高等专科学校教材

汉字信息系统处理概论

夏 耘

西安电子科技大学出版社

高等专科学校教材

汉字信息系统处理概论

夏 耘

西安电子科技大学出版社

1989

内 容 简 介

本书全面系统地论述了计算机处理汉字信息的基本原理和方法．全书共分六章，先后讨论了汉字属性、汉字的内部码、汉字的输入和输出方法、IBM-PC 及其兼容机的汉字处理系统．

这是一本适用于我国大专院校有关专业的教科书，也可作为研究汉字信息处理技术的入门指导书．

高等专科学校教材
汉字信息系统处理概论
夏 耘
责任编辑 徐德源

西安电子科技大学出版社出版
空军工程学院印刷厂印刷
陕西省新华书店发行 各地新华书店经售
开本 787×1092 1/16 印张 7.4/16 字数 163 千字
1989 年 12 月第 1 版 1989 年 12 月第 1 次印刷 印数 1—3000

ISBN 7-5606-0099-9/TP·0036 定价：1.55 元

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作，由于各有关院校及参与编审工作的广大教师共同努力，有关出版社的紧密配合，从1978年至1985年，已编审、出版了两轮教材，正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应"三个面向"的需要，贯彻"努力提高教材质量，逐步实现教材多样化，增加不同品种、不同层次，不同学术观点、不同风格、不同改革试验的教材"的精神，我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会，在总结前两轮教材工作的基础上，结合教育形势的发展和教学改革的需要，制订了1986～1990年的"七五"(第三轮)教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿，是从通过教学实践、师生反映较好的讲义中经院校推荐，由编审委员会(小组)评选择优产生出来的。广大编审者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量，作出了不懈的努力。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足之处。希望使用教材的单位，广大教师和同学积极提出批评建议，共同为不断提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前　言

本教材系按电子工业部的工科电子类专业教材 1986–1990 年编审出版规划，由大专计算机教材编审委员会软件编审小组征稿，推荐出版。责任编委朱乃立。

本教材由上海机械专科学校夏耘主编，国营北京有线电厂韩鸿文担任主审。

本课程是计算机软件专业的选修课，参考学时为 40 学时。要求学生掌握计算机处理汉字信息的基本原理和方法，理解汉字内部码的含义、作用及它与其他汉字编码的关系，为学生参与设计或分析汉字信息处理系统打下基础。讲授本课程时，应注意着重讲清汉字内部码这一章，该章是全书的重点、难点。正确理解汉字内部码，便能深入研究汉字信息处理系统。

学习本课程之前，要求先学计算机原理、操作系统、外部设备、程序设计语言以及数据库等课程。

本教材的编写工作全部由夏耘担任。初稿先后由甘圣予、毛德行、殷志鹤、韩鸿文四位同志评阅，他们都为本书提出了许多宝贵意见，这里表示诚挚的感谢。本教材有些章节参考了由赵珀璋与徐力两位老师合著的《计算机中文信息处理》一书，在此表示感谢，并在此感谢关心此书出版的阎天民、李浩、陈家正、薛桂芳等老师。由于编者水平有限，书中难免还存在一些缺点和错误，殷切希望使用此教材的老师和广大读者批评指正。

编　者

目　录

第一章 概 论

§1.1 汉字信息处理的意义和任务

§1.1.1 汉字信息处理的意义

信息是人们用以对客观世界直接进行描述的、可以在人们之间进行传递的一些知识。物质的存在伴随着信息的存在，物质的变化会引起信息的变化。它是构成客观世界的三大要素之一。

在本世纪60年代，电子计算机在非数值计算领域内得到推广和应用，它能加工和处理信息，因此相应地产生了信息处理这一新的概念。信息处理包括对信息的收集、记载、分类、排序、存贮、计算或加工、传输、制表、递交等等工作，它使有效的信息得到合理和充分的使用，并反过来促进社会生产力的发展，同时又产生出新的信息。

信息处理技术中，对文字信息的处理称为文字信息处理。文字信息处理中对汉字信息的处理称为汉字信息处理。随着计算机系统功能的不断提高，应用领域的迅速扩展，信息处理的概念、涵义、作用和涉及的范围也大大扩展了。例如：情报资料和图书的自动编目和检索；书刊和报纸的自动编辑和排版；事务处理；企业管理；办公室自动化；数据通讯等。实际上，文字信息处理技术已逐渐渗透到人类思维、生产和生活等一切方面。

事实上，计算机系统只能处理数据，而数据所表示的意义就是信息。因此对信息的处理体现为对数据的处理。表示文字信息或符号信息的数码称为代码。例如，在对西文字母以及符号的处理中，对应于26个字母(分为大写和小写体)和一些常用符号，按某种规律和约定，编成一组数码，这组数码称为字符代码。因此对文字信息的加工，就是对代码数据的加工。故对汉字信息处理的过程可分为以下3个阶段：

(1) 信息的输入：通过输入设备把文字信息转换成代码并送入计算机。

(2) 信息的加工或处理：根据各类不同的应用，借助预先设计好的程序对输入的信息进行加工和处理，从而得出结果信息。

(3) 信息的输出：通过输出设备把数据代码形式表示的结果信息，复原成文字。

汉字是一种表意文字，字量多，字形复杂，这便给汉字信息处理带来了不少困难。因此，在构成汉字信息处理系统时，需在计算机软、硬件方面做大量的工作。

汉字信息处理是中文信息处理的主要方面，故对汉字信息处理的研究是十分重要的。中文信息处理的另一方面是对少数民族文字的处理，由于我国是一个多民族国家，故这方面的工作也不能忽视，但本书只对汉字信息处理作必要的介绍。

§1.1.2 汉字信息处理技术涉及的范围

一、汉字属性

汉字信息处理技术是一项综合性的技术，其核心是计算机技术。为了合理地制定一些计算机处理汉字的技术规则，先要研究汉字的基本特性(又称汉字属性)。它主要包括：汉字字量，字形分解，汉字字体，使用频度，汉字发音，汉字索引，汉字排序，汉字标准交

换码等。在实际应用中，可根据需要进行增删。只有对汉字属性进行较彻底的研究，牢固掌握汉字的基本特性和应用规律，才有可能合理地设计出各种类型的汉字信息处理系统。

二、对汉字词组及文句结构的研究

除了汉字属性外，为了更有效地研究汉字信息，需对组成的字或字组(称为词)进行研究。所谓词是指经常使用并有特定含义的单个或多个汉字的组合。词的属性包括词的种类、组词字数、词的使用频度、词的含义、排序特性等。在汉字输入方案中，对于使用频度特别高的词，可用软件方法由用户自己定义之。若从意义和信息处理功能上分析，汉字信息处理既包含对汉字本身的处理，例如：汉字在系统中的输入、输出，以及汉字的编辑；也包含对汉字文件中的句法和上、下文结构的处理。显然，后者比前者在含义上深刻些，所涉及的范围也要宽广些。

三、汉字信息处理和计算机技术

信息处理离不开计算机技术。用作汉字信息处理的计算机系统，一般需根据系统要求扩充内、外存，添加部分汉字输入、输出设备，配备面向汉字信息处理系统的作业任务的应用程序。实际上，西文信息处理系统所用到的一些技术，在原理上都适用于汉字信息处理系统，故汉字信息处理系统应充分利用西文信息处理系统的软件资源。

除上述系统方面的工作外，如何进行汉字输入是汉字信息处理技术中的另一重要课题。目前，主要采用编码输入方法，完全依靠手工操作，故其效率低，速度远不能和计算机运行速度相适应。利用计算机直接识别汉字字模或语音的方法尚处于研究阶段，离实用尚有相当的距离。

汉字信息处理系统除了输出打印或显示的结果为汉字外，还可以语音方式输出，但这种输出方式也处于研究阶段，尚未实用。

综上所述，汉字信息处理技术所涉及的范围是很广的，必须分别解决各方面的课题，才能使汉字信息处理技术水平不断提高。

§ 1.1.3 汉字信息处理技术待解决的问题

由于汉字的特点是字量大，字形复杂，因此要建立一个汉字系统，就需要解决汉字的输入、存贮和输出等问题。此外，要把西文信息处理技术中的成熟软件用于汉字信息处理系统，使计算机能兼容西文和汉字两种文字的信息处理功能，则还有很多工作要做。

汉字信息处理技术中待解决的主要问题如下：

(1) 汉字输入技术和设备；
(2) 汉字字模的存贮；
(3) 汉字信息处理系统的内部码问题；
(4) 汉字输出技术和设备；
(5) 汉字系统的软件问题；
(6) 汉字终端技术；
(7) 汉卡技术。

上述问题目前已有专人研究，但尚未很好地解决，如解决了，汉字信息处理技术将达

到新的高度。

§1.2 汉字信息处理系统构成和分类

§1.2.1 汉字信息处理系统的构成

从信息处理角度来看，中文和西文没有什么本质区别，原则上现有的西文信息处理系统都可用来处理中文信息。但实际上，由于中文的汉字数量多，字形构造复杂，因此中文处理又有其固有的特殊性。

目前，在建立什么样的中文信息处理系统方面有两大流派。一派主张走"民族化"道路，即充分考虑中文信息的固有特性，从系统设计角度来考虑体系结构、系统软件、高级语言和应用软件，要求全部中文化，走设计中文计算机、中文操作系统、中文高级语言的道路。这样做，处理中文信息的效率一定会比用现有计算机高许多，但却失去了和国际兼容的特性，别人的成果我们不能利用，我们的成果也不易被别人共享。另一派则主张走国际化道路，即在现有西文计算机基础上，考虑中文的特点，通过对系统软硬件作适当改造，做到中西文兼容，能共享国际信息处理成果。这样做，单从处理中文信息的角度来看，可能没有前者效率高，但可以充分利用现有西文计算机的全部软、硬件资源。

就构成而论，汉字信息处理系统和通常的计算机系统是相似的，都包括硬件和软件两大部分。下面就这两大部分作一简要的说明。

一、硬件组成

汉字信息处理系统的硬件包括主处理机、常规外部设备和汉字外部设备。主处理机是通用电子计算机。根据所要求的处理能力和工作方式，它可以是大、中、小型计算机，也可以是微型机。常规外部设备主要包括外存贮器，汉字外部设备包括汉字输入键盘、打印机、显示设备等。

在汉字信息处理系统中，汉字字模库和汉字显示终端是两个重要的组成部分，对它们的设置和连接决定了该系统的汉字部分的工作方式。

通常，汉字字模库可设置在系统的3种不同的部位：

(1) 汉字字模库作为一个独立的外部设备。这种方法的优点是汉字字模库可为多台设备所共享；缺点是占用总线传送时间，影响系统效率。

(2) 汉字字模库直接和汉字打印机连接。如果系统中只配一台汉字打印机，而这台打印机的利用率又很高，就应该直接把汉字字模库和打印机相连，并可构成智能打印机。

(3) 汉字字模库设置在汉字显示终端上。

汉字终端和主处理机系统之间一般有并行和串行两种连接方式：

(1) 并行连接。这种连接方式又称外部设备型连接。如果汉字终端采用标准的8位微处理机，除了地址线和控制线外，另用8条数据线和主处理机系统相连，每次交换一字节信息。对于设在主处理机房内或距主处理机房近的汉字终端，通常采用这种连接方法。

(2) 串行连接。这种连接方式又称为通讯连接方式。这里由于信息是逐位传输的，所需传输线的数目少，故适合于长距离或远程通讯，其接口一般采用RS232-C标准或RS422标准。

二、汉字信息处理系统的软件

和通常的计算机系统相似，汉字信息处理系统的软件包括系统软件和应用软件两大类。

(一) 系统软件

汉字信息处理系统的系统软件包括以下项目：

(1) 兼容汉字和西文信息处理的操作系统。

(2) 汉字输入输出管理程序。

(3) 汉字文本编辑程序。

(4) 高级程序设计语言。

(5) 数据库管理系统。

(二) 应用软件

在汉字信息处理系统中，对各种应用项目都要配置相应的应用程序。对于一些典型的应用程序，应提供商品化的应用程序包。要尽量利用西文系统已有的一些应用程序包，使其在经过必要的改动后，成为汉字系统的应用程序包。

有些汉字信息处理系统的应用项目需配置相应的文件系统，这就是汉字文件系统。

§1.2.2 汉字信息处理系统的分类

从对系统功能和输出文字质量的要求上来区分，可以把汉字信息处理系统分成两种类型：即精密型汉字编辑排版系统(Chinese Editing and Typesetting System)和通用型汉字系统。前一种类型用于正式出版的书、刊、报纸的编辑出版；后一种类型用于汉字文件处理，统计报表，数值和数据处理等。后一种类型系统的使用范围是很广的，汉字信息处理技术的推广应用很大程度上取决于这类汉字系统的发展。

一、精密汉字编辑排版系统

这一系统最重要的技术关键是高精度汉字字模的存贮和版面输出。该系统的特点是对汉字字模点阵密度要求很高。如果要求分辨率为 30 线 / mm 以上，那么，对于一个 5 号字(尺寸为 3.675 mm^2)，就要求其点阵密度达 96×96 点。对于精密汉字字模，不仅每个字的点阵信息量大，而且由于字量多，需多种字体和字号，从而使总的字模信息量非常庞大。此外，还要兼顾有适当的字模输出速度，因此需解决一个高倍率压缩信息的课题。这类系统的版面输出可以采用两类技术，即电子束扫描输出和激光扫描输出。在扫描的同时，输出版面在感光底片上记录成像。这种输出设备又称为照排设备。整个系统除了包括上述设备外，还包括：排版用计算机；相应的外设；编辑、改错用的联机汉字显示终端；汉字数据采集用的汉字终端；校样印刷机；字模自动制作设备；图片输入设备等等。

汉字编辑排版系统要配备大量专用的软件，它们包括：专用的操作系统；编辑排版专用语言及编译系统；汉字文件系统；书、刊、报纸等各种版式的排版应用程序；图片处理软件等。

二、通用型汉字信息处理系统

它的特点是：主要用来实现数据处理或一般的汉字信息处理，使用面广；系统成本低；不太讲究汉字字模的质量。通用型汉字系统的字模点阵规格目前流行的主要有两种：15×16 点阵；24×24 点阵。对低于 15×16 点阵的字模，因质量太差，目前已很少采用。这类通用型系统的汉字印刷机，目前以针式打印机为主，今后也可能采用简易型激光扫描印刷机。这类系统对于字号尺寸的变化，不像对精密型字模那样要求严格，通常采用软件变倍或调节针头间距的方法加以解决。采用这些方法，只能变出很少几种尺寸，而且不一定符合出版印刷业定出的字号尺寸规范。

以下列举几种通用型汉字信息处理系统的例子。

(一) 汉字情报检索系统

用于书、刊、情报资料的存贮和检索等自动管理系统，进一步可发展为汉字数据库检索系统。这类系统的特点是需配备容量很大的外存贮器，以收容尽可能多的情报资料。一种类型是配置多台问答式联机汉字显示终端，供用户以询问方式向系统索取情报资料。这样的系统称为联机汉字情报检索系统。另一种类型是采用集中提问的输入方式的系统。称这种系统为批处理汉字情报检索系统。系统的响应或回答既可以由荧光屏显示，也可以由汉字印刷机印出检索结果。

(二) 企业管理系统

它用于大型工矿企业的生产管理、计划调度、行政管理、人事工资管理、设计图纸和工艺资料管理、产品和合同管理、以及供销计划管理等。工矿企业应用了计算机管理后，不仅可节省人力，而且能大大提高管理效能，便于实现现代化的企业管理体制。

(三) 事务处理系统

它用于计划拟定、公文管理、档案管理、统计报表制作等方面，主要面向各种不同的事务或业务管理。

(四) 办公用计算机

它不仅用来实现办公室范围内的文件和书信的印制，还可用作简易的文件档案管理。

(五) 汉字通讯系统

汉字信息也可以实现有线或无线传送。不过，在线路上传送的是汉字代码。在接收端，再把汉字代码转换成字形输出。数据系统的信息交换中心是一个计算机系统，它用来控制数据流的传送。

(六) 窗口系统

它可供许多公用事业机关或各种业务服务行业进行各类日常的业务处理。

(七) 文字自动翻译系统

它把不同种类的外文译成中文。

(八) 其他智能系统

利用计算机的逻辑判断和数据处理能力，还可构成除上述 7 类系统以外的其他应用系统。

§1.2.3 汉字信息处理技术标准化问题

一、标准化工作的重要性

汉字信息处理在目前还是一门新技术，它虽和传统计算机技术的关系极为密切，但却有很多独特的技术课题有待进一步去探索。而一项新技术在推广应用前，必须先确立技术标准，这是工业技术发展必须遵循的途径。如果没有或缺少技术规范或标准，各行其是，必然会影响这项技术的进一步发展。但技术标准也不可能是凭空产生的，它必须建立在一定的研制成果的生产和使用经验的基础上，并从基础的标准开始，逐步形成一个比较完整的技术标准体系。

二、标准化工作内容

汉字信息处理技术标准化工作的内容是较多的，可列举如下：

1. 有关汉字属性的标准

(1) 汉字数字化字模标准。该标准主要是对通用型的汉字字模而制定的。应先确定字形和字体。对于点阵密度可选择两种规格，即 15×16 点阵，24×24 点阵。根据需要还可选择 32×32 点阵的标准。要求在这样的点阵密度条件下，得到质量较高的字模图形。此外，制定汉字标准字根，对于以字形为主的汉字编码输入方案的设计和字根汉字字模库的设计是很有意义的。

(2) 汉字交换码标准。制定本标准是为了方便系统之间的汉字通讯。

(3) 汉字索引(Indexing)和排序标准。汉字信息处理系统中要建立汉字属性字典，以便在系统中查找汉字或由它组成的词。建立汉字索引标准能解决汉字检索中的困难。

2. 汉字系统所用的控制功能标准　控制功能标准用来规定某种控制相应的计算机系统中各项设备的特定动作，其中包括控制功能的种类、符号和含义。因为它是直接关系到各类外部设备的控制动作和软件，所以制定控制功能标准对于汉字系统的配置和应用是很重要的。

3. 汉字编码和输入方法标准　编码和输入方法有密切关系。目前各种类别的汉字编码方案很多，其输入方法也各不相同。制定汉字编码方案的标准需在优选汉字编码方案的前提下进行。为此，需要首先拟定汉字编码评测标准。制定汉字输入方法标准时，要注意分开层次等级，提高型和普及型并重的方针，以利于切合实际，推广应用。在制定输入方法标准的同时，也应包括制定笔触式汉字字盘盘外字的输入方法的标准。

4. 汉字设备方面的标准

(1) 汉字键盘标准。除了字母数字键盘外，字根式汉字键盘、笔触式汉字键盘都需有技术标准。

(2) 汉字字模库标准。对于成批生产的 ROM 汉字字模库，应有关于 ROM 器件的集成度、收容字数、编址方法、接口技术等技术标准。对于字根式汉字字模库，也要有相应的技术标准，以利于建立汉字终端设备的设计和生产的标准体制。

(3) 汉字打印机标准。对于汉字打印机，需对常用机型制定技术标准和建立型谱系列标准。

(4) 汉字显示终端标准。对于汉字显示器，应定出显示管尺寸系列、分辨率等级、所

显示的字模点阵和满屏的字数等标准。对显示终端，需要对汉字字模库的设置、扫描刷新方式、接口技术等定出标准。同时还应制定汉字显示终端的功能等级和型谱系列标准，并作出显示终端模块化结构的规定。

(5) 其他硬设备方面的技术标准。对于汉字光学字模识别技术和设备，联机手写汉字识别技术和设备，汉字语音输出设备等，都需根据研制工作的进程，制定相应的技术标准。

5. 汉字软件技术标准　汉字信息处理系统软件的基本功能和操作系统扩充汉字的功能的技术需要标准化，各种程序语言处理汉字信息的技术也要标准化。而实现上述标准化的基础乃是汉字内部码和汉字数据类型的标准化。因为汉字内部码是汉字信息系统的核心，内部码若不统一，直接影响了软硬件的兼容，严重破坏了现有计算机与信息处理标准化的成果，因此制定全国统一的汉字内部码规范已成为当前急迫的任务。

§1.3　汉字信息处理技术的现状和展望

§1.3.1　国内汉字信息处理技术的现状

"汉字信息处理"一词，是近十多年来才流行起来的，但是其活动渊源已有很久了。从广义上来说，从我们祖先创立汉字开始，就一直在进行汉字处理活动；从狭义上来说，第一部汉字字典产生以来，就一直在进行汉字信息的分析和综合处理。这里的汉字信息处理是指利用计算机对中文信息进行处理。

早在本世纪 50 年代，我国研制 103、104 电子计算机时，就在机上开始了俄汉机器翻译工作，这是我国第一次把中文信息处理和电子计算机结合起来。60 年代这方面工作不断开展，但限于当时计算机运算速度低和存贮容量小，一直进展不大。

70 年代以来，由于大规模集成电路和计算机操作系统的迅速发展，计算机的软硬件系统为中文信息处理提供了物质基础，国内汉字信息处理活动进入了新的高潮，并达到新的高度。

国内活动方面，1978 年 12 月在青岛召开了全国编码工作交流会。来自 17 个省、市、自治区的 80 多名代表参加会议，交流了 40 多种编码方案。这是我国关于中文信息处理的第一次全国性会议，会议对我国中文信息处理活动起了巨大的推动作用。

1980 年开始筹建中文信息研究会，并于 1981 年 6 月在天津正式建立。1982 年先后成立了中文信息研究会国际联络委员会、学术委员会和组织委员会。学术委员会在 1982 年先后成立了 5 个专业委员会：

(1) 基础理论专业委员会；

(2) 汉字信息处理系统专业委员会；

(3) 汉字编码专业委员会；

(4) 汉字信息处理专用设备专业委员会；

(5) 自然语言处理专业委员会。

1983 年 5 月在武汉召开了中国中文信息研究会第二次全国学术会议，来自 24 个省、市、自治区的代表，共发表了 133 篇学术论文。1984 年 10 月 7 日至 10 日在内蒙古自治区呼和浩特市召开了"全国首次少数民族语言文字信息计算机处理学术讨论会"，来自 13

个省、市、自治区的蒙古族、维吾尔族、朝鲜族、哈萨克族等11个民族的91位从事民族语言、文字及计算机处理工作的科技工作者出席了大会。会议交流了23篇学术论文，主要内容有下面几个方面：

(1) 少数民族语言文字信息处理输入编码设计及标准研究。

(2) 少数民族语言文字信息处理系统设计及字库与文献库的建立。

(3) 少数民族语言文字多文种联合检索。

(4) 少数民族语言理解的理论探讨与机器翻译等问题。

1985年10月16日至18日在内蒙古自治区首府呼和浩特市召开了中国中文信息研究会少数民族语言文字信息处理专业委员会成立大会暨第二届学术讨论会。

1986年中国中文信息研究会各专业委员会先后召开年会，主要内容为：

(1) 中文信息处理系统的研究和设计。

(2) 中文终端网络。

(3) 中文数据库。

(4) 中文办公室自动化系统。

(5) 中文键盘输入新技术。

(6) 汉字编码优化及评测标准研究。

(7) 中文信息处理设备新技术和新工艺。

(8) 中文文字识别及语音识别。

(9) 中文信息处理基础理论。

(10) 自然语言处理。

(11) 少数民族语言文字系统开发、编码字符集、字模点阵集及数据集、输入键盘布局等标准化。

1986年10月在北京召开了中国中文信息研究会全国第二次会员代表大会，选举了新的理事会，常务理事会。大会还提议组建教育委员会，加强计算机中文信息处理普及和提高教育的活动，争取早日在全国高等院校设置课程，进面设置专业。

1987年10月23日至25日在江苏省连云港市召开了中文信息技术专业委员会成立暨学术讨论会。下属学组为：技术标准学组；系统及应用学组；输入输出学组；计算机语言学组及计算机翻译学组；少数民族语言文字处理学组。

中文信息技术专业委员会拟于1988年9月至10月在杭州举行本专业委员会第一次年会。

1988年6月至7月间在四川成都举办全国中文信息处理新技术新产品交流展销会。

1988年7月至8月在延边自治州举行少数民族语言文字学处理技术讨论会。

近年来，我国中文信息处理取得了许多成就。国家颁布的有关中文信息处理标准有：

(1) GB1988-80"信息处理交换用的七位编码字符集"。

(2) GB2311-80"信息处理交换用七位编码字符集的扩充方法"。

(3) GB2312-80"信息交换用汉字编码字符集(基本集)"。

(4) GB3453-82"数据通讯基本型控制规程"。

(5) GB3454-82"数据终端设备(DTE)和数据电路终端设备(DCE)之间的接口电路定义表"。

(6) GB5199.1～5199.2-85"信息交换用汉字 15×16 点阵字模集及数据集".

(7) GB5007.1～5007.2-85"信息交换用汉字 24×24 点阵字模集及数据集".

(8) GB6345.1～6345.2-86"信息交换用汉字 32×32 点阵字模集及数据集".

(9) GB5261-86"文字和符号图形设备的增补控制功能".

(10) GB7589-87 及 7590-87"信息交换用汉字编码字符集第二辅助集和第四辅助集".

(11) 汉字属性，汉语词汇，汉字中分辨率点阵字形等国家标准正在制定.

(12) GB5119 标准汉字排列格式或增加内容.

§ 1.3.2 国际汉字信息处理的现状

1980 年 10 月在香港召开了"中文资料与文稿处理的国际电算机学术会议"，这次会议对 70 年代中文信息研究工作进行了一次展示，会上除 3 篇综合报告外又分为 8 个部分作了分题报告和讨论:

(1) 中文终端和人机联系.

(2) 按字形的编码.

(3) 系统设计.

(4) 输出设备.

(5) 语言与系统.

(6) 应用系统.

(7) 言语识别系统.

(8) 字形识别与言语识别.

1982 年 9 月在美国华盛顿召开了"中文计算机学会国际会议"，会议对下列 9 个方面作了分题报告和讨论:

(1) 输出显示.

(2) 数据编码、信息处理和压缩.

(3) 普通话分析、识别和综合.

(4) 输入方法论和系统.

(5) 中文计算机处理.

(6) 自动排版.

(7) 软件开发和字处理.

(8) 印刷体分析和识别.

(9) 中文和日文手写体计算机识别.

1983 年 10 月 12 日至 14 日于北京召开了"中文信息处理国际研讨会"，由中国中文信息研究会和联合国教科文组织共同举办. 中国、美国、日本、澳大利亚、印度尼西亚等国家和香港地区的代表参加了会议. 会议共收到 70 余篇论文. 会议讨论了汉字信息处理系统、汉字计算机输入输出技术、汉字文字识别和汉语语音识别、汉字输入编码及有关基础理论、汉语语法语义理解等方面的学术问题. 还举行了中文信息处理系统设备展览.

1983 年 10 月 17 日至 19 日在日本东京举行了"1983 年大字符集语言信息处理国际会议"，会议是由美国中文计算机学会和日本信息处理学会联合召开的. 来自中国、美国、

日本、加拿大、新加坡、澳大利亚、丹麦等国家和南朝鲜、香港、台湾等地区的100名代表参加了会议，共收到了80余篇论文，会议论题涉及到输入技术、输出技术、文本和资料编辑、文字处理机、文字识别、语言处理、计算机通讯和语音处理技术等领域。所谓大字符集语言，主要指汉语、日本语和南朝鲜语，汉语有5万个以上汉字，日语有6000个以上汉字，而南朝鲜语有1800个汉字，与西文拉丁语系或斯拉夫语系仅有几十个字符相比，显然称得上是大字符集语言了。

1986年8月，在新加坡召开了一次中文信息处理国际会议。

中国中文信息研究会和中国国际科技会议中心于1987年8月4日至6日在北京召开"中文信息处理国际学术会议"，会议期间同时举办"多文种语言文字信息处理与办公室自动化展示会"，会议内容为：

1. 输入输出和人机界面

(1) 汉字编码及其评测标准研究。

(2) 汉字I/O设备和终端设备。

(3) 汉字识别、语言理解和语言合成。

2. 系统

(1) 中文信息处理系统。

(2) 中文信息处理软件。

3. 应用

(1) 汉字数据库。

(2) 汉字应用系统设计和实践。

(3) OA系统中的中文信息处理技术。

(4) 问题回答系统(OA系统)和机器翻译系统。

4. 基础研究

(1) 汉字信息处理技术的理论基础。

(2) 中文语词处理的研究。

5. 有关中文信息处理其他学术问题。

§1.3.3 汉字信息处理技术的发展前景

一、技术开发

为了不断提高汉字信息处理技术并扩大使用范围，应做好以下几方面的工作：

(1) 加强汉字信息处理技术基础理论的研究工作，使这项技术不断向深度和广度方面发展。

(2) 重视汉字信息处理技术标准化的研究和标准的制定，逐步完善标准体系，以利于汉字设备的工业生产和推广应用。

(3) 对于各种汉字设备，应注意优选机型，扩大批量生产，不断提高质量。加强型谱系列化工作，加速研制开发新品种，建立我国完整的信息工业体系。

(4) 确定汉字系统体制，加强汉字系统软件的研制工作，做好国产计算机系统的汉字化工作。

(5) 加强汉字系统应用软件的研制工作，移植西文系统应用软件包，扩大汉字系统的