

基于国家通用盲文标调规则的汉盲转换系统^①



胡庆玲¹, 林 栋², 陈楷闻², 吴新丽², 梁宏博³, 杨文珍³

¹浙江理工大学 信息学院, 杭州 310018)

²浙江理工大学 机械与自动控制学院, 杭州 310018)

³之江实验室 类人感知研究中心, 杭州 311121)

通信作者: 杨文珍, E-mail: ywz@zhejianglab.edu.cn

摘 要: 信息时代推进盲文数字化, 关乎我国广大盲人文化素质的提高和生活水平的改善. 本文实现了一种基于国家通用盲文标调规则的汉盲转换系统, 能够快速生成海量符合国家通用盲文方案的数字化资源, 满足视障人士无障碍获取信息的需求. 此系统按通用盲文规则处理汉语文本, 将其转换为符合标调规则、简写规则的盲文结果. 测试结果表明, 此系统可以准确处理标调规则、简写规则, 可得到准确的符合国家通用盲文方案的盲文数字化结果. 声调省写覆盖率、韵母简写覆盖率和篇幅增加量均与国家通用盲文方案的理论值相当, 能够快速处理长篇语料文件, 程序执行效率高, 具有实用价值, 可以用于推广国家通用盲文, 促进我国盲文数字化无障碍建设.

关键词: 国家通用盲文; 汉盲转换系统; 盲文 ASCII 码; 盲文数字化; 盲人

引用格式: 胡庆玲, 林栋, 陈楷闻, 吴新丽, 梁宏博, 杨文珍. 基于国家通用盲文标调规则的汉盲转换系统. 计算机系统应用, 2022, 31(12): 59-68. <http://www.c-s-a.org.cn/1003-3254/8877.html>

Chinese-braille Conversion System Based on National General Braille Tone Rules

HU Qing-Ling¹, LIN Dong², CHEN Kai-Wen², WU Xin-Li², LIANG Hong-Bo³, YANG Wen-Zhen³

¹(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

²(Faculty of Mechanical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China)

³(Humanoid Perception Research Center, Zhejiang Lab, Hangzhou 311121, China)


Abstract: The promotion of digital braille in the information age can help to improve the cultural quality and living standards of blind people in China. This study implements a Chinese-braille conversion system based on the national general braille (NGB) tone rules, which can quickly generate a large number of digital resources in line with the NGB rules and make visually impaired people obtain information without barriers. This system processes Chinese text according to the NGB tone rules and converts it into braille that conforms to tone rules and abbreviation rules. The test results show that the system can accurately process the tone rules and abbreviation rules and obtain accurate digital braille that is in line with the NGB tone rules. In addition, the coverage rate of tonal and final abbreviations and the increase in length are all comparable to the theoretical values of the NGB tone rules. The system can quickly process long-form corpus files and execute programs efficiently. Furthermore, it has practical value and can be used to promote the NGB and promote the barrier-free construction of digital braille in China.

Key words: national general braille (NGB); Chinese-braille conversion system; braille ASCII code; braille digitization; the blind

① 基金项目: 国家重点研发计划 (2021YFF0600203); 浙江省自然科学基金 (LY20F020019, LQ19F020012); 浙江省基础公益研究计划 (LGF19E050005); 之江实验室科研项目 (2022MG0AC04, 2021MC0PI01)

收稿时间: 2022-04-25; 修改时间: 2022-05-22; 采用时间: 2022-06-13; csa 在线出版时间: 2022-08-24

文字是人类记录思想和传递信息的重要载体,也是人类文明传承和发展的重要工具.我国是世界上盲人最多的国家,约有1700多万视障群体,占世界盲人总数的18%^[1].盲文是专为盲人设计、靠触觉感知的文字,是盲人学习文化知识的主要媒介.在信息时代,新知识呈现爆发式增长,且快速迭代更新.为此,提高我国盲文信息化技术水平,丰富盲文数字化资源,努力为盲人创造无障碍的学习环境,具有重要现实意义^[2].

长期以来,我国使用的盲文是以汉语拼音为基础的现行盲文.现行盲文是表音文字,由声母、韵母和声调组成,需三方盲符.比如,汉字“国”的拼音为“guo2”,声母“g”、韵母“uo”和声调“2”的盲文点序分别为110110、101010、010000,盲符是.

为了节省传统盲文书籍的篇幅,现行盲文通常以“一般不标调,需要时才标调”为指导,无规则地任意省略声调,以减少盲符,造成了许多“一词多形、一形多词”的情况^[3],既不利于盲人阅读,也不利于盲文数字化.在摸读现行盲文时,盲人常常需要通过上下文猜测声调,才能准确理解语义,严重影响了学习效率.计算机也难以准确处理任意省略声调的现行盲文,难以高质量地实现“汉字-盲文”的转换^[4,5].

为了克服现行盲文任意省略声调造成的阅读障碍和盲文数字化障碍,北京联合大学、教育部语言文字应用研究所、北京语言大学、中国盲文出版社等单位的专家进行跨学科的科研攻关,形成了《国家通用盲文方案》^[6],经国家语言文字工作委员会语言文字规范标准审定委员会审定,成为国家语言文字规范,并于2018年7月1日起实施.国家通用盲文以字字标调为基础,利用零标记原理,按声母省写一个声调,标出3个声调,完善了现行盲文的标调规则,规范了声调符号用法,可以更准确地表达汉字的声、韵和调,实现了盲人学习由“猜”到“读”的转变,不仅丰富了我国盲文表音文字的内涵,也为“汉盲转换”的计算机自然语言处理奠定了基础^[7].

汉盲转换本质上是通过合适的程序结构和算法流程,按照一定的规则,快速且准确地完成普通文字字符和盲文ASCII码表的匹配,输出盲文点序,实现盲文数字化^[8-10].汉盲转换系统作为计算机自然语言处理研究领域的典型应用实例,除了涉及字符串分割、中文分词、汉拼转化等通用基础算法外^[11-13],还要处理许多盲文拼音规则,比如,声韵调简写规则、声韵调省写

规则、特殊字规则等^[9,11,14],才能获得准确的转换结果.

黄河燕等^[15]根据盲文分词连写要求建立了多条基于SC文法的特定规则,在字典匹配的分词算法基础上,对分词结果进行标记,并根据规则库进行连写处理.庄丽等^[9]对盲文分词连写规则库进行改写和补充,同时利用统计模型对分词结果进一步排歧,但算法复杂度高.杨潮等^[16]从语料中提取出现过的连写字词组合,建立连写示范库,对文本进行基于连写组合库的切分处理.吕先超^[14]基于规则分词和连写示范库相结合的汉盲转换模型开发出一款汉盲自动转换软件SunBraille,主要面向Web和服务端的应用.蔡佳等^[17]提出了一种基于汉盲对照语料库和深度学习的汉盲自动转换方法,从汉字和盲文文本中自动匹配抽取语料的方法构建了大规模的汉盲对照语料库用来训练LSTM模型,从而实现高准确的盲文分词,并建立汉字标调语料库,利用bi-gram语言模型对汉字进行标调,证明了基于汉盲对照语料库的模型效果均优于采用纯盲文语料库的方法.徐豪杰^[18]设计实现了一款带有未登录词自动识别功能的汉语盲文翻译系统,设计实现了基于改进PMI和最小邻接熵结合策略的未登录词识别算法^[19],将识别得到的未登录词改进Jieba分词系统,构建了以中文分词连写、字音转换、拼音-ASCII码转换、盲文点序生成等模块构成的汉盲翻译系统,该系统具有改善分词连写效果的功能,并实现了现行盲文全带调和不带调的转换功能,但未实现国家通用盲文的转换.

为了更好地推广国家通用盲文,促进我国盲文数字化无障碍建设,本文在带有未登录词自动识别功能的汉语盲文翻译系统^[18]的基础上,构建了国家通用盲文标调与简写规则处理模块,集成实现了基于国家通用盲文标调规则的汉盲转换系统.系统可以转换得到符合《国家通用盲文方案》的盲文结果,可以便捷、快速生成海量通用盲文数字化资源,满足信息时代盲人学习知识的迫切需求,为视障弱势群体的语言文字权益和公平教育提供技术保障.

1 系统设计与实现

汉盲转换系统的功能是把普通的汉字、字母或符号等字符转化为相应的盲文ASCII码或盲文点序,以盲文电子文件或直接输出给盲文点显器等形式,供用户使用.通常,它的输入是普通文字的字符串或文件,

输出是盲文 ASCII 码或“0”和“1”组合的盲文点序。比如,输入汉字字符串为“中国”,输出省略声调的现行盲文的盲文 ASCII 码为“/4G0”,盲文点序为“001100010011110110101010”。

本文建立了汉盲转换系统的总体流程框架,开发了高效的算法和程序结构,能把 UTF-8 编码的普通文字字符串或文件,转换为符合国家通用盲文标调规则的盲文点序。

1.1 总体流程框架与算法结构

本系统在 Spyder 4.1.5 集成开发环境下采用 Python 3 语言编写,调用了 Jieba 词库、Pypinyin 拼音库和 EasyGui 库等。系统的总体流程框架如图 1 所示。

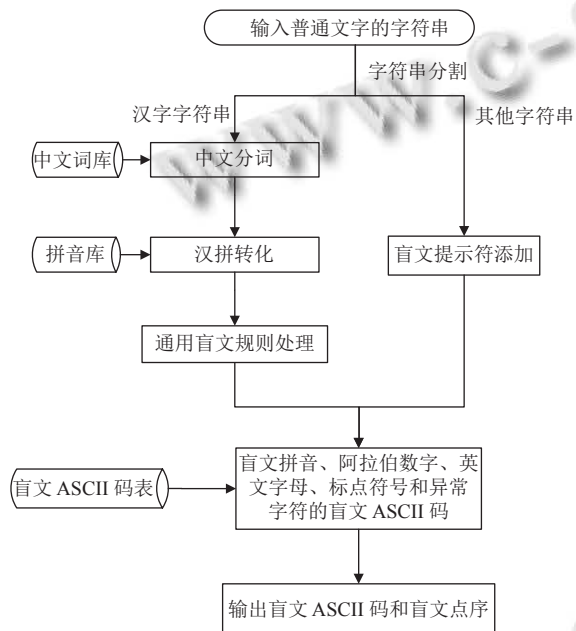


图 1 系统总体流程框架

输入普通文字的字符串后,系统首先将普通文字字符串按 Unicode 码标准分为两类字符串,一类是汉字字符串,另一类是其他字符串。对于其他字符串,系统经过盲文提示符规则处理后,直接查找对应的盲文 ASCII 码表,得到盲文点序;对于汉字字符串,系统经过中文分词、汉拼转化处理,并执行国家通用盲文规则处理程序,才能得到通用盲文拼音,再查找拼音盲文 ASCII 码,得到盲文点序。最后,系统按照普通文字字符串的先后次序,合并输出这些盲文点序,形成通用盲文的数字化资源,供盲人使用。

为了实现此系统的总体流程,本文建立了此系统的整体程序结构,如图 2 所示。

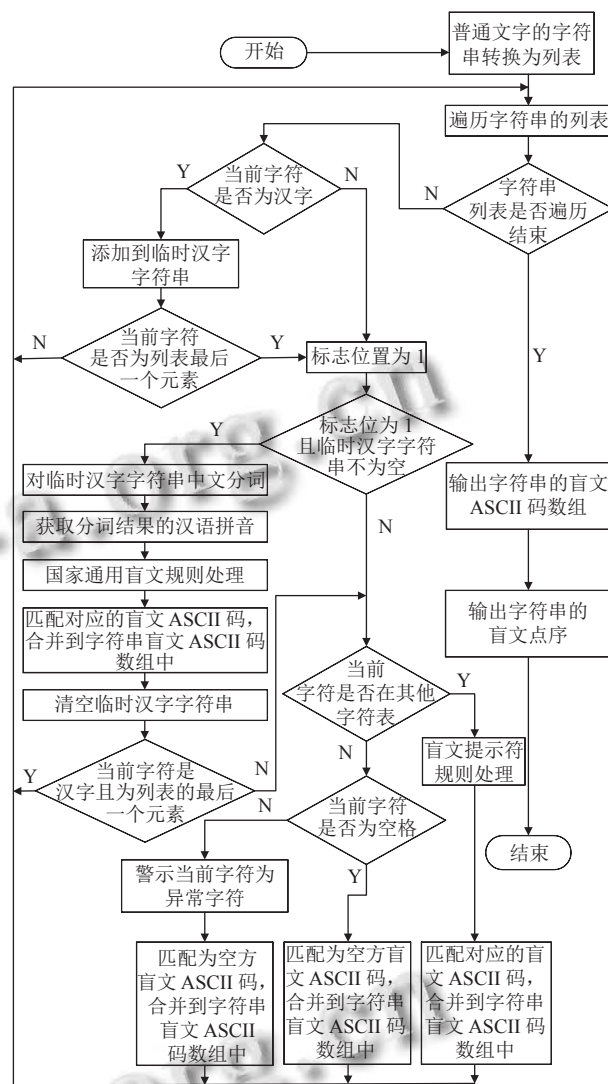


图 2 系统整体程序框图

系统将普通文字的字符串转换为字符串列表后,依次向后遍历列表,并判断当前字符的类型。在遍历过程中,系统会遇到两种情况:1)连续遍历到的当前字符都为汉字,就将当前字符依次添加到临时汉字字符串中;若字符串列表中的所有元素均为汉字,遍历完成后,将标志位置为 1。2)遍历到的当前字符不是汉字,将标志位置为 1,此时,若临时汉字字符串不为空,先经过分词和汉拼转化得到盲文拼音,而后调用国家通用盲文规则处理模块,得到符合国家通用盲文规则的盲文拼音,最后匹配相应的盲文 ASCII 码,若临时汉字字符串为空,系统进行其他字符(阿拉伯数字、英文字母、标点符号、空格、异常字符)处理,得到相应的盲文 ASCII 码。

1.2 分词与汉拼转换

基于 Jieba 分词库和 Pypinyin 汉拼转换库, 本文实现了汉字字符串中文分词和汉拼转化, 先得到汉字字符串的分词结果而后转换为汉语拼音, 经过分割后再修正, 获取到盲文拼音. 图 3 为汉字字符串转化为盲文拼音的程序流程.

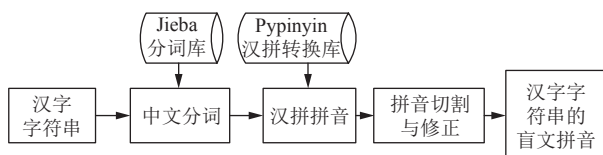


图 3 汉字字符串的盲文拼音

为了实现通用盲文的标调规则, 本文依据音节中必有韵母的特点, 定位音节字符串中韵母的第一个字母, 准确切分出一个音节字符串中的声母、韵母和声调. 声调符号用数字“1、2、3、4、5”来表示, 分别是代表“阴平、阳平、上声、去声、轻声”, 放置在音节字符串的末端. 以音节字符串“zhong1”为例, 定位字符“o”, 切分出“o”前部字符串“zh”为声母, 最后一个字符“1”为声调, 剩余字符串“ong”为韵母.

由于盲文是靠触觉感知的文字, 我国的通用盲文拼音体系与明眼人使用的汉语拼音体系存在着较大的差异. 明眼人使用的汉语拼音体系声母有 23 个, 韵母有 24 个, 然而, 盲文拼音体系的声母有 21 个, 韵母有 36 个. 因此, 转换系统不能从汉语拼音直接得到盲文拼音而是需要经过修正处理.

拼音修正处理主要包括: 1) “y、w”声母修正为“i、u”并与原韵母合并; 2) “iou、uei、uen”修正为“iu、ui、un”; 3) “yo、eg”的修正为“io、en”; 4) 声母“j、q、x”和“u”开头韵母组合, “u”修正为“v”; 5) 声母“z、c、s、zh、ch、sh、r”和韵母“i”组合省去“i”.

1.3 通用盲文标调规则与简写规则处理

基于上述盲文拼音结果, 依据《国家通用盲文方案》, 本文设计合适的算法程序结构, 处理通用盲文的标调规则和简写规则, 自动得到汉字的通用盲文拼音. 标调规则包括两个部分, 分别为基本规则和声调省写规则.

标调规则:

基本规则 (1): 字字标调.

基本规则 (2): 变调标本调.

声调省写规则 (1): 声母为 f 的音节, 省写阴平符号.

声调省写规则 (2): 声母为 p、m、t、n、h、q、

ch、r、c 的音节, 省写阳平符号; 音节 tou2 的声调符号不省写.

声调省写规则 (3): 声母为 b、d、l、g、k、j、x、zh、sh、z、s 的音节, 省写去声符号; 音节 le4、zi4 的声调符号不省写.

声调省写规则 (4): 韵母自成音节, 省写去声符号, 第 (5) 和 (6) 条规定的除外.

声调省写规则 (5): 音节 yi1、er2、wo3、ye3、you3 的声调符号省写; 音节 yi4、er4、wo4、ye4、you4 的声调符号不省写.

声调省写规则 (6): 音节 o1、o2、o3、o4 的声调符号省写; 音节 e1、e2、e3、e4 的声调符号不省写.

声调省写规则 (7): 声母自成音节后面连写韵母自成音节时, 声母自成音节不适用省写规则 (2) 和 (3).

声调省写规则 (8): 需要声调符号提供点位参照或者区分音义时, 声调符号应不省写.

简写规则:

(1) “的、么、你、他、她、它”简写. “的”简写为 , “么”简写为 , “你”简写为 , “他”简写为 , “她”简写为 , “它”简写为 .

(2) “的、么、你、他、它”后面连写韵母自成音节时不能简写.

通过分析标调规则与简写规则, 本文建立了通用盲文标调与简写规则处理模块的整体程序结构, 实现除声调省写规则 (8) 外的其他规则, 如图 4 所示. 在执行声调省写前 6 个规则后, 声调省写规则 (7) 和简写规则 (2) 会自动满足. 本文不执行声调省写规则 (8), 因为此规则还没有一个刚性的标准, 恰当执行此规则会便于摸读, 所以没有执行也不会有实质性的摸读影响^[5].

执行不同的声调省写规则和简写规则, 程序需要判断声母和韵母的逻辑关系. 由于执行声调省写规则 (7) 与简写规则 (1)、(2), 程序需要结合后一个音节进行逻辑判断, 为了便于音节遍历, 程序在拼音列表末端增加一个空格元素.

当输入汉字字符串的盲文拼音后, 程序从第一个元素开始遍历列表. 若当前音节的声母为空, 执行声调省写前 6 条规则和简写规则; 若当前音节的声母不为空, 又不是声母自成音节且不符合简写规则, 执行声调省写的 (1)–(3) 规则和简写规则; 若当前音节是声母自

成音节或对应的汉字是“的, 么, 你, 他, 她, 它”之一, 且韵母自成音节, 则不执行声调省写规则, 否则执行声调省写 (2)、(3) 规则和简写规则. 最后, 程序匹配当前音

节的声母、韵母和声调的盲文 ASCII 码并按次序合并, 得到输入汉字字符串的通用盲文的盲文 ASCII 码和盲文点序.

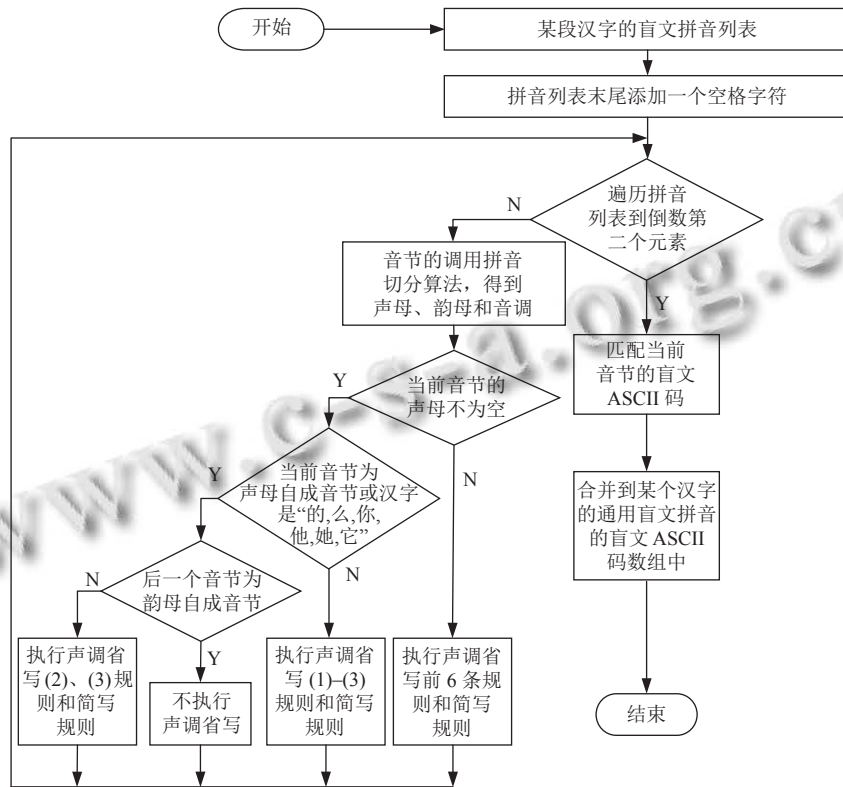


图 4 通用盲文标调规则的整体程序框图

简写规则 (1) 中提到的“它”字在声母“t”的点序前加了点序用于区分同音的“ta1”, 所以执行“它”字简写需要特殊处理. 程序将“它”的声母置为“tt”, 并在声母盲文 ASCII 码对照表中建立“tt”与盲文 ASCII 码“@t”的对应关系.

图 5-图 10 为声调省写规则 (1)-(6) 实现流程图, 图 11 为简写规则 (1) 实现流程图.

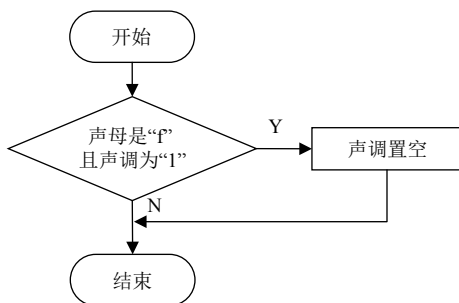


图 5 声调省写规则 (1) 流程图

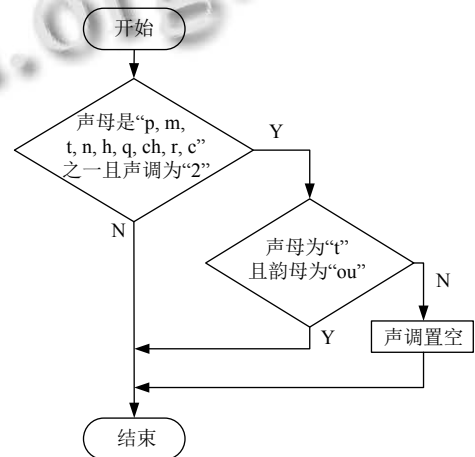


图 6 声调省写规则 (2) 流程图

1.4 盲文点序生成

将文本转换为盲文阅读设备可读取的有效数据是汉盲转换的最终环节. 依据《国家通用盲文方案》, 本

文在转换系统中构建了盲文拼音符号(声母、韵母、声调)、阿拉伯数字、英文字母和标点符号的盲文 ASCII 码对照表,利用对照表分别映射输出盲文 ASCII 码和盲文点序,如图 12 所示.每个盲文 ASCII 码的盲文点序是固定不变的,以中文标点符号“!”为例,盲文 ASCII 码是“;1”,对应的盲文点序为“0000 11010000”.拼音盲文 ASCII 码对照表包含的通用盲文拼音体系元素即声母、韵母、声调,均是汉语拼音体系修正后的形式.

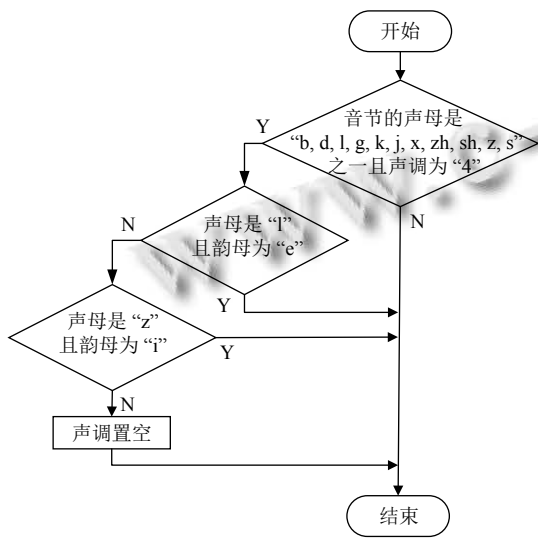


图 7 声调省写规则 (3) 流程图

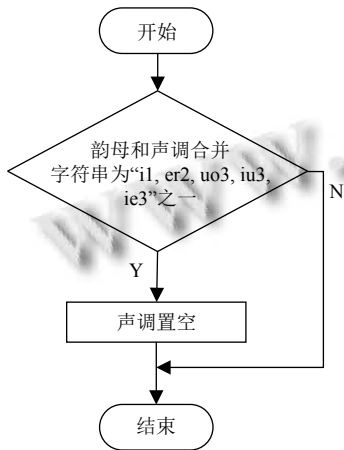


图 8 声调省写规则 (4) 流程图

对于异常字符,例如韩文、日文等字符,此系统均按空方处理,对于盲文 ASCII 码为空格,盲文点序为“000000”.

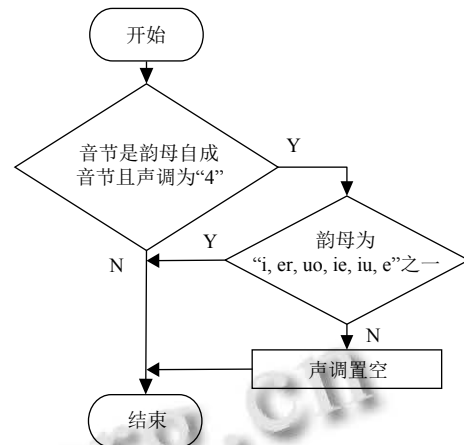


图 9 声调省写规则 (5) 流程图

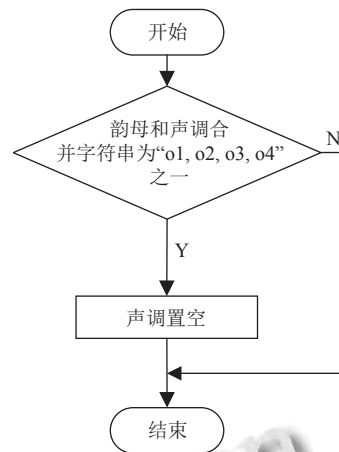


图 10 声调省写规则 (6) 流程图

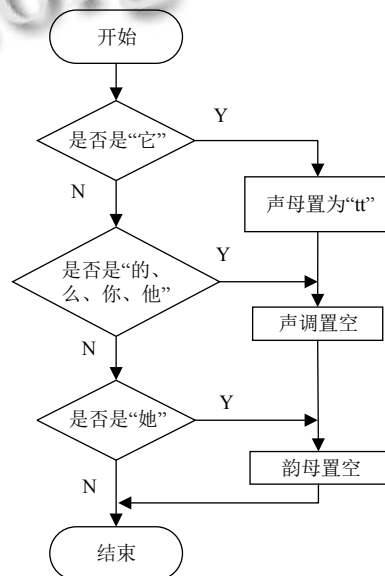


图 11 简写规则流程图

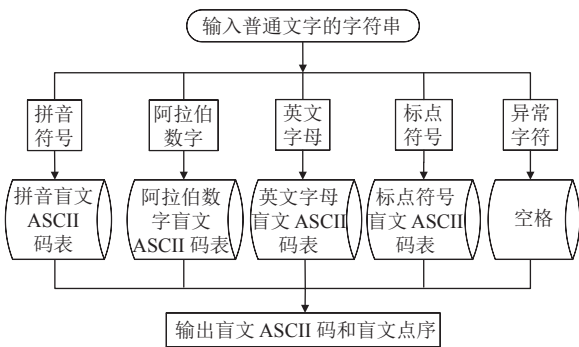


图 12 普通文字转为盲文 ASCII 码和盲文点序

本文以“Linux 诞生于 1991 年。”字符串为例. 此字符串含有汉字、阿拉伯数字、英文字母和标点符号, 其中, 汉字经过中文分词、汉拼转化、通用盲文标调规则等处理得到通用盲文 ASCII 码结果和点序结果, 程序执行过程如图 13 所示.

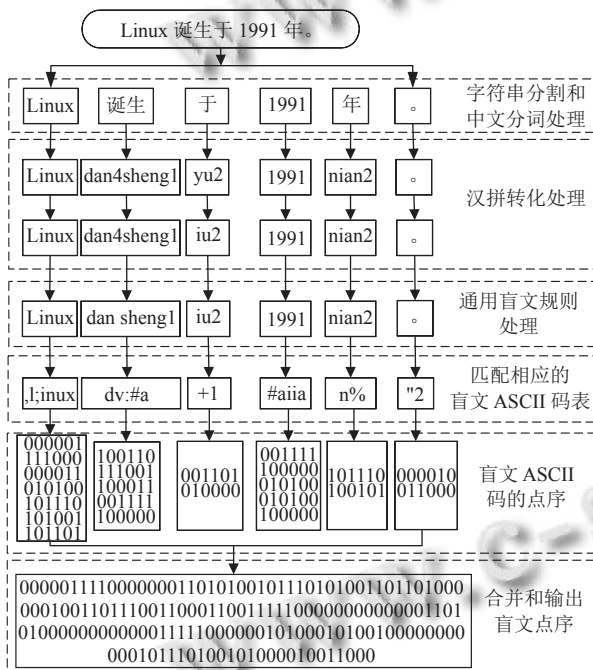


图 13 转换系统各阶段处理流程示意图

在程序执行过程中, 系统可以得到此字符串的各个运行结果, 包括此字符串的分割结果、中文分词结果、汉拼转化结果、拼音修正结果、通用盲文规则处理结果、盲文 ASCII 码结果和盲文点序结果. 此系统在字符串分割、中文分词、汉拼转化、拼音修正、拼音切割等模块基础上, 集成了本文所提出的通用盲文标调规则处理模块, 能够处理汉字、阿拉伯数字、英文字母和标点符号等字符, 可靠输出通用盲文点序.

2 系统测试与分析

本文将此系统转换结果应用于自主研发的智能盲文学习机中, 可以让盲人无障碍地摸读通用盲文, 如图 14 所示. 此智能盲文学习机能够同步呈现“视、听、触”刺激, 方便盲人学习知识^[20].

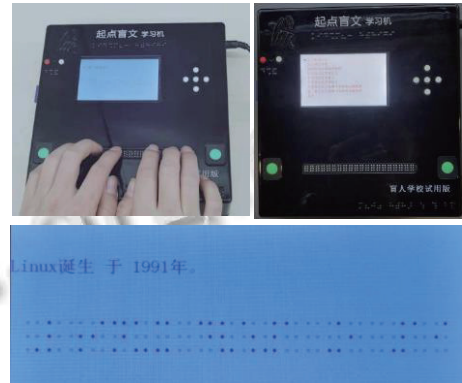


图 14 盲文学习机和盲文点序

为了验证此系统的性能, 本文开展系统标调与简写规则转换测试、系统转换准确率测试、篇幅增加量测试和转换速度的测试.

2.1 通用盲文标调与简写规则的盲文转换测试

本文选取特定的需要按通用盲文规则特殊处理的汉字或词组, 测试此系统执行通用盲文规则的正确性, 测试结果如表 1 所示. 实验结果表明, 各个测试结果的盲文点序正确, 符合通用盲文的规则, 证明本文所构建的通用盲文处理模块可在系统中准确实现《国家通用盲文方案》的相关规则.

2.2 系统转换准确率的测试

为了进一步验证集成通用规则处理模块的汉盲转换系统的可行性. 本文基于《人民日报》3 个月约 864 万字中文语料, 构建了规模较大的汉盲对照语料库. 中文语料为 2022 年 1-3 月数据, 大小为 24.1 M. 该语料包含中文、英文、数字、标点符号等多种字符类型, 统计如表 2 所示.

盲文语料是将上述中文语料以段为单位切分, 由阳光专业盲文编辑排版系统 (8.1.0 版) 转换, 并由盲文专家校对得到中文语料的国家通用盲文 ASCII 码.

将中文语料在本系统下转换得到的盲文 ASCII 码与汉盲对照语料库的盲文 ASCII 码比较得到汉盲转换准确率如表 3.

汉盲转换准确率的计算方法:

$$A = \frac{T - M}{T} \times 100\% \quad (1)$$

其中, A 表示准确率, T 表示总盲文方数, M 表示误标盲文方数。

表1 声调省写规则的测试结果

规则	项目	结果1	结果2	结果3			
声调省写规则(1)	汉字	方法	开发	风暴			
	拼音	fang fa3	kai1 fa	feng bao			
	点序						
声调省写规则(2)	汉字	平民	抬头	童年			
	拼音	ping min	tai tou2	tong nian			
	点序						
声调省写规则(3)	汉字	必备	道路	快乐			
	拼音	bi bei	dao lu	kuai le4			
	点序						
声调省写规则(4)	汉字	爱国	遗忘	运用			
	拼音	ai guo2	i2 uang	vn iong			
	点序						
声调省写规则(5)	汉字	衣	儿	我	野	友	意
	拼音	i	er	uo	ie	iu	i4
	点序						
声调省写规则(6)	汉字	哦	喔	俄	饿	恶心	
	拼音	o	o	e2	e4	e3 xin1	
	点序						
声调省写规则(7)	汉字	慈爱	事业	质问			
	拼音	ci2 ai	shi4 ie4	zhi4 un			
	点序						
简写规则(1)	汉字	怎么	他们	你的			
	拼音	/	/	/			
	点序						
简写规则(2)	汉字	他用	怎么样	她爱人			
	拼音	/	/	/			
	点序						

表2 中文语料字符统计

字符类型	统计数量
中文	7545921
英文	7973
中文标点	868085
英文标点	10372
数字	216599
特殊符号	979

从表3的测试结果可得, 此系统的汉盲转换准确率较高, 转换结果97.42%符合《国家通用盲文方案》, 可以满足盲人学习知识的实际应用需求, 出现错误的主要原因集中在多音字的汉拼转换. 这里需强调, 对于不同的测试语料, 此系统的汉盲转换准确率会有所差异.

表3 汉盲转换准确率

测试样本	汉盲转换准确率
人民日报2022年1-3月数据	97.42%

2.3 省写简写覆盖率和篇幅增加量的测试

为了提高盲文的阅读和书写效率, 通用盲文在阅读准确前提下, 要尽可能地省写声调, 简写韵母, 以减少篇幅增加量. 为此, 本文开展此系统的声调省写覆盖率、韵母简写覆盖率和篇幅增加量的测试.

在本测试中, 声调省写个数不包括轻声, 汉字个数不包括标点符号. 声调省写覆盖率和韵母简写覆盖率的计算方法:

$$S = \frac{W}{N} \times 100\% \quad (2)$$

$$E = \frac{F}{N} \times 100\% \quad (3)$$

其中, S 表示声调省写覆盖率, E 表示韵母简写覆盖率, W 表示声调省写个数, F 表示韵母简写个数, N 表示汉字个数. 声调省写和韵母简写覆盖率的测试结果如表4所示.

本文的转换系统, 经过关闭通用规则处理模块, 并将 Pypinyin 拼音库修改为无音调模式, 可输出不带调现行盲文结果. 对照转换得到的现行盲文, 继续测试此系统生成的通用盲文的篇幅增加量, 测试结果如表5所示. 篇幅增加量的计算方法:

$$V = \frac{U - C}{C} \times 100\% \quad (4)$$

其中, V 表示篇幅增加量, U 表示通用盲文方位数, C 表示现行盲文方位数.

从表4和表5的测试结果可得, 此系统在本测试样本下声调省写覆盖率为41.19%, 韵母简写覆盖率为5.18%, 篇幅增加量为14.17%, 与国家通用盲文方案的理论值相一致^[5]. 测试结果从侧面证实本文所提出的通用盲文标调规则与简写规则处理模块具有一定的通用性, 证实此系统可以生成可信度很高的通用盲文数字化资源.

表4 声调省写和韵母简写覆盖率的测试结果

参数	数值
声调省写(个)	3 108 164
韵母简写(个)	438 418
汉字个数(个)	7 545 921
声调省写率(%)	41.19
韵母简写率(%)	5.81

表5 篇幅增加量的测试结果

参数	数值
现行盲文方数(个)	15 582 100
通用盲文方数(个)	17 790 083
篇幅增加量(%)	14.17

2.4 转换速度的测试

对于汉盲转换系统,程序运行效率是一个关键性能指标.本文在 Windows 10 操作系统 (Intel Core 2.60 GHz CPU, 8 GB 内存) 下,测试此系统的执行速度.

此系统能读入 UTF-8 编码的 TXT 文件,转换为通用盲文.转换速度如表6所示.

表6 转换速度的测试结果

参数	数值
字符数(个)	8 649 894
转换速度(字符/s)	23 268.04

测试结果表明,此系统的程序执行效率较高(>20 000 字符/s),远超过盲人摸读盲文速度(<300 字符/min),适合用于实时交互的汉盲转换使用场景.

3 总结

传统纸质盲文书籍存在制造成本高、更新周期长、知识承载量少、体积大且笨重等诸多缺点,已经不能满足信息时代广大视障群体的学习需求.盲文数字化已是必然发展趋势^[6,21,22].本文依据《国家通用盲文方案》,详细分析国家通用盲文方案的标调规则和简写规则,提出国家通用盲文标调与简写规则的算法流程和程序结构,构建规则处理模块,开发出高效的汉盲转换系统.

本文所开发的系统能够便捷、快速且准确地生成海量通用盲文资源,帮助广大盲人无障碍学习知识,推进我国盲文数字化建设,也可用于自动生成“通用盲文字表或词表”,解决了“通用盲文字表或词表”需要人工标注的难题,提高了通用盲文的校对效率.测试结果表明,此系统可以准确地执行标调与简写规则,汉盲转换

准确率较高,声调省写覆盖率、韵母简写覆盖率和篇幅增加量与国家通用盲文方案研究的理论值相当,能够快速处理长篇语料文件,程序执行效率高鲁棒性好,有较高的实用性.此系统还存在少数多音字和轻声字的标调错误,以及通用盲文分词连写规则未实现的问题,是未来工作需要改进的地方.

致谢

感谢徐豪杰、汪文妃、宣建强、赵维、刘畅、陶文韬、竺志超等人对此研究工作的贡献.

参考文献

- 1 毛志伟,傅悦,崔瑶.视障群体的信息无障碍应用现状分析.信息记录材料,2019,20(7):51-53.
- 2 杜磊.信息无障碍 让盲人平等感知世界.计算机与网络,2016,42(12):1. [doi: 10.3969/j.issn.1008-1739.2016.12.001]
- 3 钟经华,韩萍,肖航,等.现行盲文简写方案.中国特殊教育,2011,(2):34-37. [doi: 10.3969/j.issn.1007-3728.2011.02.007]
- 4 阎嘉,肖阳梅,任伟宁.通用盲文与现行盲文摸读比较研究.现代特殊教育,2017,(11):47-49. [doi: 10.3969/j.issn.1004-8014.2017.11.017]
- 5 钟经华.国家通用盲文方案研究.中国特殊教育,2018,(6):42-46,41. [doi: 10.3969/j.issn.1007-3728.2018.06.008]
- 6 中华人民共和国教育部,国家语言文字工作委员会,中国残疾人联合会.语言文字规范:国家通用盲文方案.北京:中国盲文出版社,2018:2-9.
- 7 钟经华.国家通用盲文方案特色解析.现代特殊教育,2018,(23):23-25. [doi: 10.3969/j.issn.1004-8014.2018.23.009]
- 8 Zhu X, Bao T. EasyBraille: A translation system for mandarin and braille. Proceedings of the 6th Joint Symposium on Computational Linguistics in China (JSCL-2001). Beijing: China Computer Federation. 2001. 326-331.
- 9 庄丽,包塔,朱小燕.盲人用计算机软件系统中的语音和自然语言处理技术.中文信息学报,2004,18(4):72-78. [doi: 10.3969/j.issn.1003-0077.2004.04.012]
- 10 Zhang X, Ortega-Sanchez C, Murray I. Text-to-braille translator in a chip. 2006 International Conference on Electrical and Computer Engineering. Dhaka: IEEE, 2006. 530-533.
- 11 Li ZP, Wang R, Zhang TC, et al. Intelligent braille conversion system of Chinese characters based on Markov model. IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference. Chengdu:

- IEEE, 2019. 1283–1287.
- 12 杨文珍, 徐豪杰, 汪文妃, 等. 基于逆向最大匹配分词算法的汉盲翻译系统. 计算机应用与软件, 2021, 38(10): 84–91. [doi: 10.3969/j.issn.1000-386x.2021.10.014]
- 13 Wang XD, Yang Y, Zhang JC, *et al.* Chinese to braille translation based on braille word segmentation using statistical model. Journal of Shanghai Jiaotong University (Science), 2017, 22(1): 82–86. [doi: 10.1007/s12204-017-1804-x]
- 14 吕先超. 视障汉语转换软件 SunBraille 的设计实现 [硕士学位论文]. 兰州: 兰州大学, 2016.
- 15 黄河燕, 陈肇雄, 黄静. 基于多知识分析的汉盲转换算法. 语言计算与基于内容的文本处理——全国第七届计算语言学联合学术会议论文集. 哈尔滨: 中国中文信息学会, 2003: 617–623.
- 16 杨潮, 车磊. 汉字盲文转换系统的设计. 北京印刷学院学报, 2011, 19(6): 36–38. [doi: 10.3969/j.issn.1004-8626.2011.06.013]
- 17 蔡佳, 王向东, 唐李真, 等. 基于汉盲对照语料库和深度学习的汉盲自动转换. 中文信息学报, 2019, 33(4): 60–67. [doi: 10.3969/j.issn.1003-0077.2019.04.007]
- 18 徐豪杰. 基于未登录词识别的汉盲翻译系统研究 [硕士学位论文]. 杭州: 浙江理工大学, 2019.
- 19 徐豪杰, 吴新丽, 杨文珍, 等. 基于改进 PMI 和最小邻接熵结合策略的未登录词识别. 计算机系统应用, 2020, 29(6): 181–188. [doi: 10.15888/j.cnki.csa.007434]
- 20 陈楷闻, 林栋, 钟泽栋, 等. 视听触同步刺激的数字化盲文学习方法. 计算机系统应用, 2021, 30(9): 262–270. [doi: 10.15888/j.cnki.csa.008100]
- 21 Wang XD, Zhong JH, Cai J, *et al.* CBConv: Service for automatic conversion of Chinese characters into braille with high accuracy. Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility. Pittsburgh: ACM, 2019. 566–568.
- 22 Kawabe H, Shimomura Y, Seto S. Braille translation system using neural machine translation technology II—Code conversion of Kana-Kanji mixed sentences. Proceedings of the 15th International Conference on Management Science and Engineering Management. Toledo: Springer, 2021. 417–426.

(校对责编: 孙君艳)