样的新兴学科还很多。我们可以看出,这些学科都与计算科学密不可分,都属于交叉学科。这些学科与计算、可计算性、计算思维的基本理论与思维方式是紧密融合的。因此,掌握和理解计算思维的内涵,学会常规算法的设计是每个人都应该拥有的基本素质。

思考题

1. 什么是计算？什么是可计算？
2. 阐述图灵在计算机理论发展中的主要贡献。
3. 阐述图灵机模型中的主要组成部分及作用。
4. 简述什么是计算思维。
5. 阐述你对计算思维的理解。

第 2 章

计算机基础知识

电子数字计算机是 20 世纪最重大的科技成就之一。自 1946 年第一台电子计算机问世以来, 计算机得到迅速发展, 并已广泛应用于工农业生产、科学研究、国防及人们日常工作和生活的各个领域。伴随人类进入 21 世纪, 以高科技为支撑的信息化社会已经到来, 掌握计算机知识和应用能力已成为现代人一个重要标志。计算机技术的进一步发展和应用必将对社会发展和人类文明产生更大的促进作用, 也将对社会政治、经济、文化和人类生活的各个方面产生巨大而深远的影响。

2.1 计算机组成与工作过程

2.1.1 计算机简介

计算机 (computer) 是一种用于高速计算的电子计算机器, 可以进行数值计算, 也可以进行逻辑计算, 还具有存储记忆功能, 是能够按照程序运行, 自动、高速处理海量数据的现代化智能电子设备。计算机由硬件系统和软件系统所组成, 没有安装任何软件的计算机称为裸机。

2.1.2 计算机的组成

一个完整的计算机系统, 包括硬件系统和软件系统两部分, 其组成如图 2-1 所示。

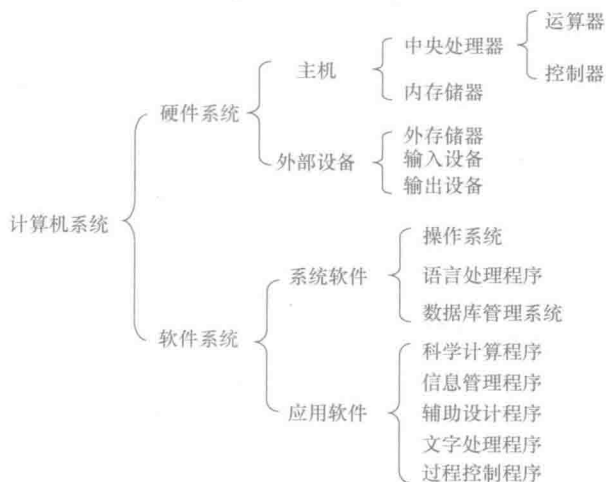


图 2-1 计算机系统的组成

硬件系统是指由电子电路和机械装置所构成的计算机实体，是计算机系统的物质基础，如CPU、存储器、输入设备、输出设备等。只有硬件系统的计算机又称为裸机。裸机只能识别由0和1组成的机器代码。没有软件系统的计算机几乎是没有用的。

软件系统是为运行、管理和维护计算机而编制的各种程序的总称。实际上，用户所面对的计算机是经过若干层软件“包装”起来的计算机。计算机的功能不仅仅取决于硬件系统，更大程度上是由所安装的软件系统所决定。

当然，在计算机系统中，对于软件和硬件的功能没有一个明确的分界线。软件实现的功能可以用硬件来实现，称作固化或硬化，例如微机中的ROM芯片中就固化了系统的引导程序。同样，硬件实现的功能也可以用软件来实现，称为硬件软化，例如，在多媒体计算机中用于处理视频信息（包括获取、编码、压缩、解压缩和回放等）的设备视频卡，一般通过软件来实现。

对某些功能，是使用硬件实现还是软件实现，取决于系统价格、速度、所需存储容量及可靠性等因素。一般来说，同一功能用硬件实现，可速度快、减少所需存储容量，但灵活性和适应性较差，而且成本高；用软件实现，可提高灵活性和适应性，但通常以降低速度作为代价。

2.1.3 计算机系统结构

1945年美籍匈牙利科学家冯·诺依曼（Von Neumann）提出了一个“存储程序”的计算机方案。这个方案包含3个要点。

- ① 采用二进制数的形式表示数据和指令。
- ② 将指令和数据存放在存储器中。
- ③ 计算机硬件由控制器、运算器、存储器、输入设备和输出设备等五大部分组成。

“存储程序”工作原理的核心是“程序存储”和“程序控制”，就是通常所说的“顺序存储程序”概念。我们将按照这一原理设计的计算机称为“冯·诺依曼型计算机”。图2-2描述了“冯·诺依曼型计算机”硬件系统的五大功能部件。图中→表示控制流，↔表示数据流。

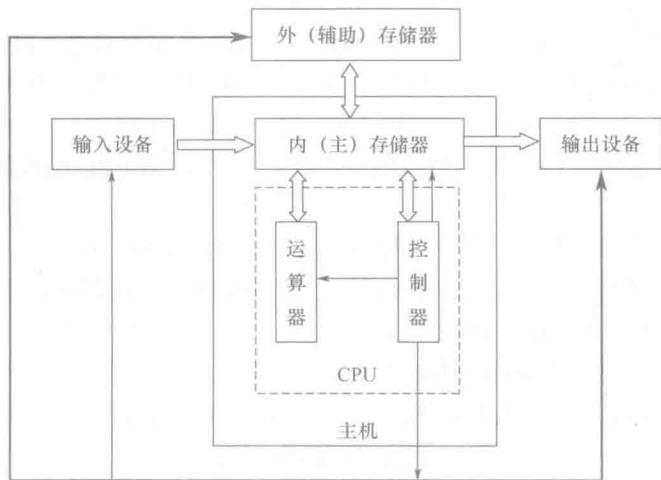


图 2-2 硬件系统中五大功能部件

1. 运算器

运算器是用来进行算术运算和逻辑运算的部件，主要由算术逻辑单元和一组寄存器构成。它对取自内存储器或寄存器中的数据进行算术或逻辑运算，再将运算结果送到内存储器或寄存器。