CC-DOS

操作系统技术大全

(续集)—CC-DOS优化版本的设计

钱培德　杨季文
吕　强　朱巧明　编著

清华大学出版社

CC-DOS操作系统技术大全(续集)

CC-DOS操作系统技术大全(续集)

——CC-DOS优化版本的设计

钱培德　杨季文
吕　强　朱巧明　编著

清华大学出版社

内 容 简 介

本书介绍CC-DOS操作系统的优化技术和方法，以CC-DOS优化版为例，全面和深入地阐述系统的显示输出管理模块、键盘输入管理模块和打印输出管理模块的设计思想，详细讨论了这三个模块的设计技术和实现方法，给出了CC-DOS优化版的全部流程图。

本书突出技术性和实用性，并兼顾理论性，可作为计算机专业人员，特别是系统设计者与开发人员的参考书，亦可作为大专院校计算机专业的参考书。

(京)新登字158号

CC-DOS操作系统技术大全(续集)
——CC-DOS优化版本的设计

钱培德　杨季文
吕　强　朱巧明　编著

☆

清华大学出版社出版
北京 清华园
国防科工委印刷厂印刷
新华书店总店科技发行所发行
☆
开本：787×1092 1/16　印张：17　字数：415千字
1992年2月第1版　1993年8月第2次印刷
印数：10001—22000
ISBN 7-302-01150-8/TP·426
定价：12.50元

序

社会信息化的基本特征是“计算机面向人人”，对计算机广大用户而言，最受欢迎的首推自然语言，对炎黄子孙来说，则首推汉语，为使计算机能在我国及其它使用汉语的国家得以普遍推广，必须采用以汉语成份为主的语言作为用户语言，必须汉化成熟的系统软件，必须研究、开发汉语软件，因此研究开发汉字操作系统就具有特殊的重要性。

《CC-DOS 汉字操作系统技术大全》的作者积多年研究汉字操作系统的经验，编写成此书，几经披读，有数善焉。

第一，应用面广，系统性强。

该书汇集、整理、提炼 CC-DOS 操作系统多种版本，结构合理，层次清晰，脉络分明，内容丰富，蔚为大观。作者的巧妙构思和安排，既为初学者奠基础，又为使用者开茅塞，更为深造者辨泾渭。应用面广，系统性强。

第二，内容先进，标志水平。

该书集中了作者在汉字操作系统方面的研究成果，并注意吸取国内外学者在这一领域的研究成果，内容先进，标志了这一领域的新近发展水平。

第三，经历考验，堪称实用。

书中所分析、论述的各种版本的汉字操作系统均已在计算机上实现，多次使用，经历了实践检验，堪称实用。

综上所述，《大全》实为一部不可多得的好书。余乐而为之序。

徐家福

于南京大学

前　言

IBM-PC 系列微型计算机是我国使用的主流机种，目前，它在我国各种行业和各个部门发挥着作用。如果说 IBM-PC 系列机的引入是促使我国计算机应用推广的重要因素的话，那么 CC-DOS 操作系统就是使 IBM-PC 系列机能在我国发挥作用的重要因素。可以说，如果没有 CC-DOS 为这种计算机建立汉字应用环境的话，那么这种计算机根本不可能在我国发挥如此大的作用。因此，CC-DOS 操作系统在我国的计算机应用方面具有不可磨灭的功勋。

我们在《CC-DOS 操作系统技术大全》一书中对 CC-DOS 的典型版本作了全面和深入的分析。大家可以从中领略到这个操作系统的风采，同时也会认识到它的不足之处。我们首先要肯定 CC-DOS，特别是它的设计思想，至今仍为系统设计者沿用。它的结构奠定了微机汉字操作系统的基础，并已成为一种事实上的标准。至于 CC-DOS 的不足之处，我们也应该看到，但是也用不着大惊小怪。一个八十年代设计的操作系统到了九十年代，当然会存在不足的地方，这是十分正常的。另外，我们也必须承认，其中有些不足之处确实是设计中的问题。

我们认为。学习 CC-DOS 的设计思想、技术和方法是十分必要的，如果只学习它的典型版本，就现在来看，就显得有些不足。我们应该在掌握其典型版本的基础上，进而学习此系统的优化技术和方法，把我们的知识提高到一个新的水平。我们在这方面进行了多年的努力，积累了一定的经验和体会。我们在 CC-DOS 的基础上，重新开发了一个 CC-DOS 的优化版本，这是一个全新的微机汉字操作系统。它基本遵循 CC-DOS 的设计思想和结构，但是又对它作了提高，因而该系统的性能要远优于 CC-DOS。在设计过程中，我们又加入了不少新技术和新方法。本书就以这个 CC-DOS 的优化版本为例，介绍 CC-DOS 的优化技术和方法。

全书共有 10 章，分为 2 篇，各篇的内容如下：

第一篇为第 1 章和第 2 章。该篇为系统概述，主要介绍 CC-DOS 优化版本的概况、汉字的特性、汉字 I/O 的数学模型，以及系统用户界面和使用方法。

第二篇为第 3 章至第 5 章。该篇为系统设计思想，主要介绍 CC-DOS 优化版的显示输出管理模块、键盘输入管理模块和打印输出管理模块的设计思想及有关原理。

第三篇为第 7 章至第 10 章。该篇为系统设计与实现，主要介绍 CC-DOS 优化版 3 个模块的设计技术和实现方法，以及汉字库的结构与管理方法。

本书向读者提供了 CC-DOS 优化版的全部流程图。我们将在《CC-DOS 操作系统技术大全(续二)》一书中，向读者提供 CC-DOS 优化版的完整源程序清单及其注释。

我国著名的计算机专家徐家福教授为本书撰写了序，我们向他表示深切的谢意！我们

还要感谢清华大学出版社的老师们对本书写作的支持！

本书由钱培德、杨季文、吕强、朱巧明共同合作完成，并由钱培德定稿。由于我们水平的限制，故书中定有不妥之处，望广大读者不吝指出。

作　者

于苏州大学计算机工程系

目　录

第一篇　系统概述

第二篇　系统设计思想

第三篇 系统设计与实现

第一篇　系统概述

第1章　绪　　论

1.1　CC-DOS优化版总述

1.1.1　汉字操作系统概况

一九八四年，原电子工业部推出CC-DOS2.1，适用于IBM－PC系列机。毫无疑问，CC-DOS2.1的推出，才真正使汉字信息在微型机上处理成为可能。可以说，CC-DOS2.1是我国第一个实用的汉字操作系统，虽然它只是针对DOS在IBM系列的微型机上实现，但其意义在于证明了汉字信息同样可以用计算机来进行处理，为我国汉字信息处理作了巨大的贡献。

CC-DOS以2.1版本为起点，陆续改进升级，直到CC-DOS4.0的推出，汉字操作系统可以说进入了第二个阶段。正如我们在本书首卷中的评价一样，CC-DOS2.1主要达到了一个可行的目标，高效和可靠的目标远未达到。在三个CC-DOS2.1的基本模块中，都表现出效率和可靠性方面的弱点。例如，在键盘输入方面不够方便和快速，在显示输出上适用面不广，在打印输出上字体字形不丰富等等的弱点。另外在一些系统资源的利用和管理上不尽合理。到CC-DOS4.0时，对上述弱点作了或多或少的改进。CC-DOS4.0初步向人们证明了，汉字信息不仅可以在计算机上进行处理，而且可以比较高效地在计算机上处理，甚至于可以更有利于西文字符信息的处理。

以CC-DOS的各种版本为基础，汉字操作系统的开发者们在不同的角度和侧面进行了开发，并且也得到了广泛的应用。例如，2.13系列的操作系统对CC-DOS的打印输出功能作了极大的提升，受到用户的青睐。然后，GWBIOS、UCDOS、联想系统、GAOK系统等以其丰富的字符功能、较强的汉字输入能力、彩色输出功能等吸引了一批用户。接着，金山DOS以其简洁有效桌面系统获得用户信赖。

可以预计，汉字操作系统还将继续发展，一是朝着系统本身标准化通用化的方向发展，二是朝着把功能强劲的应用系统柔和到操作系统的方向发展，而且，后者更能吸引用户。

1.1.2　CC-DOS优化版总述

在本书中，我们把CC-DOS的优化版本称为CCHDOS。CCHDOS对CC-DOS进行了极大的变化，主要着眼于系统方面的标准化和通用化。CCHDOS有普通版和网络版两个

版本，本书仅介绍普通版的情况。

CCHDOS 是一个组合式的系统。CCHDOS 的组合性体现为外部组合性和内部组合性。外部组合性是指 CCHDOS 的三大模块完全可以独立使用，既可以作为别的汉字操作系统的一个组成模块，也可以让这三大模块自行组合。内部组合性是指模块内部也采用了强烈的板块结构，通过拼接系统提供构件，形成不同功能和性能的模块。

CCHDOS 是一个开放式的系统。CCHDOS 采取模块式的结构，一方面是为了系统本身的有效，另一方面也支持了开放性，为模块本身的发展和利用第三方的开发成果创造了条件。本书把 CCHDOS 的重要部分予以介绍和公布，使 CCHDOS 接受用户和专家的指点。CCHDOS 还建立了自己的体系结构，积极倡导汉字操作系统的标准化，并为此进行了努力和尝试。

CCHDOS 是一个通用式的系统。我们在 CCHDOS 中，运用隔离的手段，尽量把系统与物理设备隔开，同时，努力分析与抽象各种物理设备的特性，力图建立通用的设备描述语言，使 CCHDOS 更方便和更有效地运行在目前已有的或将来发展的不同的硬件支持下。

CCHDOS 是一个高性能的系统。CCHDOS 一方面在输入汉字的考虑和设计上独具特色、方便高效，体现了 CCHDOS 作为 CC-DOS 优化版的高性能。高性能的更重要的一个方面体现在，CCHDOS 尽力利用了中高档微型机的资源，例如 286/386 的保护方式的利用、高分辨率适配卡硬件性能的挖掘等，使 CCHDOS 成为名副其实的高性能的汉字操作系统。

总之，CCHDOS 体现出以下几个特点：

(1) 键盘输入技术自成体系

系统提供了十余种汉字输入法，其中有音码，形码，流水码等，有几种编码是最新的。它独具风格，充分利用了人们早期学习的过程，考虑周到，能基本满足各种要求，和各层次的用户使用。系统能快速方便地输入汉字。

① 对大多数用户所熟悉的拼音码作了很大的改进和完善，系统不仅提供了一键的高频字，而且不用切换输入方式便可输入词组，实现了字、词混合输入。系统中的词库由精选的近五千多词条组成。系统同时还提供方便灵活的自定义词组功能，自定义词组的使用和拼音输入混合一起，异常方便，因此，一般用户只要熟悉拼音，经过几分钟的学习就可以快速地输入汉字。

② 系统提供了没有重码的双拼输入法，该输入法尤其适合于专职操作员快速地输入汉字。另外，系统还提供了笔顺和部首两种形码，对于不知字音和一时难以读正音的用户，输入汉字也较为方便。

③ 系统对区位码，电报码也作了改进，特别是区位码的 3 键提示和向前，向后翻页，便于用户记忆和检索，为制表符和图形符的输入提供了方便。

④ 系统除了有拼音词组输入外，还提供了联想式汉字输入，用户可通过功能键，方便地实现汉字联想输入或者取消联想输入。

⑤ 系统提供了与 CC-DOS4.0 外码输入模块联接的软接口。系统本身提供了一些功能键的外接口，以供系统开发使用。

⑥ 键盘管理模块采用了开放式的结构，便于用户进行二次开发。

(2) 打印输出技术完备

系统包含一个独立的打印输出子系统，该子系统能支持多种型号的打印机打印汉字，同时又能作为通用型打印机管理系统独立使用，使用面较广，操作方便。

① 本系统提供宋体，仿宋体，黑体，楷体和繁体等五种 24×24 点阵打印，它们分别具有 32 种字型，基本能满足各种需求。

② 系统提供了 16×16 点阵的汉字打印，具有 72 种字型。

③ 通过键盘或在程序中可进行字型、字体、字间距、行间距、行宽的设置，使文件、表格等输出更加分明、可读。

④ 该打印输出子系统在打印 24×24 点阵汉字时，用户可自己设置一个内部字库(包括基本字库，常用字库和活动字库)，以避免对经常用到的汉字多次读取点阵信息，同时还加快了汉字的打印输出速度。

(3) 最大可能地开发了硬件资源

由于在 DOS 3.X 以下版本只能进行 640KB 的 RAM 管理，但目前与 AT 机兼容的 286、386 机，一般都带有不止 640KB 的 RAM，本系统对这类机器提供了虚拟汉卡功能。

① 将字库装入扩展 RAM，能为用户空出约 240KB 的 RAM 空间，使得没有汉卡的用户得到了汉卡的功能。

② 将自定义词库装入到扩展 RAM，使用户能有一个很大的自定义词库，且用户可随时随地地进行词组的定义，领略 286、386 的优点，得到 286、386 的实惠。

③ 对不具备扩展 RAM 的机器，本系统可将一级字库装入内存，而把符号和二级字库留在硬盘上；或者根据用户的指定要求，把相应的字库部分驻留内存。

④ 对不具备扩展 RAM 的机器，本系统还可以根据用户某个应用程序所使用的汉字，组成一个专用字库，调入内存，做到取有用，舍无用，既留出了内存空间，又不影响系统响应速度。

⑤ 系统可配一块专用汉卡，系统自动识别汉卡的存在。另外，显示输出模块通用性好，与物理设备的隔离度高。

1.2 汉字的特性

汉字信息处理系统具有输入、输出、存储和处理汉字的功能，所以在建立汉字信息处理系统之前，有必要了解一下汉字的特性，以形成具有较佳性能的汉字信息处理系统。我们在这儿仅讨论与汉字信息处理最密切相关的汉字特性，它们是汉字字量、汉字字频、汉字字序、汉字字形和汉字字音。

1.2.1 汉字字量

汉字的字量十分庞大，据统计，至今已有六万多个形体不同的汉字，这是西文字符量无法与之比较的。但是，目前通常使用的汉字仅为这六万多个汉字中的 10%左右，也就是说，绝大多数的汉字是几乎不用的“冷字”或根本不用的“死字”。汉字信息处理系统应该支

持多少个汉字为好，这是一个值得研究的问题。显然，把六万多个汉字全部收入汉字信息处理系统是不合适的，这样会极大地浪费宝贵的存储资源。为了汉字信息处理系统的需要，国家于 1981 年正式发布了国家标准——《信息交换用汉字编码字符集基本集》，该集总共收入汉字 6763 个，将其中比较常用的 3755 个定为一级汉字，将其余 3008 个定为二级汉字。实际使用证明，基本集中的 6763 个汉字已能满足一般使用的需要。考虑到某些特殊部门和特殊方面使用汉字的需要，国家又于 1986 年审定通过了上述基本集的扩充字符集，它们是《信息交换用汉字编码字符集第二辅助集》和《信息交换用汉字编码字符集第四辅助集》。第二辅助集收入汉字 7237 个，第四辅助集收入汉字 7039 个，基本集和两个辅助集共收入汉字 21039 个。考虑到我国港、澳、台地区和海外华人界仍大量使用繁体汉字，故繁体字也应归入汉字信息处理的范畴之中。为此，国家拟定了《信息交换用汉字编码字符集第一辅助集》、《信息交换用汉字编码字符集第三辅助集》和《信息交换用汉字编码字符集第五辅助集》，这三个辅助集分别收入了基本集、第二辅助集和第四辅助集对应的繁体字集。

1.2.2 汉字字频

汉字字频是指汉字的相对使用频率，这是一项统计数据。例如对 50 万字的资料进行统计，其中的“的”字总共使用了 2 千次，则说该汉字的字频为 0.4%。若干个汉字的字频总和，称为字频的累计数。据统计所得，在汉字数较少时，字数的增加使字频累计数增长较快；当汉字数足够多时，字数的增加使字频累计数的增长极为缓慢。电子工业出版社于 1988 年出版的《汉字字频度统计》一书中，共有汉字 5991 个，它们按字频递减序排列，即使用频率较高的汉字排在较前面。前面 163 个汉字的字频累计数已达到 50%，前面 450 个汉字的字频累计数达 90%，前面 1375 个汉字的字频累计数达 95%，前面 2430 个汉字的字频累计数达 99%。我们还应该注意到，汉字的字频在不同的使用领域中是有区别的。例如，“酐”字在计算机领域中几乎不用，而在化工领域中却成为字频较高的字。人们又发现，不同应用领域中有相当一部分汉字是相同的，不同的仅是那些与专业有关的一部分字，前者是通用的，它们的字频值一般都较高，后者在本专业范围内字频值较高，在其他专业范围内字频值较低。

1.2.3 汉字字序

汉字的字序是指汉字的排列顺序。按照某种(或某几种)规则对汉字进行排序，这对于检索处理是十分必要的。最为常用的排序方法为音序法和形序法。音序法是一种很方便的排序方法，其排列方法是把每个汉字按其汉语拼音的字母顺序进行排列。形序法也是一种较为常用的汉字排序方法，它按汉字的字形排列，涉及到汉字笔画的数目、笔形、汉字的部首等，目前流行的形序有多种。《信息交换用汉字编码字符集基本集》中的汉字采用音序法和形序法排序。基本集内的一级汉字的排序主要依据音序法，即按汉字的汉语拼音字母排列，为了区别同音字，又兼用了形序法。在同音字内，以起笔笔形的顺序排列，起笔相同时，再按第二笔，依次类推。基本集内的二级汉字的排序依据形序法的部首顺序，共用 186 个部首，部首次序按部首笔画数顺序排列，起笔相同时，再按第二笔，依次类推。部首相同

的汉字按除去部首的笔画数排序，若笔画数相同，则按起笔画顺序排列，起笔相同时则按第二笔，依次类推。

1.2.4 汉字字形

汉字字形是汉字形体结构的图像，每个汉字在平面上都占有一个特定的位置，这个位置的明显特点是呈方块形。不论汉字的笔画为多少，每个字必须在一个方块内写成。为了获取汉字的字形信息，在汉字信息处理中，人们常常把方块状的汉字在不同层次上进行分解，以满足不同的加工要求。目前对汉字字形的分解方法和分解标准尚未统一，现有的字形分解方法大体上可以分为四个层次，即单字、字根、笔画和形素。单字由若干个(或一个)字根组成，根据这些字根在方块形区内的不同分布，可以把单字分成多种结构类型，通常可分独体型、左右结构型、上下结构型和内外结构型四种。字根是组成单字的基本结构单元，它本身由笔画组成。它的基本要求是组字能力强，组成的单字形体匀称。在字根层次上进行分解时，随着分解方法的不同，字根的数量差异较大，目前实际上常用的字根数量为 100 到 300 个，习惯上的偏旁部首中的大多数一般都被选择为字根。汉字是一笔一笔地写成的，每一次从落笔到提笔，便是一个笔画。按笔画层次的分解方法不同，笔画数量也不同，少则四种，多则数十种。目前在计算机汉字信息处理中较多选用五种笔画，它们是横、竖、撇、点或捺、折。形素是指把汉字的方形区域细分成若干个小方格(或点)，每个小方格便是一个基本形素。在方形范围内，凡笔画经过的小方格便形成黑点，笔画不经过的小方格便是白点。若把黑点定义为“1”，白点定义为“0”，那么一个小方格恰好与一个二进制位相对应，因此也把基本形素称作二进制位点。由二进制位点组成的汉字字形称为汉字的点阵字形。把汉字在形素层次上进行分解，完全是出于信息处理的需要。

1.2.5 汉字字音

汉字字音是汉字的语音表达形式。汉字是单音节文字，每字一个音节，组成音节的最小单位是音素。汉语普通话共有主要元音音素 6 个，辅音音素 22 个。一个元音音素可以独立构成一个音节，绝大多数的汉字音节由一个辅音音素和一个元音音素构成。其中辅音部分称为声母，元音部分称为韵母，共有声母 22 个、韵母 35 个。不过，不是声母和韵母的任意组合都能构成汉字音节的，实际上只能构成 410 余个音节，为此再用五种声调对音节加以区分。这样，汉字的带声调音节共有 1200 余个。这就是说，六万多汉字的字音仅为 400 余个(不带声调)，或 1200 余个(带声调)。这种十分规范的音节给汉语语音处理带来方便。

1.3 汉字输入与输出的数学模型

汉字输入与输出是计算机汉字信息处理系统内的最基本部分，对汉字输入与输出的数学模型作一探讨，有助于讨论汉字输入与输出模块的结构、效率和性能，并有助于汉字信息处理系统的设计与规范化。因此，在本节中将对汉字输入与输出的数学模型作一初步介绍。

1.3.1 基本定义

1. 汉字集与编码集

定义 1-1:汉字集

若干个汉字组成之集合称为汉字集。

$$H=\{h_i|i=1,2,\ldots,n\}$$

其中,H 为汉字集,h_i 为汉字,n 为正整数,汉字集大小 $|H|=n$。我们经常要以“信息交换用汉字编码字符集基本集”(GB2312)为研究对象,若用 H_0 表示该基本集,则 $|H_0|=6763$。

定义 1-2:码符集

编码组成符之集合称为码符集。

$$C=\{C_i|i=1,2,\ldots,l\}$$

其中,C 为码符集,C_i 为码符,l 为正整数,码符集大小为 $|C|=l$。

定义 1-3:编码

按一定规则组成的码符序列称为编码。

$$e=c_1c_2\ldots c_m \qquad m<=l$$

并且 $$c_j\in C \quad (j=1,2,\ldots,m)$$

其中 e 为编码,c_i 为码符。若 m 为常数,则码长 $|e|$ 为常数,这种编码称为码长为 m 的等长码。

定义 1-4:编码集

若干个编码组成之集合称为编码集。

$$E=\{e_i|i=1,2,\ldots,r\}$$

$$e_i=c_{i1}c_{i2}\ldots c_{im},\ c_{ij}\in C\ (j=1,2,\ldots,m)$$

其中,E 为编码集,e_i 为编码,r 为正整数。编码集 E 的大小 $|E|=r$,r 取决于 m 个码符的排列规则,故

$$r<Pm$$

2. 映射

定义 1-5:一对一映射

设 f 是集合 A 到 B 的一个映射,若由 $f(a_i)=f(a_j)$,可推出 $a_i=a_j(a_i,a_j\in A)$,则称 f 为 A 到 B 的一对一映射。

定义 1—6:满射

设 f 是集合 A 到 B 的一个映射,若 $f(A)=B$,则称 f 为 A 到 B 的满射。

定义 1-7:一一对应映射

设 f 是集合 A 到 B 的一个映射,若 f 既是 A 到 B 的一对一映射,又是 A 到 B 的满射,则称 f 为 A 到 B 的一一对应映射。

1.3.2 汉字输入的数学模型

1. 汉字属性序列

定义 2-1:汉字属性集

若干个汉字属性元素的集合称为汉字属性集。

$$A=\{a_i|i=1,2,\ldots,k\}$$

其中，A 为汉字属性集，a_i 为属性元素，k 为正整数。

定义 2-2：汉字属性序列

按一定规则组成的属性元素序列称为汉字属性序列。

$$S=a_1a_2\ldots a_t, \qquad t<=k$$

$$a_j\in A \ (j=1,2,\ldots,t)$$

其中，S 为汉字属性序列，a_j 为属性元素。

定义 2-3：汉字属性序列集

若干个汉字属性序列组成之集合称为汉字属性序列集。

$$S=\{s_i|i=1,2,\ldots,q\}$$

$$s_i=a_{i1}\ a_{i2}\ldots a_{it},\ a_{ij}\in A \ (j=1,2,\ldots,t)$$

其中，S 为汉字属性序列集，s_i 为汉字属性序列，q 为正整数。

汉字的属性有多种，例如音素、形素、笔画、部首、声母、韵母等，在不同的场合可以选用不同的属性。一个汉字属性序列可以描述出一个汉字，而且它们之间是一一对应的，故有

$$s_i=f_0(h_i)$$

f_0 是 H 到 S 的一一对应映射。

2. 汉字输入过程

目前，汉字输入方法有三种：语音识别输入、字形识别输入和汉字编码输入（亦称键盘输入）。这三种汉字输入方法都有一个共同之处，它们致力于抽取汉字的属性信息，并由此完成相应汉字的输入。其相异之处在于，它们可以把汉字输入过程归结为

$$d_i=f(s_i),\ s_i\in S$$

其中，d_i 为汉字内码，它属于汉字内码集 D。关于汉字内码和汉字内码集的定义，我们将在 1.3.3 内给出。这儿要指出，d_i 与 h_i 是一一对应的，汉字的输入即意味着其相应的汉字内码的输入。映射 f 为由汉字的属性序列转换成该汉字内码的过程，我们要求这种转换是唯一的，故 f 应是 S 到 D 的一一对应映射。

在汉字的语音识别输入和字形识别输入中，f 由计算机和有关装置共同完成。在汉字编码输入中，f 则由用户（人）和计算机合作完成。下面对汉字的编码输入过程作进一步的讨论。

3. 汉字输入码

根据定义 1-2、定义 1-3 和定义 1-4，可以得到输入码符集 U，输入码 R，输入码集 V 的定义。

定义 2-4：输入码符集、输入码、输入码集

$$U=\{u_i|i=1,2,\ldots,l\}$$

$$v=u_1,u_2,\ldots,u_m\ u_j\in U \qquad (j=1,2,\ldots,m)$$

$$V=\{v_i|i=1,2,\ldots,y\}$$

$$v_i=u_{i1}u_{i2}\ldots u_{im}$$