北京大学学报(自然科学版) 第 52 卷 第 1 期 2016 年 1 月
Acta Scientiarum Naturalium Universitatis Pekinensis, Vol. 52, No. 1 (Jan. 2016)
doi: 10.13209/j.0479-8023.2016.019

一种湘西民间苗文字形的动态生成方法及其实现途径

莫礼平[1,2,†] 周恺卿[3]

1. 吉首大学信息科学与工程学院, 吉首 416000; 2. 中南大学信息科学与工程学院, 长沙 410083; 3. 马来西亚理工大学计算学院, 士古来 81310; †通信作者, E-mail: zmx89@163.com

摘要 为了有效地解决湘西民间苗文字形的生成及描述问题, 提出一种字形的动态生成方法。该方法将苗文字形的生成过程表示为由苗文构件作为操作数、由构件位置关系决定运算符的组合运算表达式, 将 2~3 个构件进行不同的组合运算, 即可动态生成不同结构的苗文字形。利用操作系统自带的表意文字描述序列解释机制, 将构件组合运算表达式转换为表意文字描述序列, 即可实现该方法。测试结果表明, 根据该方法编写的映射脚本生成的湘西民间苗文字形可以满足实用要求。

关键词 民间苗文; 字形; 上下文无关文法; 表意文字描述序列

中图分类号 TP391

A Dynamical Glyph Generation Method of Xiangxi Folk Hmong Characters and Its Implementation Approach

MO Liping[1,2,†], ZHOU Kaiqing[3]

1. College of Information Science & Engineering, Ji Shou University, Jishou 416000; 2. Institute of Information Science & Engineering, Central South University, Changsha 410083; 3. Faculty of Computing, University Teknologi Malaysia, Skudai 81310; † Corresponding author, E-mail: zmx89@163.com

Abstract To effectively solve the glyph generation and glyph description problem, a dynamical glyph generation method of Xiangxi folk Hmong characters is proposed. According to this method, the glyph generation process can be described as a combination arithmetic expression. Hmong characters component acts as the operand, and the location relationship between the components decides the operator. Glyphs in different structure can be dynamically generated by combination of two or three components. Further, if combination arithmetic expression is converted to ideographic description sequence (IDS), the proposed method can be implemented with the help of the IDS explain mechanism of operation system. Test results illustrate that, the Xiangxi Hmong characters glyph, which generated by the mapping script based on the proposed method, can meet practical requirements.

Key words Folk Hmong characters; glyph; context-free grammar; ideographic description sequence (IDS)

湘西民间苗文由清朝末年以来一些民族知识分子创制, 主要包括板塘苗文、老寨苗文、古丈苗文三套方块苗文[1]。2011 年以来, 随着《湖南武陵山片区区域发展与扶贫攻坚规划(2011–2020)》的全面实施, 以世界自然遗产旅游区张家界和国家历史文化名城凤凰重点旅游景区为依托的民族文化旅游产业得到迅速发展, 非物质民族文化遗产数字化保护工作也逐渐受到重视。然而, 作为武陵山片区民族文化主要载体的湘西民间苗文, 其信息处理研究工作和成果鲜有报道。近年来, 莫礼平等[2–4]针对三套方块苗文, 在字库设计、文字输入等方面开展了一系列研究, 并取得阶段性成果。

国家自然科学基金(61462029)和中国计算机学会中文信息技术开放课题基金(CCF2014-02-01)资助
收稿日期: 2015–06–04; 修回日期: 2015–08–15; 网络出版日期: 2015–09–29

字形的生成及描述是湘西民间苗文在字层面信息处理技术研究的重要内容。本文提出一种基于构件组合运算的湘西民间苗文字形动态生成方法, 并结合 Unicode 提供的表意文字描述规范, 讨论该方法不占用编码区间的实现途径。

1 湘西民间苗文的造字原理及构字方式

1.1 造字原理

板塘苗文、老寨苗文和古丈苗文这三套湘西民间苗文均属于表意文字, 基本上都是合体字。创制者借鉴汉字的造字原理, 创造性地运用形声、会意、象形、假借等手段, 采用一字一音节的方法来标记一个语素或词。三套文字的结构类型大致分为 4 种[1]: 左右结构(最多)、上下结构(较多)、侧围结构(较少)和内外结构(极少)。表1 给出不同结构的湘西民间苗文字例及其汉义。

湘西民间苗文创制时, 遵循“取个人认为最易认易记的汉字或符号作为代表符号”的标准, 直接用含义明确、结构或笔画较简单且日常使用频率较高的汉字或偏旁, 以及极个别无音无义的纯粹符号(如“X”、“~”)作为义符、声符或形符构件[5]。表 2 按构件拼音首字母次序给出从文献[1,5–6]整理出的、作为苗文构件使用的 203 个汉字(偏旁)和两个纯粹符号。

表 1 不同结构湘西民间苗文字例及汉义

Table 1 Xiangxi folk Hmong characters in different structure and the corresponding Chinese meanings

结构类型	字例	汉义
左右结构	板塘苗文: 猭	猪
	老寨苗文: 狘	猪
	古丈苗文: 豝	猪
上下结构	板塘苗文: 蟿	认识
	老寨苗文: 虵	蛇
	古丈苗文: 霏	雨
侧围结构	板塘苗文: 氹	一个
	老寨苗文: 氹	我们
	古丈苗文: 秖	头
内外结构	板塘苗文: 闶	出去
	古丈苗文: 闯	门

1.2 构字方式

湘西民间苗文的字形由构字方式决定。同一个字形可以呈现多种风貌, 但其构字规律固定。当一个苗文由 3 个及 3 个以上部分构成时, 按照构件选取标准, 其中的某 2 个或 3 个部分通常可组成一个简单汉字。此时, 宜将此简单汉字视为一个构件。对文献[1, 5–6]所提及湘西民间苗文进行统计的结果表明, 大部分苗文均可视为二构件型, 仅个别左右结构和上下结构的苗文需当作三构件型处理。

图 1 给出不同结构湘西民间苗文字例的字形拓扑结构和构字方式。图 1(a)~(d)所示的二构件型字例的字形分别取决于构件“口”“打”、构件“尖”“口”、构件“毛”“比”和构件“门”、“竺”。图 1(e)~(f)所示的三构件型字例的字形分别取决于构件“扌”、“氵”、“井”和构件“合”“目”“目”。尽管“打”、“尖”、“竺”和“目目”均可进一步分解为两个构件, 但按照构件选取原则, “打”、“尖”和“竺”宜作为一个构件使用, 无须再拆分为虚线框内的两个部分, 而“目目”则宜继续分解为两个构件“目”和“目”。

2 湘西民间苗文字形动态生成方法

2.1 基本思想

当前计算机处理表意文字时, 主要在“单字”层面上对其逐一编码, 相应的字体设计也必须“逐字”进行。这种方法使得表意文字的字体设计工作量巨大, 同时导致编码字符集也难以瘦身。

与通过 26 个字母的自由组合即动态生成很多的英文单词字形一样, 将有限构件按照一定规则进行组合, 理应也能动态生成无穷多的湘西民间苗文字形。根据造字原理和构字方式, 湘西民间苗文的字形可视为由 2~3 个构件组合运算得到。按照构件在苗文合体字中的位置关系, 可以将组合运算分为 6 种: 左右连(left-right link, LRL)、上下连(up-down link, UDL)、左上包(left-up contain, LUC)、左下包(left-down contain, LDC)、右上包(right-up contain, RUC)和全包(all contain, AC)。两个构件经某种运算生成一个二构件型的苗文字形, 生成结果再与另一个构件进行某种运算, 即可得到一个三构件型的苗文字形。

取表2 中的 205 个构件作操作数进行 6 种运算, 不仅能够生成文献[1, 5–6]中提及的所有湘西民间苗文字形, 还能创制很多新字形。这意味着湘西民

表 2　湘西民间苗文构件

Table 2　Xiangxi folk Hmong characters components

构件拼音首字母	作为构件使用的汉字、偏旁或纯粹符号
A	安、艾、昂、敖
B	八、罢、巴、比、白、保、拨、贝、兵、疒、布、闭、百
C	厂、出、虫、匆、草、处、床、查、吹、橙
D	得、蛋、倒、呆、斗、豆、多、刀、兑、打、都
E	二、而、耳、阝
F	风、非、飞、缶、孚
G	个、工、郭、广、光、勾、敢、国、冠、贵、鬼
H	号、灰、黑、红、黄、会、火、合、后、禾
J	夹、金、架、讲、九、久、介、戒、加、句、尖、己
K	口、克、卡
L	了、列、龙、六、乐、录、绿、流、落、另、来、劣、冷、两、良、朗、略、蓝
M	米、门、毛、目、马、敏、么、木、墨、面
N	能、农、弄、牛、女、奴、年、虐、奈、鸟、闹
O	嘔
P	皮、片、贫
Q	欠、犭、气、去、青、前、七、雀、穷、千
R	日、人(亻)、柔、肉
S	食、上、豕、送、受、杀、色、身、三、四、十、山、氵、扌、饣、戍、死、水
T	土、吞、匈、头、通
W	五、务、攵、万、无
X	忄、朽、相、下、心、夕、雄
Y	一、乙、雨、礻、鱼、右、咽、雅、衣、以、月、芋、因、又、页
Z	知、竺、者、周、足、左、主、助、辶、子、隹、紫、早
纯粹符号	X、~

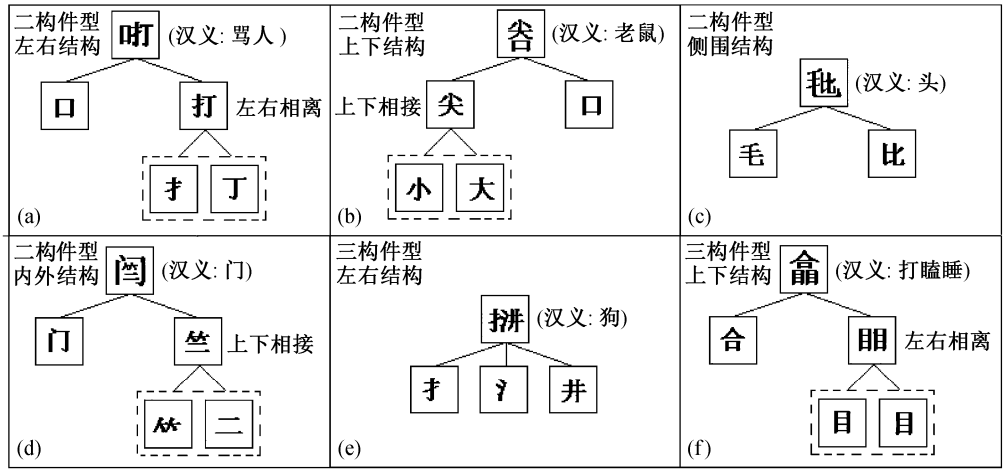


图 1　不同结构湘西民间苗文字例的字形拓扑结构和构字方式

Fig. 1　Glyph topologies and ways of glyph configuration for examples of Xiangxi folk Hmong characters in different structure

间苗文可视为一个开放的文字集合, 采用字形动态生成方法, 能够生成用户所需的各种新的苗文字形。具体例子如下。

1) 左取构件“女”、“米”、“礻”、“亻”和“蛋”, 右取构件“能”, 通过左右连运算, 可分别生成汉义为“年轻媳妇”、“稻谷”、“穿衣”、“人”、“蛋”的二构件型左右结构苗文字形。

2) 上取构件“雨”, 下取构件“加”、“龙”、“助”、“奴”、“送”、“号”、“者”、“朽”、“气”、“风”、“白”等, 通过上下连运算, 可生成表示天气的一组二构件型上下结构苗文字形。

3) 上取构件“虫”, 下取构件“~”, 通过上下连运算, 可以生成汉义为“蛇”的二构件型上下结构苗文字形。

4) 左取构件“疒”, 右取构件“相”, 通过左上包运算, 可生成汉义为“生病”的二构件型侧围结构苗文字形。

5) 左取构件“色”, 右取构件“白”、“黑” “青” “红” “绿”等, 通过左下包运算, 可生成表示颜色的一组二构件型侧围结构苗文字形。

6) 左取构件“扌”、“口”、“月”或“乙”, 右取构件“一”、“二”、“三”、“四”、“五”、“六”、“七”、“八”、“九”、“十”、“百”、“千”、“万”等, 通过左右连或左下包运算, 可生成 4 组与数字相关的二构件型左右结构或侧围结构苗文字形。

7) 左取构件“去”, 右取构件“飞”, 通过右上包运算, 可生成汉义为“飞去”的二构件型侧围结构苗文字形。

8) 外取构件“门”, 内取构件“出”, 通过全包运算, 可生成汉义为“出去”的二构件型内外结构苗文字形。

9) 上取构件“合”, 下取两个构件“目”进行左右连运算的结果, 再作上下连运算, 可生成汉义为“打瞌睡”的三构件型上下结构苗文字形。

2.2 运算符的定义

设 a 为任意湘西民间苗文构件, x 和 y 为整数, 用(x, y)表示像素点位置, 则苗文构件 a 的二值图像可用函数 $F_a(x, y)$定义。$F_a(x, y)=1$ 时, 图像为黑像素, 表示构件有笔画经过(x, y)点; $F_a(x, y)=0$ 时, 图像为白像素, 表示构件无笔画经过(x, y)点。

任意取两个湘西民间苗文构件 a 和 b, 其二值图像函数分别为 $F_a(x, y)$和 $F_b(x, y)$。将 $F_a(x, y)=1$ 和 $F_b(x, y)=1$ 时 x 和 y 的最大、最小值分别记为 maxx(a), maxx(b), maxy(a), maxy(b), minx(a), minx(b), miny(a) 和 miny(b), 则上述 6 种运算符的定义可用逻辑公式描述如下。

定义 1 若$(\text{maxx}(a)\leqslant\text{minx}(b))\wedge((\text{miny}(a)\leqslant\text{miny}(b)\leqslant\text{maxy}(b)\leqslant\text{maxy}(a))\vee(\text{miny}(b)\leqslant\text{miny}(a)\leqslant\text{maxy}(a)\leqslant\text{maxy}(b)))$, 则称 a 左右连 b, 记为 a LRL b。

定义 2 如果$(\text{miny}(a)\geqslant\text{maxy}(b))\wedge((\text{minx}(a)\leqslant\text{minx}(b)\leqslant\text{maxx}(b)\leqslant\text{maxx}(a))\vee(\text{minx}(b)\leqslant\text{minx}(a)\leqslant\text{maxx}(a)\leqslant\text{maxx}(b)))$, 则称$a$上下连$b$, 记为 a UDL b。

定义 3 如果$(\text{minx}(a)<\text{minx}(b)\leqslant\text{maxx}(b)\leqslant\text{maxx}(a))\wedge(\text{miny}(a)\leqslant\text{miny}(b)\leqslant\text{maxy}(b)<\text{maxy}(a))$, 且当 x, y 满足$(\text{minx}(b)\leqslant x\leqslant\text{maxx}(b))\wedge\text{miny}(b)\leqslant y\leqslant\text{maxy}(b))$时, $F_a(x, y)=0$, 则称 a 左上包 b, 记为 a LUC b。

定义 4 如果$(\text{minx}(a)<\text{minx}(b)\leqslant\text{maxx}(b)\leqslant\text{maxx}(a))\wedge(\text{miny}(a)<\text{miny}(b)\leqslant\text{maxy}(b)\leqslant\text{maxy}(a))$, 且当 x, y 满足$(\text{minx}(b)\leqslant x\leqslant\text{maxx}(b))\wedge\text{miny}(b)\leqslant y\leqslant\text{maxy}(b))$时, $F_a(x, y)=0$, 则称 a 左下包 b, 记为 a LDC b。

定义 5 如果$(\text{minx}(a)\leqslant\text{minx}(b)\leqslant\text{maxx}(b)<\text{maxx}(a))\wedge(\text{miny}(a)\leqslant\text{miny}(b)\leqslant\text{max}y(b)<\text{maxy}(a))$, 且当 x, y 满足$(\text{minx}(b)\leqslant x\leqslant\text{maxx}(b))\wedge\text{miny}(b)\leqslant y\leqslant\text{maxy}(b))$时, $F_a(x, y)=0$, 则称 a 右上包 b, 记为 a RUC b。

定义 6 如果$(\text{minx}(a)<\text{minx}(b))\wedge(\text{maxx}(a)>\text{maxx}(b))\wedge(\text{miny}(a)<\text{miny}(b))\wedge(\text{maxy}(a)>\text{maxy}(b))$, 且当 x, y 满足$(\text{minx}(b)\leqslant x\leqslant\text{maxx}(b))\wedge\text{miny}(b)\leqslant y\leqslant\text{maxy}(b))$时, $F_a(x, y)=0$, 则称 a 全包 b, 记为 a AC b。

2.3 运算表达式的构成

根据上述思想, 苗文字形的动态生成过程可表示为由苗文构件作操作数、由构件位置关系决定组合运算符的中缀表达式。表达式中, 所有运算符优先级相同, 均服从左结合规律, 括号内运算符优先级高于括号外运算符。

由于湘西民间苗文大部分为二构件型, 仅个别为三构件型, 所以运算表达式通常只有如下 4 种形式: 1) 构件+运算符+构件; 2) 构件+运算符+构件+运算符+构件; 3)(构件+运算符+构件) +运算符+构件; 4) 构件+运算符+(构件+运算符+构件)。其中, 形式 2 和形式 3 等价。

3　湘西民间苗文字形动态生成方法的实现途径

3.1　基于 IDS 的实现途径

湘西民间苗文字形动态生成方法实现的最直接途径是, 根据运算符定义和组合运算表达式形式, 设计构件组合运算算法和构件像素坐标提取算法, 并通过构造不同构件笔画的生成函数, 动态获取各种不同的苗文字形。但是, 该实现途径的工作量较大。表意文字描述规范[7]为湘西民间苗文字形动态生成方法的实现提供了一种简捷的途径。

表意文字描述规范最早出现在 Unicode 3.0 中。该规范定义了 12 个表意文字描述符(ideographic description characters, IDC), 给出了基于递归定义的表意文字描述算法。算法将表意文字递归地分解为部件的组合, 将文字的结构类型符作为操作符, 文字或部件作为操作数, 用操作符和操作数组成的前缀表达式表示文字字形。算法的理论依据是, 所有的表意文字都可以拆分为更小的部件, 而这些部件本身是表意文字。由于算法允许 IDS 本身继续被分解, 且 Unicode 字符集中表意文字部件存在重复出现情况, 所以, 一个表意文字的字形描述序列可能不唯一。原则上, IDS越短越好。采用 IDS 描述表意文字字形时, 限制序列长度不得多于 16 个 Unicode 码位; 若无 IDC 分隔, 构成序列的部件最多 6 个。表意文字描述规范使得表意文字字形描述与现有文字编码系统相结合, 为表意文字字形动态生成技术的实用化奠定了坚实基础。目前, IDS 已成功应于古籍数字化[8]和错字处理[9], 以及繁体汉字向量组字编辑器[10]、文字影系统(Kage System)[10−11]等动态组字技术研究中。

湘西民间苗文是一种具有固定结构特性的表意文字。虽然 Unicode 组织没有为其分配固定码位, 但 2014 年发布的 Unicode 7.0 的 CJK 区间共收录 74617 个汉字[12], 作为苗文成字构件使用的简单汉字或偏旁皆囊括其中; 不成字构件的纯粹符号也早在 Unicode 字符集中分配了码位。利用表 3 所示的 7 个 IDC 即可实现苗文字形动态生成涉及的6 种构件组合运算。显然, 不占用 Unicode 编码区间, 只需将构件组合运算表达式转换成 IDC与苗文构件的组合描述序列, 利用操作系统自带的IDS 解释机制, 就能够实现湘西民间苗文字形的动态生成方法。

表 3　IDC 与运算符的对应关系

Table 3　Correspondence between operators and IDCs

运算符	对应的 IDC	IDC 的 Unicode 码位
左右连(LRL)	左右⿲	U+2FF0
上下连(UDL)	上下⿳	U+2FF1
左上包(LUC)	左上包含⿸	U+2FF8
左下包(LDC)	左下包含⿺	U+2FFA
右上包(RUC)	右上包含⿹、	U+2FF9
全包(AC)	全包含⿴、上三包含⿵	U+2FF4, U+2FF5

3.2　实现途径的上下文无关文法表示

湘西民间苗文字形动态生成方法实现时, 苗文字形的 IDS 构成可用定义 7 给出的上下文无关文法进行描述。

定义 7　$G=(V_N, V_T, P, S)$。V_N 和 V_T 分别表示非空有穷的非终结符集和终结符集, 且 $V_N \cap V_T=\varnothing$; P 为形如“$\alpha\rightarrow\beta$”的产生式集; S 为文法 G 的开始符, $S\in V_N$ 且 S 至少在一条产生式中作为左部出现。V_N, V_T 和 P 取值如下:

$V_N=\{S, O, A, B, C\}$;

V_T={U+2FF0, U+2FF1, U+2FF4, U+2FF5, U+2FF8, U+2FF9, U+2FFA, U+4E00……U+9FFF, U+0020……U+007E};

P={$S\rightarrow OAA|OAOAA|OOAAA$,

$O\rightarrow$U+2FF0|U+2FF1|U+2FF4|U+2FF5|U+2FF8|U+2FF9|U+2FFA,

$A\rightarrow B|C$,

$B\rightarrow$U+4E00|……| U+9FFF,

$C\rightarrow$U+0020|……|U+007E。

G 中各符号的含义见表 4。

图 1(a)~(f)中 6 个湘西民间苗文字例的字形生成IDS如表 5 所示。按照构件选取标准, 图1(a)~(d)和(f)中字例的字形只有一种描述序列, 而图1 (e)中字例的字形可以有两种描述序列。

4　湘西民间苗文的字形生成测试

目前, 已完成 OpenType 字体布局表的规划, 并以湘西民间苗文的字形动态生成方法及基于 IDS 的实现途径为基础, 设计了构件组合映射脚本, 初步创建了湘西民间苗文 OpenType 字库。以微软 OpenType 字体布局设计软件 VOLT 提供的 Proofing Tool 为测试工具, 对字库进行了初步测

表 4 *G* 中的文法符号含义

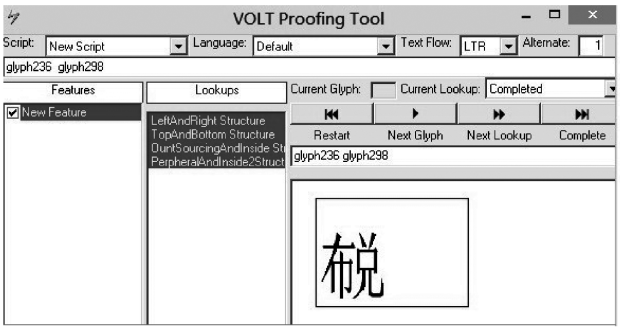
Table 4 Meanings of symbols in grammar *G*

符号	所属集合	含义
S	V_N	湘西民间苗文字形 IDS
O	V_N	对应于组合运算符的 IDC
A	V_N	苗文构件
B	V_N	常用汉字或偏旁
C	V_N	纯粹符号
U+2FF0, U+2FF1, U+2FF4, U+2FF5, U+2FF8, U+2FF9, U+2FFA	V_T	IDC 的 Unicode 编码
U+4E00……U+9FFF	V_T	常用汉字或偏旁的 Unicode 编码
U+0020……U+007E	V_T	纯粹符号的 Unicode 编码

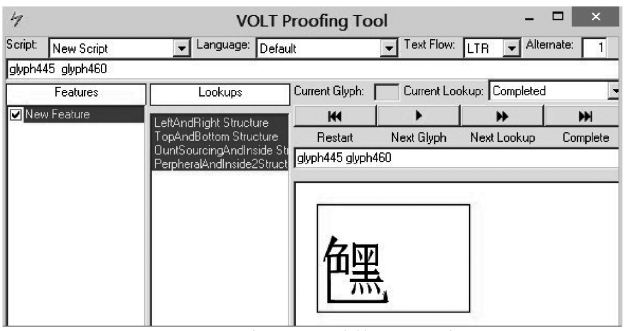
表 5 动态生成湘西民间苗文字形的 IDS 示例

Table 5 IDS examples for dynamically generating glyph of Xiangxi folk Hmong characters

字例编号	组合运算表达式	符号描述序列	编码描述序列
图 2(a)	口 LRL 打	⿰口打	U+2FF0 U+53E3 U+6253
图 2(b)	尖 UDL 口	⿱尖口	U+2FF1 U+5C16 U+53E3
图 2(c)	毛 LDC 比	⿺毛比	U+2FFA U+6BDB U+6BD4
图 2(d)	门 AC 竺	⿴门竺	U+2FF5 U+95E8 U+7AFA
图 2(e)	扌LRL(氵LRL 井) (扌LRL 氵)LRL 井	⿰扌⿰氵井 ⿰⿰扌氵井	U+2FF0 U+624C U+2FF0 U+6C35 U+4E95 U+2FF0 U+2FF0 U+624C U+6C35 U+4E95
图 2(f)	合 UDL(目 LRL 目)	⿱合⿰目目	U+2FF1 U+5408 U+2FF0 U+76EE U+76EE



(a) 左右连所生成左右结构的苗文字形



(b) 左下包所生成侧围结构的苗文字形

图 2 由映射脚本生成的湘西民间苗文字形测试结果

Fig. 2 Test results of folk Hmong characters glyph generated by the mapping script

试。测试结果表明，映射脚本生成的湘西民间苗文字形整齐规范，基本上达到实用要求。对应左右连运算和左下包运算的组合映射脚本所生成的二构件型左右结构和侧围结构苗文字形的测试结果如图 2 所示。

5　结语

以往研究将湘西民间苗文编码限定在 Unicode 私用区[U+EF00~U+FFFF]，每个字形占用一个码位，已创建的苗文 TrueType 字库中的每个文字皆以独立的字形轮廓进行描述。这种方式虽然可行，但不利于移植。本文提出的方法，通过 2 个或 3 个构件和 6 种组合运算符构造运算表达式，再利用 IDC 及苗文构件的 Unicode 编码的组合描述序列来实现运算表达式，借助操作系统自带的 IDS 解释机制，便能动态生成用户所需的各种湘西民间苗文字形，这对于实现湘西民间苗文字形的高效存储和快速显示技术有重要作用。

下一步，拟研究湘西民间苗文字形动态生成方法在无字库苗文处理系统中及互联网上跨平台苗文信息传播中的应用技术。

参考文献

[1] 赵丽明，刘自齐．湘西方块苗文．民族语文，1990, 12(1): 44–49

[2] 莫礼平，周恺卿，蒋效会．板塘苗文的计算机编码及字库创建．吉首大学学报：自然科学版，2013, 34(2): 31–35

[3] 莫礼平，周恺卿，张兆海．基于 Windows IMM-IME 的接口式方块苗文输入法的实现．计算机应用与软件, 2014, 31(3): 64–66, 81

[4] 莫礼平，曾水玲，周恺卿．音形结合的方块苗文输入编码方案研究．计算机科学与探索，2014, 8(8): 1017–1024

[5] 杨再彪，罗红源．湘西苗族民间苗文造字体系．吉首大学学报：社会科学版, 2008, 29(6): 130–134

[6] 龙正海．渝、湘、鄂西水流域方块苗文造字法再探．重庆教育学院学报, 2012, 25(5): 56–59

[7] Lu Qin, Chan Shiutong, Li Yin,et al. Decomposition for ISO/IEC 10646 ideographic characters [EB/OL]. (2004–06–01) [2015–01–06]. http://www.aclweb.org/anthology/W/W02/W02-1209.pdf

[8] 肖禹，王昭．动态组字的发展及其在古籍数字化中的应用．科技情报开发与经济，2013, 23(5): 118–122

[9] 李小庆．面向汉字教学的错字处理工具设计与实现[D]．内蒙古：内蒙古师范大学, 2010: 17–29

[10] 百度百科．动态组字[EB/OL]. (2010–07–10) [2015–01–19]. http://baike.baidu.com/view/908298.htm?fr=aladdin

[11] Miyazaki I, Tomabechi T. Omega/CHISE: a type-setting framework based on the character information service environment [EB/OL]. (2004–05–13) [2015–02–08]. http://coe21.zinbun.kyoto-u.ac.jp/papers/ws-type-2003/077-Omega-CHISE.pdf

[12] The Unicode Consortium. Unicode7.0 character code charts [EB/OL]. (2014–07–16) [2015–03–29]. http://www.Unicode.org/Public/UCD/latest/charts/CodeCharts.pdf