

● 萧启宏著

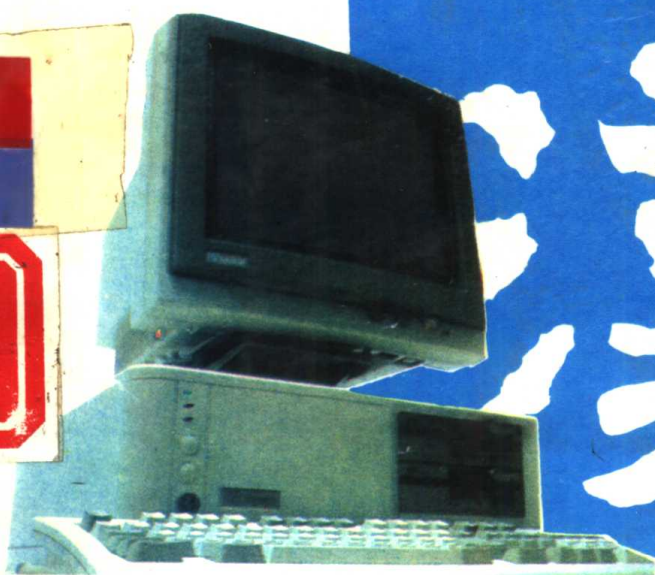
启宏

全息码

全汉字编码输入系统

● 电子工业出版社

全汉字



全汉字编码输入系统

启 宏 全 息 码

萧 启 宏 著

电子工业出版社

(京)新登字 055 号

内 容 提 要

作者历时七年,潜心研究,发明了一种全汉字编码输入技术。该技术专利名称为“汉字全息码计算机输入系统”,专利号:91109162·9-3。在不建词库,无非常规编码的情况下,在解决繁简一体、全汉字输入方面,该项发明处于国内领先地位。本书是作者阐述这种编码技术的专著,全书共五章。前二章,作者追本溯源,介绍文字的起源和三大文字体系,并对汉字的结构、特点从文字学理论出发作了精辟的分析;第三章,作者在集其它编码技术之长的基础上,提出了“三案合一”的启宏全息码基本思路;第四章,详细讲授启宏全息码方案和使用规则;第五章,给出国家一、二级字库 6 763 个汉字的启宏全息码字典,并宣告符合 ISO 10646 CJK 字集的启宏码汉字处理系统已进入实用化阶段。

该书可作为学习全汉字编码输入技术的教材,也可供研究和应用中
文信息处理技术的人员参考。

启宏全息码

萧启宏 著

责任编辑 史明生

*

电子工业出版社出版(北京市万寿路)

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

北京市顺义县李史山印刷厂印刷

开本:787×1092 毫米 1/32 印张:6 字数:125 千字

1993 年 3 月第 1 版 1993 年 3 月第 1 次印刷

印数:5000 册 定价:4.70 元

ISBN7-5053-1864-0/TP·443

信息大世

文化盛事

周有光



注：周有光研究员是国家语委著名语言文字学家、《汉语大词典》顾问

序

所谓中文信息处理,就是要实现用中国的语言文字跟计算机打交道这个目标。首先是让计算机能处理汉字,接着是让它能理解汉语语言,直接和我们对话。这也就是通常所说的人机联系。中文信息处理是手段,目的在于为电子计算机在我国的应用铺平道路。

经济活动是人类一切活动的基础。然而,任何一个经济活动都是一个物质流程,原料进来,产品出去,就是一个物流过程。要想物流畅通无阻,就必须有信息流作保证。随着社会现代化大生产的迅速发展,信息量愈来愈大,靠人力加工信息已经不可能了,必须用电子计算机来参与管理。计算机应用的广度和深度,是衡量一个国家现代化水平高低的重要标志。

由于汉语、汉字所具有的特殊性,怎样让电子计算机能处理汉字、理解汉语,在使用汉语言文字的国家里,建立起人机联系的便利通道,这是一个极其复杂的科学难题。从七十年代起,随着国家现代化经济建设高潮的到来,一个以研究计算机处理汉语言文字为中心内容的中文信息处理学科在中国兴起。

有的外国人曾预言:“使用古老汉字的国家,是不可能实现现代化的”。正是中文信息处理学科所取得的成就,证明了使用汉字的国家不能实现现代化的预言是种偏见,开拓了我国电子计算机应用的各个领域,展示了我国现代化建设的美好前景。

同其它学科发展一样,中文信息处理每向前跨出一步,都会碰到新的技术难关。在解决了 GB-2312 基本字集(常用汉字)处理之后,要迈向全汉字系统处理,由于处理字数从几千字到几万字的数量级变化,新的难关自然形成,只有在克服难关上取得突破,中文信息处理才能获得重大的发展。

九十年代起,在国际标准组织(ISO)分技术委员会和 14 个跨国公司组成的计算机产业集团动议的推动下,中文信息处理迎来了两个喜人的变化:其一,计算机产业界放弃已用多年的八位代码方式,升为十六位代码系统,使仅含有 128 个字符的单字节 ASCII 码,宣布完成了它的历史使命;其二,将 ISO 确认的 ISO 10646 CJK 字集作为中文国际信息交换字符集,使国内已用多年的 GB-2312 基本字集的 6 763 个汉字字数限制成为历史。这两个变化,一方面使汉字以同等身份和西方拼音文字一道,跻身于计算机应用领域;另一方面在汉字文化圈,进行了一次跨国界的统一行动,为汉字在国际上的流通铺平了道路。

中文信息处理发展中的这一重大变化,随之而来的就是解决大型汉字字符集或者是全汉字系统的输入方法问题。

《启宏全息码》一书,是作者肖启宏同志应用他的“91109162.9-3”号发明专利,对《汉语大字典》54 620 个汉字实施编码,建立“全汉字编码输入系统”的基础上写成的启宏全息码理论分析和技术说明专著。这是中文编码字数从几千字迈向几万字的全汉字系统,编码方法从查字法编码迈向文字学编码的一个里程碑。

《启宏全息码》一书,作者用人类文字发展史的丰富史料,论述了汉字在人类文明史上的重要地位,得出了“汉字是我们的祖先用方块符号建筑起来的文化长城”的结论。从分析我国

百余年来汉字改革运动的因果中,指出汉字在教学和信息处理两个方面,与西方拼音文字相比存在差距的原因,是由于没有把汉文字学规律应用于教学和技术处理的缘故。作者站在汉字发源的起点上,提出了以文字学理论为基础的全息理论框架,成功地解决了全汉字系统编码的理论和技术问题,一举实现了“中西文字兼容汉字文化圈资源共享,繁简古今一体炎黄子孙群文化合流”的目标,成为我国第一个建立全汉字系统工程拓荒者。本书作为文字学编码和全汉字系统工程的一个理论和技术专著,具有重大的学术价值和经济意义。

作者把建立通用键盘,全汉字(包括简体、繁体、异体、讹体以及外造汉字)编码输入系统工程看作是一项在大科学时代,不使汉字沉沦,继承和弘扬民族文化,促进国家统一民族团结的伟大事业。这种思想认识也是值得称赞的。

希望《启宏全息码》一书的出版,能为繁荣我国中文信息处理事业、配合全汉字系统的开发和应用起到积极的作用。是为序。

陈力为

一九九二年六月

注:陈力为教授是中国计算机与信息标准技术委员会主任、中国中文信息学会理事长。

前 言

《启宏全息码》是一部研究汉字信息处理键入技术的高科技专著。汉字全息码是用代码替换古人创造汉字的全部材料与方法,复写出全部汉字的汉文字学编码。这里的“全息”概念,不是光学上反映单个汉字字形的每个网点的全息,而是指捕捉古人创造汉字的全部信息和实现对全部汉字编码的全息。

汉字是我们祖先创造出来的、人类最为精确简明的文字符号系统。在人类进入电子计算机广泛应用的信息时代,怎样使这饱含人类智慧、浩如烟海的古老汉字返老还童,迸发出科学的青春,继续为人类社会的进步作出新的贡献,这是时代赋予我们的使命!

“废除汉字。”谈何容易,子孙不容!

“限制汉字数量。”谁给你这样的权力!

“谁主汉字沉浮?”我们,生活在汉字这块故土上的炎黄子孙!

《启宏全息码》一书,便是对当代炎黄子孙如何解决汉字信息处理技术的历史记录和疑难解答。

为了表明中国人在汉字信息处理中忧思深远、深谋远虑的思考,使本书具有广博的知识性、趣味性、可读性,笔者从远古的结绳记事、实物传信到文字的产生;从刀刻笔画到活字印刷的发明;从电磁脉冲电报到电子计算机汉字信息处理,均用信息科学一条红线贯通连接。

01-000000

为了说明汉字信息处理键入技术的核心内容,本书记述了中国信息科学工作者如何从电报数字区位编码到人工查字法编码,迈向文字学造字法编码的艰苦历程,最终使汉字编码走上了理论化、科学化、适用化的胜利坦途!

为了使本书具有适用性,书中详细阐述了全息码方案研究的战略目标,设计思路,理论根据和技术细则,使之成为一本启宏全息码全汉字系统的实用教材。

科学检测的数据表明,获国家专利(专利号:91109162.9-3)的全息码全汉字系统,是高速适用的全汉字系统。它的推广与应用,是使我国中文信息化事业实现“繁简古今一体炎黄子孙群文化合流,中西文字兼容汉字文化圈资源共享”伟大目标的一条途径。

正如著名诗人歌德所说:“每当一种新事物出现的时候,人们总是要问:它有什么用途?人们这样做是没有错的,因为人们只有通过每一事物的用途才能发现它的价值”(《歌德的格言和感想集》第48页)。这正是笔者撰著出版本书的目的。

在本书出版过程中,得到中国中文信息学会理事长、中国计算机与信息标准技术委员会主任陈力为教授,国家语委著名语言文字学家周有光研究员的关心和指导;得到陈昌鑑、尹向阳、黄金平、焦焕文和邹春龄同志在工作上的帮助。谨此致谢!

作者

一九九二年六月一日

目 录

第一章 方块筑起的文化长城	(1)
第一节 蒙昧时代的实物信息	(3)
一、实物信息.....	(3)
二、结绳记事.....	(3)
三、烽火狼烟.....	(4)
第二节 开启文明的文字发明	(5)
一、文字的起源.....	(5)
二、三大文字体系.....	(9)
第三节 举世无双的文化长城	(14)
一、汉字的历史	(14)
二、汉字的特点	(15)
三、汉字的现状	(19)
第二章 汉字长城的危机	(21)
第一节 国外汉字文化圈的断缺	(22)
一、朝鲜改制谚文	(22)
二、越南走向拉丁化	(23)
三、日本采取混用法	(24)
第二节 国内汉字改革的兴起	(25)
一、汉字改革兴起的背景	(25)
二、汉字改革的发展过程	(25)
三、汉字改革成果与汉字信息处理的联系	(27)
第三节 修复长城的战斗	(28)
一、中文信息处理异军崛起	(29)

二、打破神话的汉字编码	(35)
三、徘徊中的困惑和选择	(39)
第三章 回到源头上坦途	(43)
第一节 总体战略 寻根求解	(44)
一、战略目标,双重任务	(44)
二、舍本求源,全息架构	(46)
三、分类聚合,发散推演	(51)
第二节 材料模块 代码固化	(55)
一、字根、字元模块与代码	(55)
二、部首、部件模块与代码	(56)
三、笔画和方案规则代码	(57)
第三节 全息并用 坦途轻骑	(58)
一、图表一览,一目了然	(58)
二、三案合一,避难就易	(58)
三、如来佛掌,坦途轻骑	(60)
第四章 启宏全息码教程	(65)
第一节 键位中心字元及其相关字码释例	(66)
一、键位中心字元分布	(66)
二、键位中心字元及其相关字代码	(67)
练习一	
第二节 键位左上角字元及相关字码释例	(70)
一、键位左上角字元分布	(70)
二、左上角字元代码及其相关字代码	(71)
练习二	
第三节 键位右上角字元及相关字码释例	(74)
一、键位右上角字元分布	(74)
二、键位右上角字元代码及其相关字代码	(75)
练习三	
第四节 键位左下角字元及相关字码释例	(77)

一、键位左下角字元分布	(77)
二、键位左下角字元代码及其相关字代码	(78)
练习四	
第五节 键位右下角部件及相关字码释例	(80)
一、键位右下角部件分布	(80)
二、键位右下角部件及其组字代码	(80)
练习五	
第六节 部首代码及其使用规则释例	(83)
一、部首在键位上的分布	(83)
二、部首组字使用字例	(84)
三、笔画在数字键上的分布	(85)
四、部首组字规则	(85)
练习六	
第七节 36 键位键盘图及规则总表	(87)
一、键盘图	(87)
二、规则总表	(88)
第八节 全息码汉语拼音用键对照表	(89)
练习八	
第九节 一级字元音序排列表及组字释例	(91)
一、按音序排列的一级字元	(91)
二、字例	(93)
练习九	
第十节 一级字元转注部件对照表及其释例	(95)
一、字元代码转注部件	(95)
二、字元转注部件,组字释例	(96)
练习十	
第十一节 一级字元繁简对照表及其释例	(97)
一、一级字元繁简体对照	(97)
二、字例	(97)

练习十一

第十二节	一级字元综合信息笔顺排列检索表	(99)
------	-----------------------	------

第十三节	二级字元音序排列表及组字释例	(102)
------	----------------------	-------

一、按音序排列的二级字元	(102)
--------------------	-------

二、字例	(104)
------------	-------

练习十三

第十四节	二级字元繁简对照表及其释例	(106)
------	---------------------	-------

一、二级字元繁简体对照	(106)
-------------------	-------

二、字例	(106)
------------	-------

练习十四

第十五节	二级字元综合信息笔顺排列检索表	(108)
------	-----------------------	-------

第十六节	笔画代码与方案规则解析	(111)
------	-------------------	-------

一、笔画代码及使用规则	(111)
-------------------	-------

二、方案规则	(112)
--------------	-------

三、方案字码样句	(113)
----------------	-------

练习十六

第十七节	启宏全息码系统一览表	(115)
------	------------------	-------

第五章	启宏全息码字典	(117)
-----	---------------	-------

第一章 方块筑起的文化长城

我们的祖先不仅用方砖在丛山峻岭上，建起了雄冠中华大地的万里长城，而且还用方块符号筑起了世代维系中华民族团结统一的文化长城。这就是五千年来一脉相承的巨大的汉字符号系统，是举世无双、旷世难以再造的人类文明史上的奇观。

文字是人类记录语言、用来交际的信息符号系统。它的发明与应用，是人类社会由原始野蛮时期过渡到文明时期的分界线。正是有了文字，人类的知识得以积累和传播，这才有了社会的快速发展和当今享有的一切文明成果的创造。

东方是人类文明的摇篮。汉字是人类可通向文化源头的时间隧道，它既是贯穿历史几千年的红线，又是千百年来积累人类智慧的知识宝库。

在大科学时代,彻底开发和利用这座宝库的知识资源,对于开创造福于人类的盛德大业有着不可估量的意义。

为此,我们应当了解,我们的祖先建筑起如此辉煌的知识库是一个多么漫长艰难的历程。

第一节 蒙昧时代的实物信息

一、实物信息

人种志的研究成果告诉我们：实物交际出现在文字发明的几千年以前。实物传信也是一种极其简单的方法。比如：印地安人，用短斧作为宣战的信号，用烟管象征和解；苏门答腊人，用盐表示友谊，用辣椒表示仇恨；斯拉夫人用面包和盐表示致敬等等。

后来人们把实物信号移位到图画上，比如：北美印地安人用弓箭表示战争，用插羽毛的烟管表示和平，用人吸烟管表示讲和等等（见图 1-1）。这种从实物传信到实物画面传信的发展，成为



图 1-1 表示信息的图画

文字由来的一个重要源泉。

二、结绳记事

在中国《易经》上有“上古结绳而治，后世圣人易之以书契”的记载。结绳记事跟其它实物记事一样，只起到帮助记事的作用，跟语言不发生联系。结绳记事最为发达的时期，绳子分出粗细不同，颜色各有区别，疙瘩分出大小和不同结法。问题解决了，疙瘩解开了。发展下去，诸如刻痕、串珠、烽火、击鼓等等。

三、烽火狼烟

烽火传信，流传下来为军事所用，用作设防地区的报警信号。相传古代烽火用狼粪，狼烟有直而聚，风吹不散不斜的性能。唐朝李商隐《李義山诗集》里有：“鸡塞谁生事，狼烟不暂停”的记载。明代钱允治有诗云：“野夫却喜狼烽息，醉饮将军宝纛旁”。

传说周幽王为博得褒姒一笑，多次谎报敌警，举烽火征兵取乐，由此失信于诸侯，后被犬戎杀于骊山下。这就是千金一笑失天下的故事。可见，人类社会信息联系是一件多么重要的事情。