

图书馆科技成果介绍

一等奖

汉字属性字典及汉字信息处理支撑软件

(北京图书馆) 朱 岩等

电子计算机产生在西方,它的汉化,即开发处理中国汉字信息的功能是近十年的事。但截止该软件研制成功之前,无论是引进的或国产的大、中、微型计算机,在解决象符合中国传统要求的汉字排序、字音转换、不同汉字编码转换等这些汉字信息处理的基本功能上,在软件上还是空白,而不解决这些问题,就难以在我国涉及汉字信息部门推广应用。因此,该汉字属性库及软件研制成功,并装在大、中、小微机上运行,填补了国际上电子计算机在汉字信息处理支撑软件方面的空白,该软件系统设计思想正确,实用结果表明具有很好的实用性,规范性和可移植性。并且提高工效显著、处理结果准确,软件功能具有中国特色。它的效益表现在:

1. 第一次使中国汉字在计算机中成为有序的集合,解决了中国汉字参加比较、计算的问题,因而使各行业普遍要求的诸如人名、团体名、产品名、题名、词语按照中国多种传统规范方法进行排序。实用证明,提高工效比手工排档快200倍以上。这些排序

功能还为计算机软件系统、数据库系统进一步汉化提供了依据。

2. 实现了汉字与其拼音的自动转换,用于中国ISDS中心的汉字罗马化转换,以及情报检索系统检索点拼音索引的自动生成,转换出的结果准确、可靠。

3. 解决了中国汉字标准码与日本、台湾、美国一些计算机编码的自动转换问题,为计算机用户同上述国家和地区交换信息提供可能条件。

由于编制汉字属性和软件时,考虑了国内各行业的使用要求,依据了中国历年来文字规范,因而软件系统具有通用性和规范性。用户反映为他们“节省了大量的开发工作量,而且具有权威性,符合实用要求”。它已被国家标准局指定为制订国家汉字属性的依据,从1986年开始,该软件已向馆外大、中、小微机移植,国外一些计算机公司表示在它成为国标后将积极采用。它的应用推广,将促进我国文字信息系统的建立,产生更大的社会效益。

汉字属性字典及汉字信息处理

支撑软件应用效果显著

〔本刊讯〕 据悉北京图书馆于1985年研制的汉字属性字典及汉字信息处理支撑软件,除在本馆中型机上的一些业务处理系统中应用以外,还移植到馆外许多拥有大、中、小、微型计算机系统上,一年多来已取得了

满意的使用效果。

第一,提高工效显著。如对国家标准汉字编码字符集(GB2312—80)中6763个汉字进行排序处理,部首法排序,在该馆70万次/秒的中低档机上所用主机时间为37秒,

汉语拼音排序法所用主机时间为58秒,如将上机准备工作和输出打印时间计算在内,比手工排序提高工效也在200倍以上,用计算机编排的这些字表,已被计算数据录入员检查汉字编码、编辑字典索引等工作采用,成为有效的工具。用于以人为键的职工档案排序、文献题名和责任者的书目排序,以及《日英汉科技大词典》26万个汉语词条的排序,都取得了同样令人满意的效果。

该软件用于解决国际连续出版物数据系统 (ISDS) 中国国家中心的中文书目数据向拼音罗马化形式自动转换,取代了手工注音、数据录入、反复校验的繁琐操作,效率高,转换结果准确无误。用于中文机读目录中汉语拼音字段的自动生成,也取得了同样的效果。字音转换在批处理时是按照汉字的常用音转换的;虽然有些汉字有多音的特点,有时需要人工加以修正,但根据使用情况统计,这种需要修正的字仅为万分之四左右。而联机处理则可按指定音转换。

第二,处理结果准确、规范。无论是排

序结果或字音转换结果,经用户反复验证均正确无误。而且这些处理结果符合图书馆情报界多年形成的规范,在利用汉字属性元素的深度上又超过了手工处理的规范,处理结果更为精细。

第三,该软件的编码转换、排序处理、字音转换等主要功能对许多行业都具有通用性,为计算机系统的进一步汉化,为建立更多的实用汉字信息处理系统打下了基础。

第四,可移植性好。由于提供给用户使用的软件是数值型的,因而对大、中、小和微机不论机型如何,都可移植。可以整体移植,也可以根据需要部分移植。目前,除M系列机外已移植到IBM、王安、UNIVAC等机型。该软件目前占外存空间3.5兆字节,内存需128K,可在各种操作系统下运行。所用代码为EBCDIC或ASCII。

汇集汉字各主要属性的属性字典即将由书目文献出版社出版。

该软件的一些处理结果样张见封二。

汉字基本属性

【部首及序号】 禾 160 【部首】 禾 5 【部首外起末笔形】 11134 【异体字】	稷 QIN 5 【区位号】 3962 【国际码】 475E 【电报码】 1392 【台码】 21362E 【笔画数】 13 【四角码】 30147 【起末笔形】 454425114554 【部首外起末笔形】 4425114554 【异体字】
琴 QIN 2 【区位号】 3957 【国际码】 4759 【电报码】 3839 【台码】 21484D 【笔画数】 12 【四角码】 11207 【起末笔形】 112111213445 【部首及序号】 王 084 【部首】 王 4 【部首外起末笔形】 11213445 【异体字】	沁 QIN 4 【区位号】 3961 【国际码】 475F 【电报码】 3084 【台码】 21466B 【笔画数】 7 【四角码】 33100 【起末笔形】 4444544 【部首及序号】 氵 145 【部首】 氵 3 【部首外起末笔形】 4544 【异体字】
勤 QIN 2 【区位号】 3958 【国际码】 475A 【电报码】 0530 【台码】 213439 【笔画数】 13 【四角码】 44127 【起末笔形】 1221251121153 【部首及序号】 力 037 【部首】 力 2 【部首外起末笔形】 12212511211 【异体字】	青 QING 1 【区位号】 3964 【国际码】 4760 【电报码】 7235 【台码】 215F70 【笔画数】 8 【四角码】 50227 【起末笔形】 11212511 【部首及序号】 青 242 【部首】 青 8 【部首外起末笔形】 【异体字】
芹 QIN 2 【区位号】 3959 【国际码】 475B 【电报码】 5367 【台码】 21543E 【笔画数】 7 【四角码】 44221 【起末笔形】 1223512 【部首及序号】 艹 211 【部首】 艹 3 【部首外起末笔形】 5312 【异体字】	轻 QING 1 【区位号】 3965 【国际码】 4761 【电报码】 65 5 【台码】 215B43 【笔画数】 9 【四角码】 47512 【起末笔形】 152154121 【部首及序号】 车 076 【部首】 车 4 【部首外起末笔形】 54121 【异体字】
摘 QIN 2 【区位号】 3960 【国际码】 475C 【电报码】 2350 【台码】 214223 【笔画数】 15 【四角码】 58027 【起末笔形】 15134413432554 【部首及序号】 扌 115 【部首】 扌 3 【部首外起末笔形】 344134522554 【异体字】	甄 QING 1 【区位号】 3566 【国际码】 4762 【电报码】 841 【台码】 21445B 【笔画数】 9 【四角码】 80117 【起末笔形】 311554121 【部首及序号】 气 117 【部首】 气 4 【部首外起末笔形】 54121 【异体字】
离 QIN 2 【区位号】 3961 【国际码】 475D 【电报码】 4419 【台码】 214F21 【笔画数】 12 【四角码】 80227 【起末笔形】 344134522554 【部首及序号】 人 020 【部首】 人 2 【部首外起末笔形】 4134522554 【异体字】	倾 QING 1 【区位号】 3167 【国际码】 4763

按汉字拼音排序法对人名排序

按汉字拼音排序法对人名排序

姓 名	编 号	性 别	部 门 作 业	科 研 名 称
艾 芳	BA05		图书馆	金石
李 芳	BA05		图书馆	金石
安 芳	HA13	女	中文	中文
安 芳	HA13	女	中文	中文
安 芳	LA09		图书馆	金石
安 芳	P 15		图书馆	金石
安 芳	b 02	女	图书馆	金石
安 芳	3F05		图书馆	金石
安 芳	DE12		图书馆	金石
安 芳	V604	女	图书馆	金石
安 芳	S625	女	图书馆	金石
安 芳	U 26		图书馆	金石
安 芳	V 03		图书馆	金石
安 芳	b 04		图书馆	金石
安 芳	IC03		图书馆	金石
安 芳	NE08		图书馆	金石
安 芳	NA02		图书馆	金石
安 芳	VA01		图书馆	金石
安 芳	V 63		图书馆	金石
安 芳	VE14		图书馆	金石
安 芳	WE01	女	图书馆	金石
安 芳	E 02		图书馆	金石
安 芳	0A27		图书馆	金石
安 芳	Q 08	女	图书馆	金石
安 芳	JA17		图书馆	金石
安 芳	JE44		图书馆	金石
安 芳	Y 06	女	图书馆	金石
安 芳	Y 56	女	图书馆	金石
安 芳	G 14		图书馆	金石
安 芳	ID05		图书馆	金石
安 芳	RD05		图书馆	金石
安 芳	OD10		图书馆	金石
安 芳	V 45		图书馆	金石
安 芳	JC07		图书馆	金石
安 芳	Q 06	女	图书馆	金石
安 芳	HA20		图书馆	金石
安 芳	Q 01		图书馆	金石
安 芳	LD13	女	图书馆	金石
安 芳	LD09	女	图书馆	金石
安 芳	JA12		图书馆	金石
安 芳	HC17		图书馆	金石
安 芳	D 06	女	图书馆	金石
安 芳	DB15	女	图书馆	金石
安 芳	DB35		图书馆	金石
安 芳	RB08		图书馆	金石
安 芳	DA17	女	图书馆	金石
安 芳	SE08		图书馆	金石
安 芳	IA04	女	图书馆	金石
安 芳	V 27		图书馆	金石
安 芳	NA27		图书馆	金石

新闻稿

北京图书馆馆长任继愈同志已于1987年5月7日到职视事。

≡ 汉字属性字典及汉字信息处理支撑软件系统处理结果样张 ≡

Page No. 1
01/01/80

按汉语拼音排序的结果

题名和责任人
情报学/曹保明编
采油地质工程/金毅孙主编
超导电子学/〈英〉廷哈母(T. H. M. M.)著;邵惠民等译
痴情/王默风著

Page No. 1
01/01/80

按部首排序的结果

题名和责任人
中华人民共和国人事制度概要/曹志主编
中华人民共和国邮票目录/中国集邮出版社编
中国珍稀野生动物/中国野生动物保护协会主编
中国植物志第三十七卷/中国科学院中国植物志编委会编
易经正传/〈清〉冯武撰著

Page No. 1
01/01/80

按笔画排序的结果

题名和责任人
工业心理学/〈苏〉齐斯著;俞祖元、姚锡家译
工业锅炉产品样本/机械工业部编
小型水利水电工程设计图例:倒虹吸管分册/湖南省水利水电勘测设计院编制
马列文论研究第七集/全国马列文艺论著研究会主编
少女漂泊记/赵世菊著

Page No. 1
01/01/80

按四角号码排序的结果

题名和责任人
痴情/王默风著
新定三礼图/〈宋〉聂崇义集注
工业心理学/〈苏〉齐斯著;俞祖元、姚锡家译
工业锅炉产品样本/机械工业部编

中文15DS数据

aaa 00	1 0 0 0 0 4 0 2 *	222	@ a 自然 @ 科学史 @ 研究六
001	2 6 0 3 2 9 c 1 9 7 9 9 9 9 9 c h n q 2 2 p	245 0	@ a 自然 @ 科学史 @ 研究六
008	e c h i *	245 1	@ a Studies in 1
022	@ a 1 0 0 0 - 0 4 0 2 *	260	N a t u r a l S c i e n c e
052	@ b 0 2 0 *	550	@ a 北京 @ 科学 @ 出版社六
222	@ a 图书 @ 学 @ 通 @ 六		@ a 中国 @ 科学 @ 自然 @ 科学
245 0	@ a 图书 @ 学 @ 通 @ 六	aaa 00	
260	@ a 北京 @ b 文 @ 行 @ 出版 @ 社六	001	1 0 0 0 0 1 4 3 *
550	@ a 中国 @ 图书 @ 学 @ 会 @ 六 *	008	2 6 0 3 2 9 c 1 9 7 9 9 9

罗马化5角代码

aaa 00	1 0 0 0 0 1 9 4 *	aaa 00	
001	2 6 0 3 2 9 c 1 9	001 10	000402H
008	c	008 66	0828c19799999chnq22p
022	@ a 1 0 0 0 - 0 1	022	@ a 1000-0402H
052	@ b 9 3 1 *	082	@ b 020H
222	@ a 文 @ 物 @ 天 @ 地 @ 六	222	@ a Tushuguanxue tongxunH
245 0	@ a 文 @ 物 @ 天 @ 地 @ 六	245 0	@ a Tushuguanxue tongxunH
260	@ a 北京 @ b 文 @ 行 @ 出版 @ 社六	260	@ a Beijingabwenwu ChubansheH
	@ a 中国 @ 图书 @ 学 @ 会 @ 六 *	550	@ a Zhongguo Tushuguan XuehuiH
aaa 00	1 0 0 0 0 1 5 1 *	aaa 00	
001	2 6 0 3 2 9 c 1 9	001 10	000194H
008	c	008 66	0828c19819999chnb22p
022	@ a 1 0 0 0 - 0 1	072	@ a 1000-0194H
052	@ b 9 3 1 *		
222	@ a 文 @ 物 @ 天 @ 地 @ 六		
245 0	@ a 文 @ 物 @ 天 @ 地 @ 六		
260	@ a 北京 @ b 文 @ 行 @ 出版 @ 社六		
779 2	@ t 研 @ 究 @ 六 *		