

电视讲座教材

北京计算机教育培训中心

清华大学出版社

联合推出

# 自然码 6.0 实用教程

周山笑 著



清华大学出版社

电视教育讲座教材

# 自然码 6.0 实用教程

周山芙 著

清华大学出版社

JS121/15  
(京)新登字 158 号

## 内 容 简 介

自然码 6.0 是自然码汉字输入法的最新版本,它继承了以前版本的优点,增加了大量新功能,重新设计了全新的屏幕界面,使整个系统更加友好,更易于使用。

本书以教科书的风格向读者全面介绍自然码 6.0 的使用,尊重介绍其中的新增功能,语言通俗,层次清楚。

本书不仅适用于自然码 6.0 的初学者,而且对于已经使用过自然码以前版本的老用户,也不失为一本理想的教科书和使用手册。

**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: 自然码 6.0 实用教程

作 者: 周山芙 著

出版者: 清华大学出版社(北京清华大学校内,邮编 100084)

印刷者: 北京丰台丰华印刷厂

发行者: 新华书店总店北京科技发行所

开 本: 850 × 1168 1/32 印张: 3.75 字数: 96 千字

版 次: 1997 年 8 月第 1 版 1997 年 8 月第 1 次印刷

书 号: ISBN 7-302-02636-X/TP · 1356

印 数: 0001~6000

定 价: 4.50 元

# 序

人们在兴高采烈地迎接 21 世纪，所有的人都认识到：即将到来的 21 世纪将是一个高度发展的信息时代。计算机是处理信息的主要工具，计算机知识已成为当代人类文化不可缺少的重要组成部分。为此，我国近年来掀起了一个全国范围的学习计算机热，各行各业的人都迫切地要求学习和掌握计算机知识。

北京计算机教育培训中心的宗旨就是推动全社会的计算机普及。我们拥有一批具有丰富经验的计算机教育专家，他们熟悉广大读者的特点和认识规律，具有数十年的教学经历，撰写过多种受读者欢迎的计算机著作。北京计算机教育培训中心的理事长是我国著名的计算机教育专家、全国高等院校计算机基础教育研究会理事长谭浩强教授，副理事长是著名计算机专家、《个人电脑》主编、南开大学教授刘瑞挺和清华大学教授、中国计算机学会普及委员会主任吴文虎等。

为了帮助广大读者学习计算机知识，北京计算机教育培训中心特意组织编写了一些计算机读物。这些读物的特点是：面向广大初学者，以应用为出发点，力求能使读者学以致用。它们不是系统的理论教材，只是帮助读者较快地学习计算机在某一个方面的应用，因此它们不要求读者有系统的计算机理论知识基础。

此外，北京计算机教育培训中心还录制了不同层次的计算机应用系列讲座的录像带，并先后在中央电视台等多家电视台播出，社会反映良好，收到了显著的效果。

应当强调说明的是：计算机技术是实践性很强的一门学科，

必须通过大量的实践应用才能真正深入地掌握它。只靠看书和录像是难以真正掌握计算机应用的。正如在课堂上是无法学会游泳的一样，要学游泳必须下到水中去。同样，要学习计算机应用，就必须亲自上机、反复操作，所谓熟能生巧，只要方法得当，您在计算机上花的时间愈多，收获就愈大。

《自然码 6.0 实用教程》一书是为配合自然码汉字输入系统 6.0 版而编著的。自然码 6.0 是在原自然码 5.61 的基础上，结合目前广大用户的普遍需要，由周志农、王磊等合作设计制作的、具有许多独创性功能的新软件，是非常优秀的汉字输入方法之一。本书配有教学讲座录像带，该讲座由自然码的发明人周志农先生本人主讲，并且将在中央电视台、中国教育电视台、四川电视台、陕西电视台等台播出。

我们希望和同行们以及广大读者共同努力，为我国的计算机普及事业做出贡献！

北京计算机教育培训中心

1997 年 6 月

# 前 言

计算机技术的发展日新月异，微型计算机的使用已深入到社会的各个领域，特别是网络通信技术的迅猛发展已经把人类社会无情地推上了信息高速公路的飞车。不能跟上这伟大的时代飞车，必然会被时代抛弃。具有五千年文明历史的中华民族必然要站在时代的前列。代表着中华文明的汉字也必然要深入高科技的各个领域，特别是计算机信息处理领域。要让计算机处理汉字，首先必然要解决计算机的汉字输入问题。自然码汉字输入法就是千百种汉字输入法中十分优秀的一种。

自然码是周志农先生于 1989 年推出的汉字输入系统。1989 年推出 2.1 版，1990 年推出 3.1 版，1991 年推出 4.1 版，1992 年推出 5.0 版，1994 年又连续推出 5.2 版、5.4 版和 5.6 版：几乎是一年一版，升级换代十分迅速。在 5.6 版推出以后，很多用户感到各种能够想到的功能几乎都已经实现了，因此认为自然码不会有太大的变化了。谁想到在一片新的汉字输入法不断问世又迅速销声匿迹的火爆声中，经过两年多的默默无闻和埋头苦干，自然码的 6.0 版本终于推出了！

新版本全面继承了原版本的优点，坚决废除了一些华而不实或过时的东西，增加了大量新功能，重新设计了全新的屏幕界面。特别是为适用于 Windows 系统而独创了“剪切板”造词、多任务多用户操作等许多特殊功能。打开新的自然码系统就会有一种耳目一新的感觉，使用自然码竟然是一种享受！

本书既考虑到使用户尽快学会操作使用的目标，又注重基本

概念、基本知识的掌握。本书以教科书的风格全面介绍自然码内容，层次清楚，深入浅出，无论对于初学者还是自然码的老用户都不失为一本理想的教科书和使用手册。

本人为配合自然码汉字输入法的学习，于 1994 年 12 月在清华大学出版社出版了《自然码实用教程》一书，受到了广大自然码用户的热情支持和欢迎。本书应当说是上一本书的姐妹篇。上一本书主要介绍的是自然码 5.6 版本，这一本书主要介绍的是自然码 6.0 版本。它们既可以单独学习使用，也可以互相参考使用。由于要赶上自然码 6.0 版的发布，时间紧张，书中必然会存在一些差错和不足，恳请广大的读者和用户批评指正。

周志农先生已经授权北京计算机教育培训中心和清华大学出版社发行部代销自然码 6.0 软件。读者购买软件可以和以上单位联系。

北京计算机教育培训中心地址：北京地安门黄化门街 5 号

电话：64031608 64078663

清华大学出版社发行部地址：北京清华大学校内

邮政编码：100084

电话：62784891 62781728 62781737

周志农先生电话：68348899 呼 61628。

编 者

# 目 录

|                             |    |
|-----------------------------|----|
| 第一章 计算机汉字处理简述 .....         | 1  |
| 1.1 计算机的汉字处理 .....          | 1  |
| 1.1.1 汉字系统 .....            | 1  |
| 1.1.2 汉字的编码 .....           | 2  |
| 1.1.3 汉字库 .....             | 8  |
| 1.1.4 汉字的处理方式 .....         | 10 |
| 1.2 汉字输入法概述 .....           | 11 |
| 1.2.1 输入方法的分类 .....         | 11 |
| 1.2.2 汉字键盘输入的发展过程 .....     | 13 |
| 第二章 自然码 6.0 操作基础 .....      | 15 |
| 2.1 自然码 6.0 的安装与驻留 .....    | 18 |
| 2.1.1 自然码 6.0 的安装 .....     | 18 |
| 2.1.2 自然码 6.0 的启动和驻留 .....  | 20 |
| 2.2 自然码 6.0 的设置 .....       | 28 |
| 2.2.1 在 DOS 状态下设置 .....     | 28 |
| 2.2.2 在 Windows 状态下设置 ..... | 32 |
| 2.3 单字的输入 .....             | 42 |
| 2.3.1 简码输入法 .....           | 42 |
| 2.3.2 标准输入法 .....           | 45 |
| 2.4 词的输入 .....              | 49 |
| 2.4.1 两字词输入 .....           | 50 |
| 2.4.2 三字词输入 .....           | 53 |



|              |                      |            |
|--------------|----------------------|------------|
| 2.4.3        | 四字词输入 .....          | 53         |
| 2.4.4        | 多字词输入 .....          | 54         |
| <b>第三章</b>   | <b>自然码操作进阶 .....</b> | <b>56</b>  |
| 3.1          | 单字输入的技巧 .....        | 56         |
| 3.1.1        | 非标准发音输入法 .....       | 56         |
| 3.1.2        | 特殊字的输入 .....         | 59         |
| 3.1.3        | 利用音节查找韵母 .....       | 60         |
| 3.1.4        | 简明部首输入法 .....        | 61         |
| 3.2          | 词输入的技巧 .....         | 64         |
| 3.2.1        | 自动顶词 .....           | 64         |
| 3.2.2        | 字、词的叠加输入 .....       | 66         |
| 3.2.3        | 全拼和双拼混合状态 .....      | 67         |
| 3.2.4        | 词中选字 .....           | 69         |
| 3.3          | 自然码 6.0 的特殊操作 .....  | 72         |
| 3.3.1        | 语句输入功能 .....         | 72         |
| 3.3.2        | 自造词与自造短语 .....       | 74         |
| 3.4          | 辅助功能 .....           | 88         |
| 3.4.1        | 常用中文标点的输入 .....      | 88         |
| 3.4.2        | 特殊符号的输入 .....        | 89         |
| 3.4.3        | 中文数字、日期等的输入 .....    | 91         |
| <b>附录</b>    | <b>.....</b>         | <b>93</b>  |
| 附录 A         | 词频自动调整程序的使用 .....    | 93         |
| 附录 B         | 旧自造词库的转换 .....       | 97         |
| 附录 C         | 简码字库的修改 .....        | 98         |
| 附录 D         | 简码词库的修改 .....        | 100        |
| <b>练习与测试</b> | <b>.....</b>         | <b>103</b> |

# 第一章 计算机汉字处理简述

计算机是西方人的发明，它能够方便地处理西文，但是却不能处理中文。在中国要想普及和发展计算机技术，就必须使计算机具有中文信息的处理能力。所谓中文信息处理能力，就是使计算机能够输入汉字、存储汉字、传输汉字、编辑修改汉字和输出汉字。

汉字输入是计算机中文信息处理的关键。随着计算机技术的普及和推广，计算机的应用已经深入到企业、公司、机关、学校及家庭等各个领域。计算机汉字输入早已不是少数专业录入人员的工作，而成为了千千万万个普通干部、技术人员、职工、作家、记者、教师和学生不可缺少的日常工作。如果不能高效率地解决他们的汉字输入问题，要真正实现中文信息处理的现代化将是不可能的。本书的目的就是为广大非专业汉字录入人员开辟一条迅速通往计算机汉字输入理想境界的金光大道。

## 1.1 计算机的汉字处理

### 1.1.1 汉字系统

DOS 汉字系统是加载在英文 DOS 操作系统之上的一个汉字处理模块；Windows 汉字系统是加载在英文 Windows 操作系统之上的一个汉字处理模块。汉字系统不能脱离英文操作系统而独立使用，所以我们常常把它叫做中文环境、中文平台或中文外壳，其实质就是在英文操作系统功能的基础上，增加了汉字的输入、

识别、存储、处理和输出功能。如果把英文操作系统和汉字系统融合成一个整体，就叫做汉字操作系统。例如：DOS 方式下的 PDOS，Windows 方式下的 PWindows 3.2 和 Windows 95 中文版等。

目前使用比较广泛的 DOS 汉字系统有北京希望电脑公司的“UCDOS”、北京怡江新技术实业发展公司的“天汇系统”及香港金山公司和北大方正公司的“SPDOS”。其中 UCDOS 的 3.1 版本和 5.0 版本是最流行的汉字系统。使用比较广泛的 Windows 汉字系统主要有方正新天地公司的“中文之星”，四通利方公司的“利方多元系统 RichWin”等。

### 1.1.2 汉字的编码

为了规范我国汉字的使用，国家颁布了汉字的国家标准“中华人民共和国国家标准信息交换汉字编码”，它的代号是“GB2312—80”，一般地把它简称为“国标码”。

在国标码中把常用的 6763 个汉字分成两级：一级是最常用汉字，共有 3755 个；二级为次常用汉字，共有 3008 个；再加上 682 个图形符号，一共收录了 7445 个汉字图形符号。

国标码把全部汉字符号按顺序排列在一张 94 行乘 94 列的二维表中，这张表叫做国标图形代码表，如表 1-1 所示。在表中，一级汉字排在前面，二级汉字排在后面。一级汉字按汉语拼音字母的顺序排列，多音字取常用发音，同音字再按起笔的笔形(按横、竖、撇、点、折)顺序排列；若第一笔相同，则按第二笔顺序，依此类推。二级汉字按部首顺序排列。

国标图形代码表中的每一行称为一个“区”，每一列称为一个“位”；所以国标码表中有 01～94 共 94 个区，而每个区又有 01～94 共 94 个位。行与列(或者说区与位)的交点就是每个汉字图形符号的坐标，也可以说是它在表中的位置地址。

表 1-1 国标图形代码表(GB2312 - 80)

|                                                 |   |   |   |   |   |   |   |      |    |    |    |    |      |      |      |     |    |    |   |   |    |   |   |   |   |   |   |
|-------------------------------------------------|---|---|---|---|---|---|---|------|----|----|----|----|------|------|------|-----|----|----|---|---|----|---|---|---|---|---|---|
| 例如：<br>唉 = 16 05<br>自 = 55 52<br><br>( 用区位码表示 ) |   |   |   |   |   |   |   |      |    |    |    | 21 | 22   | 23   | 24   | 52  | 53 | 54 | 7 | 7 | 7E |   |   |   |   |   |   |
|                                                 |   |   |   |   |   |   |   |      |    |    |    |    |      |      |      |     |    |    |   |   |    |   | C | D |   |   |   |
|                                                 |   |   |   |   |   |   |   | 第二字节 | b7 | 0  | 0  | 0  | 0    | 1    | 1    | 1   | 1  | 1  | 1 | 1 | 1  | 1 | 1 | 1 | 1 | 1 | 1 |
|                                                 |   |   |   |   |   |   |   |      | b6 | 1  | 1  | 1  | 1    | 0    | 0    | 0   | 0  | 0  | 0 | 0 | 0  | 0 | 0 | 0 | 0 | 0 | 0 |
|                                                 |   |   |   |   |   |   |   |      | b5 | 0  | 0  | 0  | 0    | 1    | 1    | 1   | 1  | 1  | 1 | 1 | 1  | 1 | 1 | 1 | 1 | 1 | 1 |
|                                                 |   |   |   |   |   |   |   |      | b4 | 0  | 0  | 0  | 0    | 0    | 0    | 0   | 0  | 0  | 0 | 0 | 0  | 0 | 0 | 0 | 0 | 0 | 0 |
|                                                 |   |   |   |   |   |   |   |      | b3 | 0  | 0  | 0  | 0    | 1    | 0    | 0   | 1  | 0  | 0 | 1 | 1  | 1 | 1 | 1 | 1 | 1 | 1 |
|                                                 |   |   |   |   |   |   |   |      | b2 | 0  | 1  | 1  | 0    | 1    | 1    | 0   | 1  | 1  | 0 | 0 | 0  | 0 | 0 | 0 | 0 | 0 | 0 |
| b1                                              | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0    | 1  | 0  | 0  | 1  | 0    | 0    | 1    | 0   | 0  | 1  | 0 |   |    |   |   |   |   |   |   |
| 第一字节                                            |   |   |   |   |   |   |   | 位    | 01 | 02 | 03 | 04 | 50   | 51   | 52   | 92  |    | 94 |   |   |    |   |   |   |   |   |   |
|                                                 |   |   |   |   |   |   |   |      | 区  |    |    |    |      |      |      |     |    |    |   |   |    |   |   |   |   |   |   |
| 21                                              | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 01   | SP | ,  | .  | .  | ∫    | §    | ≡    | ↑   | }  | ≡  |   |   |    |   |   |   |   |   |   |
| 22                                              | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 02   |    |    |    |    | (14) | (15) | (16) | XII |    |    |   |   |    |   |   |   |   |   |   |
| 23                                              | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 03   | !  | "  | #  | ¥  | R    | S    | T    |     | }  | —  |   |   |    |   |   |   |   |   |   |
| 30                                              | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 16   | 啊  | 阿  | 埃  | 挨  | 耙    | 坝    | 霸    | 包   | 褒  | 剥  |   |   |    |   |   |   |   |   |   |
| 31                                              | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 17   | 薄  | 苞  | 保  | 堡  | 币    | 庇    | 痹    | 秉   | 饼  | 炳  |   |   |    |   |   |   |   |   |   |
| 57                                              | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 55   | 住  | 注  | 祝  | 驻  | 滓    | 子    | 自    |     |    |    |   |   |    |   |   |   |   |   |   |
| 7D                                              | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 93   |    |    |    |    |      |      |      |     |    |    |   |   |    |   |   |   |   |   |   |
| 7E                                              | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 94   |    |    |    |    |      |      |      |     |    |    |   |   |    |   |   |   |   |   |   |

比如“自”字，处在码表中 55 行 52 列的位置。如果用十进制数表示，它的地址是 55 52，如果用二进制数表示，可以用两个字节，地址是 00110111 00110100(用十六进制数表示是 37 34)。

我们把表中每个汉字的十进制数地址称为汉字区位码，利用区位方式在计算机键盘上输入汉字的方法，就叫做区位码汉字输入法。为避开 ASCII 编码中 00H ~ 20H 一段的控制码，在每个汉字区位码的二进制数形式的每个字节上都加上二进制数 00100000(用十六进制表示是 20)，就变成了汉字图形符号的对应国标码。比如：我们已经知道“自”字区位码用二进制表示是 00110111 00110100(十进制是 55 52，十六进制是 37 34)，那么它的国标码就应当是 01010111 01010100(十六进制是 57 54)。这正是 GB2312 - 80 图形字符代码表中给出的形式。

在国标码码表中，01 ~ 15 行是图形符号区；16 ~ 55 行是一级汉字区；56 ~ 87 行是二级汉字区；88 ~ 94 行是空的，为自定义区。

#### 1. 汉字的内码

计算机只能识别二进制数码 0 和 1，任何信息如果要用计算机处理，就必须用 0 和 1 对它们编码，所以汉字信息也必须用二进制数 0 和 1 来编码。既然国标码的每个汉字都可以用它的两个字节的二进制地址唯一地识别，那么是否就可以直接用国标码作为计算机内部使用的汉字编码呢？答案是否定的，因为在计算机中除了要使用汉字编码以外，还要使用 ASCII 编码(已被国际标准化组织认定为国际标准的美国标准信息交换码)，而 ASCII 编码是很容易与国标码混淆的。

比如计算机要识别处理英文和汉字的混合信息“F面”，其中字母“F”的 ASCII 码是 01000110，“面”字的国标码是 01000011 01100110，“F面”的二进制表示是就 01000110 01000011 01100110。但对于这样一组信息，计算机却可以理解成是三个独立的 ASCII 码字符“FCf”，或前面两个字节为汉字后面一个为 ASCII 字符，即为“喻 f”。为了避免这种混淆，就必须对国标码进行必要的改造。

通过分析,我们发现 ASCII 码实际上只使用了一个字节中的低(后)7 位,最高位是用来检查代码正确性的校验位,这一位总是 0。如果使汉字编码每个字节的最高位总是 1,就可以使它区别于 ASCII 码字符了。这也就是要在国标码的两个字节上各加上二进制数 10000000(用十六进制表示是 80)。

经过改造,国标码就变成了目前在计算机中统一使用的汉字编码。因为这种两字节编码是汉字在计算机内部的代码,所以就称为汉字机内码,或者简称内码。

汉字内码与国标码的对应关系可以写成下面的十六进制公式:

内码 = 区位码 + 20H + 80H = 国标码 + 80H (H 表示十六进制)

汉字内码与国标码的对应关系也可以写成下面的十进制公式:

内码 = 区位码 + 32D + 128D = 国标码 + 128D (D 表示十进制)

比如“自”字在国标图形代码表中的国标码是: 01010111 01010100B(B 表示二进制),转换成相应的内码就应当是: 11000011 11100110B(C3E6H)。转换方式如下:

01010111 + 10000000 = 11010111                      高字节

01010100 + 10000000 = 11010100                      低字节

## 2. 汉字外码

有了汉字内码以后,在计算机内部就可以处理汉字了。但是要让计算机处理汉字信息,首先必须将汉字信息送入计算机。要将汉字图形符号输入计算机,可以用语音转换输入、光电输入、联机手写输入和键盘输入等许多方法,但目前最成熟、最简便易行和实用化的还是键盘输入。

所谓键盘输入,就是通过敲击计算机键盘上的键钮,把全部汉字或者各种符号输入计算机。我们知道,计算机键盘上标出的都是英文字母或符号,我们可以直接敲击相应键位来输入英文字

母或符号；但是汉字不同于英文字符，在计算机键盘上是找不到相应汉字的键位的。为了能在英文键盘上输入汉字，人们借鉴在计算机内部使用 0 和 1 就可以对全部汉字编码的思想，先用键盘上一串固定的英文字母键及数字符号键对每一个汉字进行编码，再以键入这一串键序列来间接地输入汉字；这就是目前键盘汉字输入的基本方式。这一套为每一个汉字定义的键序列就叫做汉字输入码，或称汉字外码。

因为汉语拼音字母和英文字母十分相似，人们首先就想到了用汉语拼音输入汉字。我们若把汉语拼音的字母看成是英文字母，再把汉语拼音中的特殊字符用相应字母代替(如韵母 u 用字母 V 代替)，就可以在计算机键盘上用拼音输入汉字了，这样的键盘编码叫做汉字的拼音码。

在键盘上利用拼音输入汉字，既直观又简单方便，但是也存在一些明显的问题，比如：由于许多汉字的拼音很长，输入时需要多次操作键盘按键，而且汉语中同音字很多；同时，由于我国地域广大，各地方言千差万别，许多人也很难准确使用拼音，有时甚至还会碰到不会读音的生字。所以，人们又把汉字按笔划的构成拆成许多基本部件，再把每个部件对应到一个固定的键位上，通过键入相应的键位输入各种汉字，这样的键盘编码就是汉字的字形码。

由于不同目的和环境的要求，根据汉字的不同特性，制定不同的编码规则，便可以产生许许多多不同的汉字输入编码规范。也就是说，可以有许许多多不同的汉字编码系统和汉字输入法。这些编码规范都是人和计算机交换信息时使用的，我们把它们通称为汉字外码。虽然目前有众多的汉字输入法，但是真正比较实用的、用户较多的汉字输入方法也就是十几种。

### 3. 汉字的输出码

汉字的输出主要是指在显示器上或打印机上输出汉字的字

形。汉字字形的表现形式主要有：铅字形式和点阵形式。铅字形适用于打字机和印刷机，而点阵形式则特别适合于计算机的汉字处理。

汉字是一种象形文字，每一个字都可以看成是一个特殊的图形。这种图形很容易用点阵来描述。描述时，把汉字放在一个网状的方格内，凡是有笔划通过的网点为黑点，否则为白点。这样从网状方格的整体来看就是一个汉字。我们把这种网状方格称为点阵。比如图 1-1 就用 16 行，每行 16 个点的点阵描述了一个“王”字。

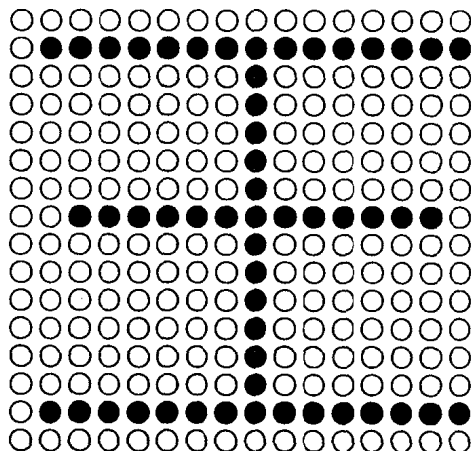


图 1-1

每一个这样的汉字点阵叫做一个汉字的字模(或字形)，字模中的每个点叫做位点。若用 1 表示黑点，0 表示白点，则黑白点信息就可以用二进制数的形式表示。一般说，一个方格就是一个  $m$  行  $n$  列的点阵，其中行数  $m$  总是字节(8 位二进制数)的倍数， $m \times n$  的积就是该点阵的总点数(二进制信息数)。上例是一个  $16 \times 16$  的点阵，它需要用  $2(\text{字节}) \times 16(\text{行}) = 32$  个字节的信息来存



储一个汉字。8000 个汉字就需要  $8000 \times 32(\text{字节}) = 256\text{KB}$ (B 表示字节)的存储容量。用二进制数表示的汉字点阵信息叫做汉字的输出码。汉字输出码又叫做汉字字形码或汉字发生器编码。

#### 4. 汉字控制功能码

能够直接影响汉字数据的格式处理、传输方式和解释执行的代码叫做汉字控制功能码。它总是和汉字数据代码混合组成汉字数据流,并且贯穿于汉字输入、机内处理、汉字输出等整个汉字处理的全过程。

### 1.1.3 汉字库

在计算机内部处理的是汉字内码,而在显示器、打印机上输出的是汉字的字形,所以必须事先把全部的汉字字形都放在计算机的外存储器中,再通过汉字处理系统的专门处理程序自动把要输出的汉字的内码转换成对应的汉字字形后输出。

全部汉字字形的集合叫做汉字字形库(简称汉字库)。汉字库总是以二进制的字形码形式存放在计算机的存储器中。汉字的字符集很大,不包括罕见字、古字的现用汉字就有一万多个。目前我国的汉字系统一般都使用 GB2312 规定的两级常用汉字建立汉字库,即一共有 7445 个汉字图形字符。

上一节已经讲到一个 16 点阵的汉字库就需要大约 256KB 的存储容量,而这样低点阵的汉字是很难看的,一般只在显示时使用;打印输出常用的是  $48 \times 48$  点阵的汉字库。一个 48 点阵汉字库的容量应当是  $6(\text{字节}) \times 48 \times 8000 = 2304\text{KB} = 2.3\text{MB}$ 。显然所需存储容量是很大的。

一般说,表现汉字时使用的点阵数越高,则汉字字形的质量也越好,当然,每个汉字点阵的存储量也越大。通用型汉字点阵常分为三种,它们所需的存储量可由表 1-2 看出。