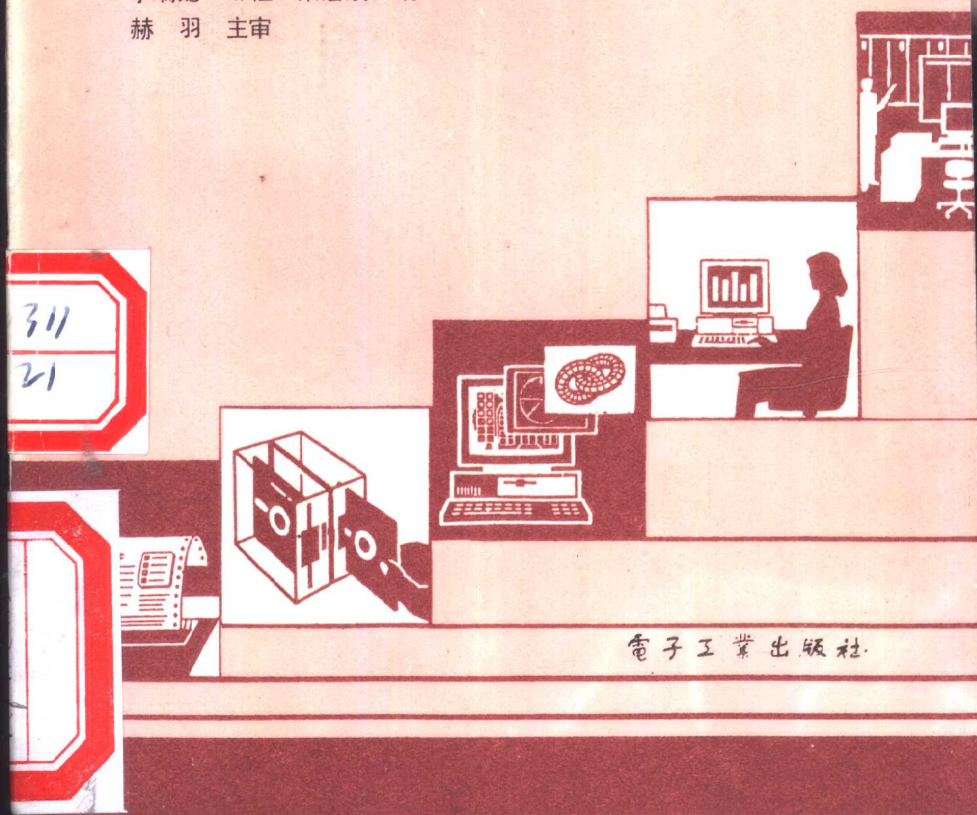


职业高中教材 (计算机技术专业)

微型计算机

汉字处理与录入

李瑞龙 邢程 宋宏斌 编
赫 羽 主审



电子工业出版社

内 容 简 介

本书是职业学校计算机技术专业教材,全书共分四章,主要讲述微型计算机汉字处理的基本原理,汉字录入技术,汉字编辑制表软件的使用等。

本书遵循应用原则,力求逻辑严谨,深入浅出,通俗易懂。它不仅是中等职业技术教育的专业教材,也可以作为计算机文字处理培训班的教材及一般文秘人员的自学读物。

微 型 计 算 机
汉 字 处 理 与 录 入
李瑞龙 邢程 宋宏斌 编
赫羽 主审
责任编辑 吴明卒

电子工业出版社出版(北京海淀区万寿路)
电子工业出版社发行 各地新华书店经销
电子工业出版社计算机排版室 排版
北京顺义李史山胶印厂印刷

*

开本:787×1092毫米 1/32 印张:3.125 插页:1 字数:70千字

1990年9月第一版 1990年9月第一次印刷

印数:1—15000册 定价:1.50元

ISBN 7-5053-1125-5/TP·185

出版说明

根据 1986 年全国职业技术教育工作会议关于“职业技术教育管理职责暂行规定”的分工精神和国家教委的要求,为了满足职业高中、职业中专等职业学校的迅速发展对教材的需要,我部组织了职业学校电子类教材的编审与出版。通过全国部分省、市及参与编审工作的有实践经验的教师,从事电子技术工作的工程师,职业教育研究工作者的共同努力,已编审出版的教材深受职业学校师生的欢迎。我们在总结前段工作的基础上,为推进该层次的教材建设,成立了有北京、上海、天津、江苏、浙江、山东、辽宁、吉林、黑龙江、河北、河南、四川、新疆、甘肃共十四个省、市、自治区的教师和职教主管部门领导参加的职业学校电子类教材工作领导小组和编审委员会,制订了“实用电子技术”及“计算机技术”两个专业的参考性教学计划和 1988~1990 年教材出版规划。根据教学计划的需要,列入规划的教材共 23 种。

我们组织编写的这套教材,是以实用电子技术和计算机技术专业的教学计划为依据。为突出职业学校着重职业技能训练的特点,侧重于教材的实用性、科学性以及增强学生实验和操作技能训练的内容。为适应各地电子工业发展的需要,教材除注意基础知识外,也适当反映了电子专业的现代技术。另一方面,由于电子类专业分支多,教材编写还立足于宽口径,以方便不同专业选用。

编写职业学校教材是一个新课题,经验不足,希望全国电子类职业学校广大师生积极提出批评建议,共同为进一步提高教材质量而努力。

机械电子工业部电子类专业教材办公室
一九八八年十二月

职业学校电子类教材工作领导小组

组 长：姚志清 （以下以姓氏笔划为序）
副组长：王世华 孙金兰 宫玉发 赵家鹏
组 员：于润发 王仲伦 王绍发 刘庆春 杨玉民
 苏 丹 何肃波 李宏栋 张荫生 费爱伦
 梁 义 葛玉刚 褚家蒙 翟汝直
秘书长：邓又强

职业学校电子类教材编审委员会

主任委员： 杨玉民 （以下以姓氏笔划为序）
副主任委员：于润发 （兼实用电子技术编审组组长）
 张荫生 （兼计算机技术编审组组长）
委 员：实用电子技术编审组
 来岳舟 陈其纯 张晓明 （以上为副组长）
 万相众 王条鑫 白春章 朱晓斌 沈大林
 杨荫彪 袁是人 徐洪吉 崔玉春
 计算机技术编审组
 王道生 王 森 栾宏为 （以上为副组长）
 马忠裔 刘永振 吕旭东 朱晋蜀 严振国
 陈继国 李海田 郑子罕
秘 书：王昌喜 吴浩源

1988~1990 年计划出版的职业学校教材

实用电子技术专业

1. 电子技术工艺基础
2. 模拟电路
3. 微型计算机应用基础
4. 制图与钳工工艺基础
5. 收录机原理与维修
6. 黑白电视机原理与维修
7. 录像机原理与维修
8. 家用电器原理与维修
9. 单片微型计算机原理与应用
10. 电子测量仪器
11. 维修电工技术
12. 电机的结构与维修

计算机技术专业

1. 计算机电路基础
2. 微型计算机原理与实验
3. BASIC 语言程序设计
4. 微型计算机磁盘操作系统的使用
5. 数据库应用基础
6. 微型计算机汉字处理与录入
7. 微型计算机外设结构与维护——打印机
8. 微型计算机外设结构与维护——显示器与键盘
9. 微型计算机外设结构与维护——软磁盘驱动器
10. 微型计算机接口技术
11. pascal 语言程序设计

前 言

本教材系由机械电子工业部职业学校电子类教材编审委员会计算机技术编审组评审、推荐出版的,作为计算机技术专业微型计算机汉字处理与录入课程的教材。

该教材由河北省石家庄市第二职业中学李瑞龙担任主编,山东海洋大学赫羽副教授担任主审。

本课程的参考教学时数为 64 学时(理论 30 学时,实践 34 学时),主要介绍微型计算机汉字处理的基本原理,汉字录入技术以及汉字编辑制表软件的使用等。本书重点介绍了微型计算机的汉字录入技术,其中包括拼音码输入,“五笔字型”,区位码输入等。

本教材由李瑞龙编写第一章,宋宏斌编写第二章和第四章的第一节,邢程编写第三章和第四章的第二节。王森副教授为本书的编写提出过许多宝贵的建议和意见,这里表示诚挚的感谢。由于编者水平有限,书中难免存在缺点和错误,殷切希望广大师生批评指正。

编者

1989 年 6 月

目 录

第一章 概述.....	(1)
第一节 什么是汉字处理	(2)
第二节 汉字处理的主要内容	(3)
第三节 国内外汉字处理技术的发展概况	(5)
第二章 汉字处理的基本原理.....	(7)
第一节 汉字处理中使用的代码	(7)
第二节 字库.....	(13)
第三节 汉字处理过程.....	(15)
第三章 汉字录入技术	(17)
第一节 APPLE- II 机软汉字输入	(17)
第二节 拼音码输入.....	(25)
第三节 五笔字型输入.....	(37)
第四节 区位码输入.....	(63)
第五节 其它录入方式简介.....	(64)
第四章 汉字编辑制表软件的使用	(67)
第一节 汉字编辑软件 WORDSTAR	(67)
第二节 XE 多窗口字处理软件简介	(86)
附录 汉字区位码表	

第一章 概述

第一节 什么是汉字处理

随着大规模集成电路和计算机技术的迅猛发展,人类开始进入信息时代。人类社会进入信息化时代的一个重要特征是采集信息、贮存信息、传播和利用信息。信息的采集、贮存、传播和利用是离不开文字的,大量复杂的文字信息的处理又离不开电子计算机。电子计算机已经不单是数值计算的工具,而且是一种信息处理的电脑。

汉字是世界上最悠久的文字之一,它记载着我国五千年光辉灿烂的文明史。汉字是世界上公认的重要的文化交流工具之一,地球上五分之二的人们用它来交流思想与文化,传授知识,传播信息。

在我们进入信息时代的时候,人们已经深刻地认识到信息资源的重要性。在一个现代化的社会里,有大量的信息需要交流、传递、存储和处理。人们如果能充分地开发和利用信息资源,那么信息可以转化为巨大的效益。因此,信息处理科学的发展水平反映着一个国家科学技术的发展水平。

人脑在信息的识别、分析、综合、推理和联想等智能方面有着较强的功能,但记忆信息的功能较弱,而且处理信息的速度也较慢。人们要从大量而又经常变化的信息中识别、分析、

综合关键性信息并使其变化为生产力,不是一件简单的事。电子计算机的产生和发展,使人们有了信息处理的强有力的工具。它具有信息转换快、存储容量大、处理速度高、传递速度快、重复性强和不易疲劳等优点,这正好弥补了人类在处理信息方面的不足,从而给人类对信息资源的开发与利用开创了一个崭新的时代。

在我国,要进行处理的信息主要是汉字信息,电子计算机的汉字信息处理对我国的现代化建设非常重要,对世界的经济发展也有一定的影响。近几年来,国内外掀起了研究汉字信息处理技术的热潮。我国是汉字的发源地,对汉字的结构、特性及应用更为熟悉,对发展汉字信息处理技术的要求更为迫切,受益也最大。因此,我们应该积极开展计算机汉字处理的研究,走在世界的前列。

电子计算机诞生在西方,其硬、软件系统适应于西方的拼音文字。西方的拼音文字,字符数量少(如英文字母 26 个),结构较汉字简单,所谓“见字即能念”,“见字知其意”,“会说即能写”。国外拼音文字国家早已实现了计算机文字处理,并且已经收到了效益。而当今的汉字,是由象形文字演变而来的,基本上属于一种图形文字,所谓“重形轻音”,“由形达意”。它字数多,字体多,一字表一意,一字表数意,在印刷系统中字号也多,这就给汉字的计算机处理带来了困难。

实现电子计算机的汉字处理的方法很多,现在被普遍采用的是汉英文兼容技术,改造计算机的结构体系、输入输出设备、系统软件和应用软件等。近几年,我国的科学工作者和计算机应用人员在汉字处理方面进行了艰苦的工作,为我国的计算机应用事业开创了新局面;在我国,计算机汉字处理技术

已经显示出它的强大生命力。

第二节 汉字处理的主要内容

一、汉字的特点

要想了解微型计算机的汉字处理,我们首先要了解汉字的基本特点。

汉字是从象形文字演变而来的,它起源于图画,但又区别于图画,方块汉字是一种特殊的图形。

汉字是一种表意文字。一个汉字通常表示汉语里的一个词或词素,只读一个音节,所以汉字被看作是形、音、义的统一体。一个词是可以由不同的字来表示的,一个字又存在着大量的同音字。这样,就使汉字形体复杂,数目繁多。这些均说明,微型计算机汉字处理较拼音文字确有一定的难度。尽管如此,微型计算机信息处理的汉字化还是可行的,这是因为汉字较拼音文字至少有以下两个优点:

汉字码较拼音字码短。汉字码平均不大于4个字母,而拼音字码较长,如英文,其平均字长6.7个字母;

汉语的造句比英语紧凑、精炼、含义深广。

二、汉字的输入和存储

汉字的输入是微型计算机汉字信息处理系统中的关键之一,也是难点之一。

汉字的输入通常可分为自然输入和编码输入两大类。

自然输入指的是汉字的文字识别和语音识别。文字识别

有印刷体识别和手写体识别；语音识别有孤立词语的语音识别，断续词语识别和连续词语识别。自然输入是汉字输入的最终和最理想的形式，这种输入方式还在研究和试用阶段。

编码输入是目前普遍采用的汉字输入方法。它是将汉字编成代码，将汉字符号化，以便通过适当的方法将汉字输入微机。目前，我国已有数百种汉字输入编码方案。这些编码方案有的是以字形属性编码，有的是以字音属性编码，也有的是以形、音、义三个属性编码。本书介绍的APPLE-Ⅱ机软汉字输入，拼音码输入，五笔字型，区位码输入都是常用的汉字输入方法。

为了要在计算机间进行汉字数据通信，必须制定统一的汉字代码标准。国家标准信息交换用的汉字编码字符集称为国标码。国标码属编码输入类。

此外，汉字输入还有整字输入。这种输入方法既可以列入自然输入类，也可以列入编码输入类，如汉字大键盘输入。日本普遍采用大键盘输入。

汉字的存储有两种形式：一种是汉字的代码信息存储，一种是汉字字模点阵信息存储。前者与汉字数据库设计有关，后者与汉字字形库设计有关。

汉字数据库中的代码存储采用不等长码，以节省数据库存储空间。

汉字字形库通常有低、中、高分辨率三大类。 32×32 点阵以下为低分辨率； 32×32 点阵以上， 64×64 点阵以下为中分辨率； 96×96 点阵以上为高分辨率。对于由点阵构成的字形，我们根据汉字出现的不同频度将汉字字形分成若干级。因为汉字字型库所占据的存储空间非常大，所以对字形要采取压

缩还原技术。

三、汉字的输出

汉字的输出主要靠汉字的显示设备和印字设备。目前,国产微型计算机在设计中都注意到了输出设备适应于汉字处理:显示设备具有高分辨率并带有智能的汉字显示终端,它既能显示西文,又能显示汉字,输出方式有单显和彩显;印字设备一般是点阵针式打印机。我国的微型计算机的显示设备和印字设备都有较强的图形表格处理功能和文字编辑功能。

第三节 国内外汉字处理技术的发展概况

日本在微机汉字处理的研究与应用方面起步较早,60年代就已经研制出高速汉字印字机和计算机编辑系统。70年代中期,由于微型机及软盘技术的迅速发展,日本的汉字信息处理技术进入普及应用阶段。70年代后期,日本制定了《信息交换用汉字字符集用的控制文字符集》,统一了汉字处理的标准。目前,日本在微机汉字处理的输入技术上,主要研究汉字识别;在输出设备方面,已开始使用激光印字机。

在美国,王安公司研制了整套适用于办公事务处理的汉字处理系统,并能提供多种应用软件;IBM公司已经研制出汉字智能终端及汉字联机系统。

我国在计算机汉字处理方面起步也不晚,60年代初,在国产104计算机上进行俄语翻译过程中,提出汉字处理方案。70年代初,我国科学工作者为了解决汉字的计算机输入问题,提出了多种汉字编码方案。1976年,我国提出国内使用的

汉字频度表。1978年,我国召开了第一次汉字编码学术会议。1980年,中国仪器仪表学会成立了汉字信息处理系统研究会,此后,汉字处理技术从单纯编码研究转向汉字处理系统的研制。1981年,在天津成立了中国中文信息研究会,1983年,在桂林又成立了中国微电脑应用协会汉字信息处理专业委员会。这些学术团体的建立使我国计算机汉字处理系统的研制和应用进入了蓬勃发展的时期:优选编码方案,制定点阵字库的国家标准,研制先进的汉字处理硬件设备,开发汉字系统软件和应用软件,进行文字识别和语音识别系统的研究。总之,我国的计算机汉字处理技术已经达到国际领先的地位。

近几年,国际间经常进行汉字处理学术和技术研讨合作会议。1983年,在北京和东京分别召开了“中文信息处理国际研讨会”和“大字符集语言信息处理国际会议”。1986年的新加坡国际会议和1987年北京“第二次中文信息处理国际研讨会”,将国际间汉字处理工作推向了一个新阶段。这些国际会议,在计算机汉字输入输出设备 and 人机接口、操作系统、应用软件、汉字处理技术的理论研究方面都取得了丰硕的成果。

第二章 汉字处理的基本原理

电子计算机自 1946 年问世到今天,经过了半个多世纪。在这段时间里,它经历了几代的发展,技术日渐成熟,在促进人类适应自然、改造自然中发挥了举足轻重的作用。它不仅应用于科学界,而且已经渗透到各行各业。

电子计算机首先由西方人研制成功,其应用适合于西方拼音文字。如何用先进的微型计算机技术来处理汉字呢?这是我们十分关心的问题。怎样才能屏幕上显示出优美的汉字,让打印机打出漂亮的汉字呢?

其实,汉字的微机处理并不神秘,下面我们来学习有关这方面的知识。

目前,国内、国外的汉字处理方法很多,不可能一一列举,我们仅用长城系列机(IBM 系列微机)使用的 CC-DOS 作概略说明。

第一节 汉字处理中使用的代码

在汉字处理过程中,我们使用了许多代码。这些代码有些是操作员使用的,有些是计算机内部处理汉字使用的;使用的代码有汉字输入码、汉字内部码、汉字地址码、汉字字形码、汉字交换码。

下面我们来逐项学习这些代码的知识。

一、汉字输入码

微机键盘通常为标准的打字机键盘,它处理西方文字是一键一字,键与字符有直接的对应关系。由于汉字的字符多,形体复杂,不可能利用现有的微机小键盘实现一键一字。如果利用标准打字机键盘上键的不同组合为每一个汉字编制一个由几键组成的代码,我们就可以将汉字输入微机了。

用户在微机键盘上输入汉字时给汉字的编码称为汉字输入码。对同一个汉字可以根据不同的原理设计出不同的汉字输入编码方案来。每一种汉字输入编码方案自成一个体系,经常见到的汉字输入有拼音输入、区位输入、快速输入、首尾输入、见字识码、五笔字型等。

对同一个汉字,由于编码方案不同,它的汉字输入码也不同。例如,对于汉字的“啊”,拼音码是小写的英文字母“a”;区位码是数字 1601;见字识码是大写的英文字母:“KEKG”,五笔字型是大写的英文字母“KRAK”。可见,对不同的编码方案,其输入码的差别是很大的。虽然这些编码设计原理和方法千差万别,但在设计上都考虑到中国汉字的字数多、笔划多、字音多等特点。

这些编码方案由于设计上的原因,使用时有的输入速度快,有的输入速度慢;有的容易学,记忆规律简单,有的记忆规律复杂,很难学;有的汉字输入码长,要多按几个键,有的汉字输入码短,可以少按一些键。这些汉字输入码直接与用户接触,处于人机联系的界面上。

在输入汉字时,看不到计算机内部是如何对各种汉字输入码进行复杂处理的,所以我们对汉字输入编码的优劣性最

为关心。一般地讲,人们希望汉字输入码具有如下特点:

- 1、输入速度快,
- 2、容易掌握,
- 3、可靠性高,
- 4、重码率低。

输入速度快是指单位时间内输入的字数多。输入速度快主要涉及到编码的长度及组字规则问题。对一个汉字来说,它的编码越短,需要按动的键就越少;组字规则越简单,大脑思考时间就越少,反应速度就越快。英文字母输入的平均速度为8~12个/秒,最高可达14个/秒。由于汉字和西文的输入方式不同,所以输入速度比西文要慢,目前最快的输入速度已达到3个/秒左右。

容易掌握是指记忆规律简单,好学好记。目前已有几百种编码方案,常用的大约有几十种。它们的规则各有特色,有的十分简单,特别容易记忆;有的比较复杂,需要进行学习才能掌握,而且长时间不用还容易忘记。规则简单的汉字输入码如拼音码,它利用汉语拼音来输入汉字,只要我们认识汉字并且会使用汉语拼音就会用微机输入汉字;见字识码也是较简单的一种输入码。五笔字型是规则较为复杂的一种输入码,它需要特殊的训练才能较好地掌握这种输入方法。

可靠性高是指误码率低。误码是指键入的汉字输入码对应了不该对应的汉字,也就是说,根据某输入码敲了相应的键后,屏幕上出现的不是相应的汉字,而是其他的什么东西。我们不希望出现误码,否则它就没有实用价值了。

重码率低是指具有相同编码的不同汉字的个数越少越好,如拼音输入,它的重码率就相当高。拼音码是按发音来输