

“计算机科学技术创新”科普教育丛书

ASCII

信息王国中的代码

董百年 等 编著 / 吴洪来 吕传兴 审



清华大学出版社

<http://www.tup.tsinghua.edu.cn>

信息王国中的代码

董百年 等 编 著
吴洪来 吕传兴 审



A1008594

清华大学出版社

(京)新登字 158 号

内 容 简 介

在信息时代中,人们经常是向计算机输入信息,然后再由计算机输出处理后的信息。

本书从什么是信息、信息有什么用谈起,讲到信息要交流、处理就必须将其代码化,进而要用计算机对其进行处理还必须数字化,最后讲到计算机如何处理这些数字化的信息。全书以“信息→代码化→数字化”为主线,讲述计算机处理信息中信息表示的问题。

本书语言通俗易懂,文句流畅,并配有生动有趣的插图,是一本适合广大青少年及计算机爱好者阅读的科普读物。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: 信息王国中的代码

作 者: 董百年 等 编著

出版者: 清华大学出版社(北京清华大学学研大厦,邮编 100084)

<http://www.tup.tsinghua.edu.cn>

印刷者: 世界知识印刷厂

发行者: 新华书店总店北京发行所

开 本: 787×960 1/16 **印张:** 14.25 **字数:** 258 千字

版 次: 2002 年 9 月第 1 版 2002 年 9 月第 1 次印刷

书 号: ISBN 7-302-04586-0/TP·2718

印 数: 0001~5000

定 价: 19.90 元

“计算机科学技术创新”科普教育丛书

编委会名单

主 编：李三立

副主 编：吕传兴 吴洪来

编 委 （按姓氏笔画为序）：

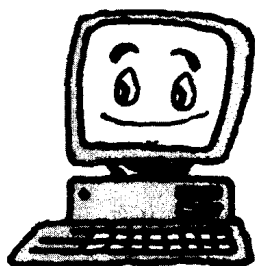
毛国平	石 磊	叶金霞	孙元清	苏芳来	张 权
张世正	张令毅	林奇清	陈海洋	陈春法	周卓伦
郑增仪	郭 鸿	郭善渡	唐 玲	徐桂珍	高黎新
董百年	蒋敦杰	蔡建民			

执行编委：

吕传兴 吴洪来 俞嘉惠 缪淮扣 薛维明

邓小平同志早在 1984 年就提出,“计算机普及要从娃娃抓起”,这是一个非常有战略远见的思想。现在看来,将计算机信息技术课作为中学阶段、首先是高中阶段的必修课势在必行,中小学计算机(信息技术)教育的优劣将直接影响到我国 21 世纪的经济发展和科技竞争实力,会影响到一代人甚至几代人的终身学习和发展。从国内实际情况看,虽然我们做了不少卓有成效的工作,但仍然赶不上时代前进的步伐,与发达国家的差距甚至还在拉大。我们必须奋起直追,争取在一段较短的时间内使我国的中小学计算机教育上一个大的台阶,直至成为世界最好的中小学计算机教育的国家之一。

我国是一个人口众多的发展中国家,这一客观条件决定了我国在青少年信息技术普及过程中应该有其自己的特色,在“教什么”和“怎么教”的问题上应与西方发达国家有所不同,不能一成不变地照搬外国。必须自己组织力量,下苦功夫,编写出一套甚至一批适合我国青少年需要的信息技术科普读物和教材!这是一项难度很大的工作,但又是一项意义重大、影响深远的工作。如果做得好,将给我国亿万中小学生的学习和发展建造起一个扎实的平台,将会使他们在基础教育阶段就能对计算机与信息科学的基本原理和发展动态有一个正确的理解,建立起良好的信息意识和信息素养。信息科学与技术的发展史就是一部激动人心的创新史,把信息科学家们的创新故事告诉我们的孩子们,把强烈的创新意识和



创新精神根植于他们的心中,也是我们的一个非常重要的任务。这些就是组织编写这套书的来由和初衷。

我赞成编委会提出的“丛书以激发和培养读者(特别是青少年)的创新精神为主旨”,要通过知识的重新整合,深入浅出地讲清道理,力求通俗易懂;要写薄书,写浅书。关于丛书内容,我认为应以中学生主要关心的基本问题为主,如什么是计算机,为什么计算机有如此强大的功能,计算机为什么能联网,网络为什么有如此巨大的潜力等等。这些问题大体上也是成年人学习计算机和信息科学时提出的,这些问题能解决也都是科学创新的结果。

万事开头难。将这些重大的发明和创新过程进行重新研究和整合,以准确的事实、清晰的概念、浅显的文字、生动的插图展现出来,实非易事。对于参与组织编写这套书的专家学者们所做的工作,以及他们承担这样一个艰巨的任务所表现出来的强烈的历史责任感和钻研精神,我感到十分敬佩;同时还要感谢上海市计算机学会和清华大学出版社,对于他们能够花大力气来做这样一件有意义的事,我感到十分高兴。尽管首批8本书中还存在一些不足之处,但毕竟是走出了第一步。我希望这套书能一直做下去,并不断吸引全国热心于科普事业的专家、学者参与进来,不断修订、更新、扩充,精益求精,使我们的青少年读者能从中真正受益。

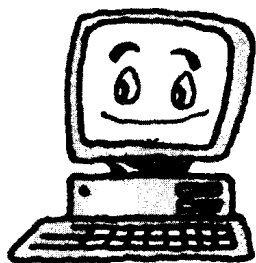
教育部副部长

2000年11月

尽管从第一台电子计算机诞生至今只有 50 多年,但计算机却已经走入了我们生活的每一个角落。计算机知识已成为人类重要的文化基础,计算机科学技术一跃成为推动人类社会向前发展的最活跃、最积极的动力。科学技术发展的灵魂在于创新。计算机科学技术的发展之所以气势磅礴、一日千里,其根本的原因就在于这一领域中人类创新意识的空前宏扬和创新精神的淋漓尽致的发挥。因此,计算机科学技术的发展史本身也是一本最精彩的创新精神的教科书。

国家把希望寄托在年轻一代身上。教育部领导十分重视向青少年一代普及计算机科学技术的基础知识,培养青少年的创新精神,认为这是一项十分重要而且迫切的任务。吕福源同志提出要编写一套适合中学生阅读的“计算机科学技术创新”科普教育丛书,并在百忙中亲自主持召开作者座谈会,反复强调丛书要以宏扬创新精神为主线;在内容编排上要将计算机科学的有关知识进行重新整合,要围绕计算机科学的几个最基本、最重要的问题展开。关于丛书的具体编写,他也提出了许多具体意见,如丛书一定要做到图文并茂、通俗易懂、重点突出;在叙述前人的创新事迹时,同时也要指出其局限性等。

这些意见给丛书编写工作指明了方向。五位执行编委经多次研讨,以软件技术的进步、计算机体系结构的演变、人机界面的改进、代码技术的发展、网络与信息高速公路等方面的有关知识、技术和创新事迹为重



点,拟定了八个题目。我们认为,通过这八个题目的讲述,应能使读者对计算机科学技术的基础知识有一个较为全面的了解。在写作过程中,我们力求言必有据,概念准确,把计算机科学技术的基础知识和计算机发展史上的创新火花穿插在一起,用通俗易懂的语言,生动有趣的情节和插图展现给读者,使读者的思绪同计算机科学技术大师们的灵感与智慧一起涌动,从而能够潜移默化地学到知识,自然而然地把创新精神融化于自己的思想和行动之中。

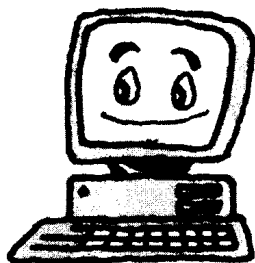
来自全国各地的计算机专家、教授和教育工作者参加了本套丛书的编写。经过两年的艰苦努力,终于使这套丛书得以同广大读者见面了。如果丛书能在普及计算机科学知识,启迪青少年的创新精神,激发青少年深入学习计算机科学技术知识的兴趣等方面有所收获,我们将会感到无比欣慰。

在丛书的编写过程中,上海市计算机学会和清华大学出版社在各方面都给予我们很大的帮助;上海科诺科技服务公司协助我们组织了插图创作队伍,在此一并致谢。

此次付印的书稿虽经反复修改,但错误与疏漏仍在所难免,诚恳希望广大读者及学界前辈不吝指正。

丛书编委会

2000年12月



什么是信息？人类已经在不知不觉中同信息打了几万年的交道。我国古代著名军事家孙子可以算是最早认识信息重要性的一个。事实上无论古代、现代，信息都很重要。香农(Shannon)第一个把关于信息的学问系统化、理论化，提出了“信息论”。随着计算机的出现，信息技术得到了巨大的发展，人类步入了信息社会。

信息本身是看不见、摸不着的，所以要利用它，就必须用某种方式把它表现出来，这就是代码。“代码”的含义非常广，凡是用于表示某个信息的东西都可以称为代码。从古到今，人类发明了各种各样的“代码”。近代社会，人们逐渐探索出一种新的“代码”——数字化代码(布尔代数、二进制数字)。

信息有一个特点就是共享，而人类对信息的认识和利用离不开信息的传播。制定代码的规范和标准，对于信息的传播是很重要的，秦始皇颁令“书同文”是一次很成功的书写文字规范化的改革。人类文明史中曾经出现过许多种计数的符号——数码，到了16世纪前后，差不多都被源自印度的一种现在都称之为阿拉伯数字的符号所取代，形成了一个非常好的数字表示规范。规范化不但有利于信息的传播，而且有利于信息的处理。在信息社会中，信息的共享达到了空前的程度，所以标准就显得越来越重要了，无论是电报中的摩尔斯码，还是计算机中的交换码ASCII, GB2312等等，都在信息处理中起着非常重要的作用。

现在的计算机已经能处理字符、声音、图像等多媒体信息，但其内部最终还是使用布尔代数和二进制代码，所以要利用计算机处理各种信息，前提是信息代码要数字化，也就

是编码。字符的数字化相对比较简单。而声音、图像视频信息的数字化则是近二三十年计算机应用技术发展的核心问题之一。

随着计算机的广泛应用,所需要处理的信息量也越来越大,特别是数字化的声音、图像要占用巨大的存储空间,使得数据的压缩成为必要。压缩后的数据不能直接使用,因此必然有一个相应的解压缩的问题。压缩方法好坏的标准是什么?有哪些通用压缩方法?

计算机网络的出现使得信息的共享和传播更加方便,但有的信息只能和一部分人共享,对其他不相干的人保密。如何防止信息代码在传播、存储过程中可能产生的错误,这就是信息的安全问题。从编码的角度出发看信息安全,就提出了检测码、纠错码、加密、解密等技术。

汉字,是中华民族的文化瑰宝。中国要进入信息社会,必须解决如何用计算机处理汉字问题,其关键是汉字的编码问题。汉字的输入曾经是华夏子孙使用计算机的一个瓶颈问题,但经过多年的实践,人们已不仅能用键盘输入汉字,而且还可以用手写和声音输入。

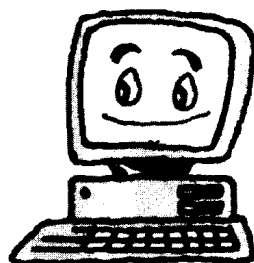
本书以计算机技术中的“信息——编码——数字化”为主线,介绍了上述相关技术及其发展创新历程。

本书的第1,3章由董百年执笔,第2,4,5,7章由蒋冠雄执笔,第6章由黄松英执笔,董百年负责全书的统稿。

本书从立题开始到确定提纲,收集素材,初稿形成到现在的样子,自始至终得到吕传兴、吴洪来两位老师耐心的指导,从总的框架到文字上的修改意见,他们为本书付出了大量的精力。笔者愿借此机会表示感谢和敬意。顾家琪为本书创作和绘制了插图,使本书增色不少,在此一并致谢。

作 者

2001年2月



第 1 章	啊,奇妙的信息世界	1
1.1	信息 = X	1
	知彼知己、百战不殆	3
	借东风	4
	捕捉信息,翻新产品	6
	望、闻、问、切——中医诊断的手段	7
1.2	信息社会中计算机的作用	9
	诊断疑难病的利器	9
	石油勘探和大范围天气预报	11
	战场上“一盎司硅”的作用大于“一吨铀”	14
	在教育领域里大显身手	16
	在金融业中大出风头	18
	电子商务	20
1.3	认识“信息”	22
	信息到底是什么?	22
	信息论和信息科学	25
	信息的特点	27
	信息的载体	29
第 2 章	信息的代码技术	31
2.1	丰富多彩的信息代码	31

非语言信息代码	32
语言和文字的诞生	34
数字化信息代码	36
逻辑代数	41
2.2 信息代码的存储	44
早期的存储器	44
纸张和印刷术的发明	47
奇妙的缩微存储	49
声像代码存储	50
计算机中的存储器	53
2.3 信息代码的传递	56
古人信息传递法——费时费力	56
电信时代的到来——方便快捷	60
从模拟通信到数字通信	62
第3章 代码规范(标准和协议)	68
3.1 从秦始皇的“书同文”说起	68
温州方言成了一种“新密码”	69
汉字的形成	69
秦始皇下令：全国文字统一	71
3.2 记数法的形成及阿拉伯数字的推广	72
各国古代记数法	72
优胜劣汰，全世界流行了源于印度的阿拉伯数字	74
十六进制数简介	75
3.3 计算机里的 ASCII 码和交换码	77
早期计算机尚无“规矩”可言	77
计算机发展之初就呼唤代码标准	78
ASCII 码成了事实上的国际通用编码	78
现在计算机内部都用二进制码	79
3.4 分久必合——多媒体信息处理标准的问世	80

	从胶卷、录音带到录像带	80
	从 VCD 到 DVD	81
	DVD 标准之争	81
3.5	因特网简介	82
	初期的计算机网络	82
	一大批协议几乎是与 Internet 同步问世	83
第 4 章	“万物皆数”——代码数字化	86
4.1	一切都向数字化迈进	86
	多媒体电脑	86
	最初的计算机只是为了数值计算	88
	数学的力量促进了现代的科学研究	90
	解决问题的前提是代码数字化	93
4.2	“数”中自有书万卷	96
	超级图书馆和字符数字化	96
	电报中的字符编码	98
	计算机中的字符编码	100
	和计算机交流字符信息	105
4.3	“数”中自有天籁乐	107
	能力有限的电脑	108
	变化无穷的声音	111
	声音采样的技术	113
	CD 中的数字化音乐	119
	另辟蹊径的 MIDI 音乐	121
	声卡的作用	122
4.4	“数”中自有颜如玉	123
	“真我本色?”——计算机颜色系统	124
	“图不厌精”——计算机图像分辨率	128
	图像和数码之间的转变	130
	另一种图像的数字化	131

计算机视频中的“动静统一”	133
第 5 章 “能屈能伸”——代码压缩	136
5.1 日益膨胀的数据量	136
信息社会信息多	137
数字化声像数据多	138
虽快仍慢的网络	139
5.2 代码也能压缩?	141
如果能够压缩	141
为何能够压缩?	142
如何评价压缩方法	144
5.3 通用压缩方法	148
最简单的方法	148
也可以“长”话短说	151
“决不当头”的前缀码	153
5.4 声像代码压缩	157
压缩音频信号	158
压缩静态图像	161
压缩动态图像	164
VCD 中的动态图像压缩	165
第 6 章 “伪装揭秘”——计算机密码	169
6.1 密码趣谈	169
密码介绍	170
战争促进了密码技术的发展	173
经济领域中的密码技术	174
6.2 密码学基础	175
密码技术发展过程	175
密码技术的设计原则、设计要求、设计范畴	176
6.3 现代密码技术简介	177

密码系统原理	177
传统密码体制	177
公开密钥加密体制	178
6.4 计算机安全中的密码技术	180
密码技术中的新挑战	180
数字签名	180
认证 / 公证技术	181
第 7 章 “文化瑰宝”——汉字的数字化	183
7.1 汉字与计算机	183
第五大发明——汉字	183
汉字的特点	186
国家标准——区位码	187
7.2 趋于统一的汉字内码	189
难点在中西文共存	189
麻烦在两岸不统一	192
前途在多文种编码	194
7.3 形形色色的汉字外码	196
“艰苦卓越”——早期的输入法	197
“万码奔腾”——目前的输入法	200
“天外有天”——未来的输入法	206



啊,奇妙的信息世界

1.1 信息 = X

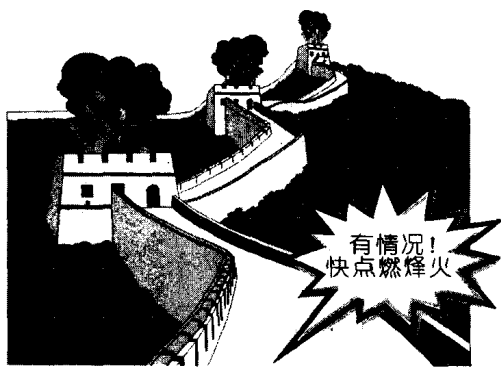
翻开形形色色的报纸,选看五光十色的电视频道,步入色彩斑斓的网中天地……到处都可以看到关于信息重要性的论述,如“信息=商机”,“信息=战机”,“信息=金钱”,等等。

这就让人要思考:信息究竟是什么?为什么今天人们把信息看得如此重要?

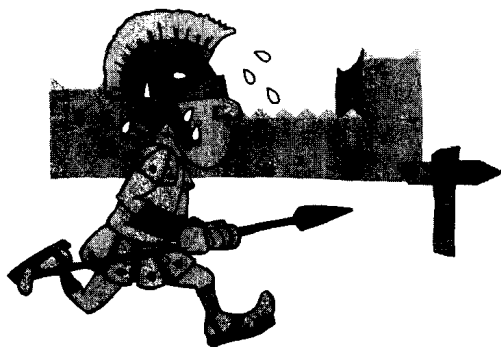
我国的《辞海》(1999年版)里关于“信息”有两种解释:(1)音讯;消息。(2)通信系统传输和处理的对象,泛指消息和信号的具体内容和意义。

《辞海》中的解释准确,但比较深奥,还是让我们先来看看自然界的动植物。你可能听说过一种植物,它的名字叫含羞草。当你用手轻轻地碰一下它的叶片,或用嘴轻轻地往它的叶片上吹一口气,它的叶片会害羞似地慢悠悠合拢起来。如果你触得重,它的动作反应也快,不到十秒钟,所有的叶片全部都合拢了。还有一种叫猪笼草的能吃小虫的植物,它会把飞到它的花里面的小昆虫卷在中间,并分泌出一种液体把“猎获物”消化吸收掉。这些植物能感觉到某种信息,并会做出相应的反应。

在动物界,则更有趣,如果你观察过蚂蚁觅食,一开始蚁群会派出



烽火

希腊士兵从马拉松前线
向雅典报捷

一批“侦察兵”，当它们在路上相遇时，相互之间一定有肢体或触须相碰，其作用就是在交换信息，不久也许你就会看到有大批蚂蚁有目标地开始行动，把所找到的食物搬回蚁穴。有的时候，你也许能看到蜜蜂在飞舞中的轨迹图形是各式各样的，据研究，其中有的蜜蜂其实是在为与其伙伴交换信息而“舞”。

动植物尚且如此，人类社会在生存、发展中就更离不开信息。

让我们从远古时代说起。为了尽快传递消息，在驿站出现之前，史书就记载了“烽燧”（即烽火，又因是烧狼粪而称为狼烟）。平时，每隔一定距离修筑一个高台——烽火台，当边疆有外敌来犯时，烽火台立即点燃起烽火，相邻的烽火台一见到有烽火点燃了，马上也把自己台上点燃狼烟示警。这样一来，消息就一级一级地传递过去，前方有来犯之敌的信息很快就传开了。这里要事先有一个约定：把烽火台上的烽火点燃与否作为有无敌情的信号，也就是规定了：烽火点燃则表示有敌情，而烽火没有点燃则表示平安无事。这样靠烽火台就可快速传递敌情消息。用现代信息技术观点可把这种模式看成是最早一种二态信息代码及其传递，它也可算是简化信息使之便于传递、处理的最早的范例。有名的“烽火戏诸侯”发生于商周年代（公元前一千多年），可以推测，至少在此之前应该已经有了“烽火”。

按时间算，在发生“烽火戏诸侯”之后好几百年，在古希腊的马拉松平原一带，希腊人与来犯的波斯人进行了一场激战，希腊最终战胜了入侵之敌。胜利之时，在马拉松前线的希腊统帅，派了一位名叫菲迪皮茨的士兵跑了四十多公里，向雅典报捷，这个士兵在完成了送递战场获胜的喜报的任务后，献出了他年轻