

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2014.024

基于语义构件的甲骨文字库自动生成技术研究

吴琴霞^{1,2,†} 栗青生^{1,2} 高峰^{1,2}

1. 安阳师范学院计算机与信息工程学院, 安阳 455002; 2. 甲骨文数字化工程技术研究中心, 安阳 455002;
† 通信作者, E-mail: wqx0218@163.com

摘要 针对甲骨文字形多变、异体字多等特点, 通过对甲骨字的构成分析, 提出一种基于语义构件的甲骨文字库自动生成方法。该方法以动态描述库为基础, 通过算法提取甲骨字的构件特征信息, 重组笔元生成语义构件, 再给语义构件加上特征描述生成构件知识库。通过仿射变换重复使用语义构件自动生成任意甲骨字。实验表明, 该方法能有效解决甲骨文无字库输入的实现, 还可以解决甲骨字编码、构件统计、未释字的考释等。

关键词 甲骨文; 构件; 字形描述; 有向笔段; 仿射变换

中图分类号 TP391

Study on the Technique of Automatic Generation of Oracle Characters Based on Semantic Component

WU Qinxia^{1,2,†}, LI Qingsheng^{1,2}, GAO Feng^{1,2}

1. School of Computer and Information Engineering, Anyang Normal University, Anyang 455002; 2. Institute of Digital Inscriptions on Bones/Tortoise Shells, Anyang 455002; † Corresponding author, E-mail: wqx0218@163.com

Abstract By analyzing the composition of oracle characters, aiming at the oracle character glyph is changeable, and some of them have many variant forms, The author put forward a kind of automatic generation method of oracle character stock based on semantic component. The method is based on the dynamic description library, Firstly an algorithm is used to get the feature information of the component oracle characters; secondly the semantic component is made by reorganization strokes, finally the component library base was generated by feature descriptions for component. The semantic component library was designed with affine transform, from which each oracle character could be generated. Experiments show that this method can effectively solve the problem of oracle character input without word stock, and can also solve the oracle character encoding, component statistics, and interpretation of the incognizant oracle character, etc.

Key words oracle character; component; glyph description; stroke-segments-vector; affine transform

从 1899 年发现甲骨文至今已经有一百多年的历史, 甲骨文已经成为与考古学、历史学、文字学等有密切关系的一门学科——甲骨学。一百多年来, 出土的甲骨片大约有 10 万多片, 经过几代学者的共同努力目前认读的甲骨文字约占总数的 40% 左右。甲骨文是我国迄今发现的最早的一种成熟文字系统, 甲骨文具有极其重要的学术价值和文化保护价值。如何借助先进的计算机技术对其进行数字化展示、有效保护和方便使用, 具有重要的现实意义。

采用计算机去处理甲骨文需要甲骨文文字专家和计算机专家的通力合作, 这给甲骨文研究领域以及中文信息处理领域提供了新的研究方向。设计甲骨文字库是计算机处理甲骨文以及进行甲骨文研究的前提, 然而甲骨文的特征与现代汉字的特征不同, 它难以定形和输入。许多专家学者通过研究和设计字形库来建立甲骨文字库, 然而字库的形式并不适合甲骨文。首先, 甲骨文字字形复杂, 存在较多的异体字和合体字; 其次甲骨文难以定量, 大量

的未释字和新出土的甲骨文难以表示。分层造字是邓章应等^[1]探讨东巴文、水文等字符形成的过程中提出的,该理论认为文字系统中字符的形成具有层次性和理据性。根据该造字理论结合甲骨文的字体特征,通过分析甲骨文字的有关结构知识,提出一种甲骨文字分层字库的构造方法。该方法将甲骨字表示成由 415 个语义构件和 18 个笔元作为操作数,构件间的 5 种位置关系作为运算符的数学表达式。该方法摆脱了字库的限制,存储量小,方便了文字的跨平台的传播,并且每个构件附加上对应的语义信息,为认知和考释甲骨文带来了方便。

1 甲骨文字形构成分析

参照邓章应提出的分层造字机制,认为在所有的文字系统中字符的构成有具有一定的层次性和理据性。根据字符在形成时是否参考了文字系统中已经存在的字符,可以将字符分成两个层次即初造字符和新造字符。初造字是指某字符在形成的过程中没有参照文字系统中已有的字符;新造字是指某字符在形成的过程中参照了文字系统中已有的字符。在甲骨文文字系统中大多数字符为合体字,王宁^[2]在对汉字的构件研究中提出将汉字构件按照其结构功能划分为 4 类:表形、表义、示音和标示。然而对构成甲骨文字的成字部件与汉字的构件是有区别的,在甲骨部件中其表形和表义很难区分,而且标示的功能几乎不存在。在此基于分层造字原理,根据新造字符形成时参与造字的构件的功能,将参与构件分为表义和表音两大类。理论上可能存在 3 种组合类型:义义组合、义音组合和音音组合。而甲骨文字体中存在的大多数为义义组合^[3]。

义义组合是指某字符形成时两个或两个以上表义字符参与了该字符的构成。如:彡(束)→彡(丝),像两束丝,表示丝线;木(木)→森(森)或𣎵(森),以三木组合表示树多为森。以形状组成的合体字如火(火)→焱(焱),以三个火组合表示大火;新造字表示动作如隹(隹)→𠂔(隹),以手捕鸟之状,表示捕获。匕(爪)+角+牛→𠂔,手抓住了牛角的部分,表示解牛;匕(爪)+皿→𠂔,像手放入器皿中洗涤,表示洗手。

在甲骨字构成系统中,甲骨字大多为基字之间通过某种组合规则相互组合获得的新字符,新造字符是由文字系统中其他字符发生符号联系而成的。大多数合体新造字符都是通过基字和构件之间建立

联系而生成的,通过参与的基字为中介,又可与其他字符发生关联。通过这种联系性可以将所有的字符都联系起来,这样整个文字系统中的字符就形成了一个有机整体,联系关系如图 1 所示。

在甲骨文文字系统中有大量的新造字符富有层次性,即新造字符的形成过程中所参照字符也为新照字符。此类字符表现出构件具有一定的层次性,且随层次的增多,构件数量也逐渐增多,字符显得更加复杂。层次性新造的甲骨字的构件含义不是一次性表现出来的,而是随层次逐级生成的。甲骨文字系统中层次组合结构绝大多数为二次组合,占总字数的 13%左右。例如图 2 中的甲骨字。

除此之外,甲骨文中的合体字还具有以下特殊性:1)所使用的构件区别方式更为多样,有些原本不一样的部件可能混同,致使某些部件类化;2)合体字所产生的异体字类型众多,如笔画差异、部件的多寡、部件位置的变化等。

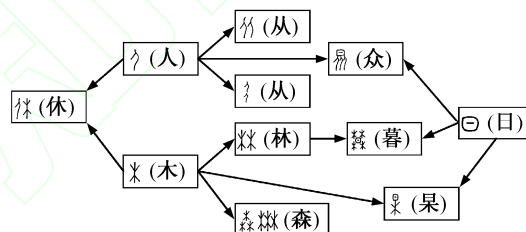


图 1 新造字由基字生成
Fig.1 New character generated by the base word

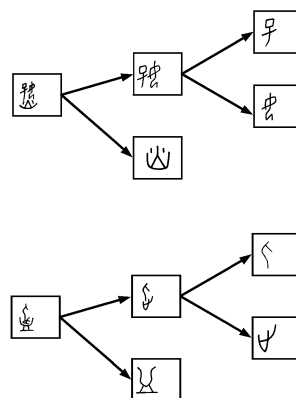


图 2 层次性的新造字
Fig.2 Hierarchical structure of the new character

2 分层甲骨文字库的原理

与汉字不同,甲骨文字是契刻文字,不像现在的汉字一样有完整的笔画结构。根据甲骨字的特殊

性引入了笔元和笔段参与甲骨字合体字的构成。对目前 1024 个可释甲骨文的笔元进行统计，将甲骨文的笔元分成 3 类基本的笔元：第一类是直线笔元，第二类是折线笔元，第三类是弧线笔元，如表 1 所示。根据上述甲骨字体构成分析，由于构件组字时会发生一系列变化，所以在甲骨文构件特征提取中，以有向笔段和笔元对甲骨文字形构件进行动态的矢量化描述。将甲骨文字形构件用字形骨架来表示，然后在骨架上增加特征点，最后通过存储特征点形成构件特征库，以后的字形的编辑和输出都是由构件来动态地生成甲骨文字。笔元与笔段的定义^[4]如下：

有向笔段：即有方向的线段，设 (X_1, Y_1) 是线段的起点， (X_2, Y_2) 是线段的终点，则一个完整的有向笔段的描述为

$$B12 = \{(X_1, Y_1)|(X_2, Y_2)\}。 \quad (1)$$

笔元：笔元是由一个或多个有向笔段组成的一个完整的笔画结构，设一个笔元由 n 个有向笔段来组成，则笔元的描述^[5]为：

$$SS_n = \{(X_{11}, Y_{11})|(X_{j1}, Y_{j1}), (X_{12}, Y_{12})| \\ (X_{j2}, Y_{j2}), \dots, (X_{in}, Y_{in})|(X_{jn}, Y_{jn})\}。 \quad (2)$$

通过上述甲骨字形成的分析，可以看出甲骨字像汉字一样具有相同的组成原理。汉字组字的最小单位为笔画，笔画组成部件。部件是汉字体系结构的基本单元^[3]。由于大部分甲骨字含有一定语义，因此这里提出的甲骨文构件不同于汉字字形的部件，而是指具有独立的语义信息的最小单位。这样每个甲骨字可以表示为由几个构件和笔元构成。许多构件和笔元在不同的甲骨字中经常出现，所以甲骨字库中只存储这些构件和笔元。每个甲骨字的数学描述如下。

一个甲骨字 JGZ 由若干个构件 $C_i(i=1, 2, \dots, n)$








构成，每个构件由若干笔元构成 $SS_i(i=1, 2, \dots, n)$ 构成，每个笔元由若干个笔段 B_{ij} 构成。即

$$JGZ = \{C_1, C_2, \dots, C_n\}; C_i = \{SS_1, SS_2, \dots, SS_n\}; \\ SS_i = \{(X_{i1}, Y_{i1})|(X_{i2}, Y_{i2})\}; \quad (3)$$













3 构件知识库的建立与应用

通过分析 1024 个已释甲骨字，提取 415 个具有独立语义单位的甲骨文构件。甲骨字是由构件和基本笔元组成，不同的组合和位置关系组成不同的甲骨字。为实现甲骨字体的笔元分解，将所有的构件组成的规则有机组织起来，形成甲骨字构件知识库。将甲骨字体的笔元进行特征提取归类，搜索部件知识库，可以得到构件的特征信息。根据构件和构件间的位置关系，从整体上把握甲骨字的结构。

3.1 构件知识库的建立

虽然有很多甲骨字使用相同的构件，但这些构件是有区别的。如、、、、和这几个甲骨字中都含有“”这个构件，但在不同的甲骨字中构件的外观是有区别的，但它们表达的语义信息是相同的。因此设计的构件基本库中只包含一个构件信息，为了能够动态地修改成多种外观，构件知识库里需要保存构件的如下特征信息^[6]。



















笔元分界点：为了区分构件中不同笔元，以便组字时方便修改和动态缩放，每个构件的基本笔元必须进行区分。

笔元间的组合关系：甲骨字形中因其在字中的位置和组合关系不同分为相离(、、、)用“disjoint”表示，相接(、、、)用“join”表示，相交(、、、)用“cross”表示。

笔元数：每个构件的笔元的个数。

部件的语义：甲骨字都具有一定的含义，甲骨字构成中义义组合占有绝大部分。为了更好的释读

表 1 甲骨文的基本笔元
Table 1 Jiaguwen's basic strokes

笔元类型	基本笔元						
直线笔元							
	横线	竖线	撇线	捺线	点线		
折线笔元							
	上折线	下折线	左折线	右折线	撇折线	捺折线	
弧线笔元							
	左弧线	右弧线	上弧线	下弧线	横波浪弧线	竖波浪弧线	椭圆线

甲骨文,同时也为了进一步考释甲骨文需要给部件添加语义信息。

构件的骨架提取以甲骨文动态描述库^[4]为基础,提取步骤如下:

1) 加载动态描述库,通过字型目录找到甲骨字对应的骨架信息,如查找 𠄎 对应的骨架信息为 50, -64, 0, 2, -12, -7, -6, -8, 9, -64, 0, 2, -12, 8, -6, 8, 9, -64, 0, 2, -7, 3, 7, -1, 8, -2, 12, 3, 10, 1, 13, -64, 0, -2, -3, 5, -4, 4, -1, -2, 1, -4, 3, -64, 0, -2, -3, -2, 4, -64, -64, , , , , ,^[7];

2) 分析动态描述库中的特征点,执行 Split(笔元, "-64, 0, ")得出每一个笔元信息;

3) 合成有联系的笔元,生成对应的语义构件;

4) 保存语义构件的特征信息。

以甲骨文语义构件 𠄎 为例,进行特征提取,提取的特征信息如表 2 所示。

表 2 “𠄎”构件对应的特征信息
Table 2 Feature information of “𠄎” component

笔元顺序	对应的特征点	笔元外观	笔元组合关系	部件语义	部件编号
第一笔元	(22, 13, 23, 27, 19, 28, 18, 32, 23, 30, 21, 33)		一二笔元 cross		
第二笔元	(18, 17, 25, 16, 24, 19, 12, 21, 16, 23)		二三笔元 join	女性	005
第三笔元	(18, 17, 18, 24)		一三笔元 disjoint		

3.2 件外观的动态修改

为了满足不同甲骨字中构件的外观不同,先对构件的骨架信息进行编辑和修改^[7]。构件的编辑算法算法如下。

步骤 1 单击编辑进入构件编辑界面。

步骤 2 选择操作类型。若操作类型为“Ins”,则转步骤 2.1;若操作类型为“Mov”,则转步骤 2.2;若操作类型为“Del”,则转步骤 2.3;若操作类型为“Edit”,则转向步骤 2.4;若操作类型为“NoOper”,则转向步骤 3。

步骤 2.1 插入笔元。插入具有 n 个笔段的笔元 E_n 是通过逐个插入该笔元的笔段来实现的,设插入位置为 pos :

步骤 2.1.1 在插入位置插入一个界点 $(m, 0)$;将笔段计数变量 i 置 1;插入位置 pos 自加 1。

步骤 2.1.2 在插入位置插入 E_n 的第 i 个笔段

$(B[i].x, B[i].y, Z[i].x, Z[i].y)$ ($B[i]$ 和 $Z[i]$ 分别为第 i 个笔段的势点和驻点);在 $B[i]$ 和 $Z[i]$ 之间画线段;插入位置 pos 自加 2;笔段数 i 自加 1。

步骤 2.1.3 若 $i < n$, 转步骤 2.1.2; 否则, 笔元数 ele_num 自加 1, 转步骤 2。

步骤 2.2 整体移动笔元。整体移动笔元是通过修改笔元中各个点的坐标(界点除外)来实现的。设 $Vertex[i]$ 是待移动笔元中的第 i 个点, dx 和 dy 分别是水平增量和垂直增量, $VerCount$ 是待移动笔元中除界点之外的点数:

步骤 2.2.1 置 i 为 1。

步骤 2.2.2 $Vertex[i].x$ 和 $Vertex[i].y$ 分别增加 dx 和 dy ; i 自加 1。

步骤 2.2.3 若 $i < VerCount$, 转步骤 2.2.2; 否则, 转步骤 2。

步骤 2.3 删除笔元。删除笔元是通过删除笔元的各个点来实现的。设 $Vertex[i]$ 是待移动笔元中的第 i 个点, $VerCount$ 是待删除笔元中除界点之外的点数。

步骤 2.3.1 删除一个界点 $(m, 0)$; 置 i 为 1。

步骤 2.3.2 删除 $Vertex[i]$; i 自加 1。

步骤 2.2.3 若 $i < VerCount$, 转步骤 2.3.2; 否则, 转步骤 2。

步骤 2.4 编辑笔元。编辑笔元主要是指移动笔元中的某个笔段, 这主要通过移动相应笔段的势点或驻点来实现。设待移动的笔段的势点或驻点为 $Vertex[i]$, dx 和 dy 分别是水平增量和垂直增量, 移动该点的方法为:

$$Vertex[i].x = Vertex[i].x + dx;$$

$$Vertex[i].y = Vertex[i].y + dy;$$

转步骤 2。

依据上述算法, 可以动态地生成不同外观的甲骨构件, 如表 3 所示。

表 3 构件的不同外观
Table 3 The different appearance of the component

库存基字形	变形 1	变形 2	变形 3	变形 4	变形 5

3.3 于仿射变换的构件复用

仿射变换是在平面几何上的一种线性变换, 可以将二维图形旋转、缩放、切变和投影等。所以仿射变换可以达到构件复用的功能^[8]。仿射变换的一

般形式如下：

$$\begin{cases} X = a_1x + b_1y + c_1 \\ Y = a_2x + b_2y + c_2 \end{cases} \quad (4)$$

在对构件进行仿射变换生成甲骨字时，只需要确定 (a_1, b_1, c_1) 、 (a_2, b_2, c_2) 这 6 个参数，部件上每一个点 (x, y) 就会通过式(4)生成甲骨字上的一个点 (X, Y) 。在构件特征库里和甲骨字的动态描述库中各选取 3 个点分别是 (x_1, y_1) 、 (x_2, y_2) 、 (x_3, y_3) 和 (X_1, Y_1) 、 (X_2, Y_2) 、 (X_3, Y_3) ，将这些点的坐标代入下面的方程组：

$$\begin{cases} \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \\ \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \end{cases} \quad (5)$$

求解式(5)可以得到

$$\begin{aligned} a_1 &= \frac{(X_1 - X_2)(y_1 - y_2) - (X_1 - X_3)(y_1 - y_2)}{(x_1 - x_2)(y_1 - y_3) - (x_1 - x_3)(y_1 - y_2)}, \\ a_2 &= \frac{(Y_1 - Y_2)(y_1 - y_3) - (Y_1 - Y_3)(y_1 - y_2)}{(x_1 - x_2)(y_1 - y_3) - (x_1 - x_3)(y_1 - y_2)}, \\ b_1 &= \frac{(X_1 - X_2) - a_1(x_1 - x_2)}{y_1 - y_2}, \\ b_2 &= \frac{(Y_1 - Y_2) - a_2(x_1 - x_2)}{y_1 - y_2}, \\ c_1 &= X_1 - a_1x_1 - b_1y_1, \quad c_2 = Y_1 - a_2x_1 - b_2y_1. \end{aligned}$$

将计算所得的 6 个参数代入式(4)，甲骨字动态描述库中甲骨字骨架上的每一对 (X, Y) 都可以在构件骨架上找到一对 (x, y) 与之对应。这样就使得同一个构件可以改变位置、大小和形状，变化出一个与不同甲骨字相应部位都可以相似的构件。从而实现了使用 415 个构件和 18 个笔元可以生成所有甲骨字的目的。

仿射变换的关键是在甲骨字和构件上选取 3 个对应点，3 个点在变换后与原甲骨字字形位置不变，它完全决定了变化后的甲骨字构件的位置、大小和形状。选点不当会造成字形不理想^[9]。针对这种情况，通过 200 个仿射变换模拟甲骨字的实验，总结出以下选点规律。

- 1) 直线笔元：选开始节点的两个分叉点和末节点。两个分叉点决定笔元的宽度，末节点决定笔元的长度。
- 2) 折线笔元：选开始节点、中间拐点和末节点。
- 3) 弧线笔元：选开始节点、中心点和末节点。
- 4) 方形的构件：选 4 个角中的 3 点。需要上部对齐则选上部左右两角和下部一点，需要下部对齐则相反。
- 4) 三角形的构件：一般选最上点和最下两点。但如果构件没有对正，则可适当在构件或甲骨字上向左或右偏移下面的某一点，直至对齐。
- 5) 在左边的构件：选最上一点、最下一点和最左一点，在右边的偏旁则反之。

4 甲骨字构形系统分析

系统以甲骨文的笔元矢量图为基础的对象，经过特征提取整合等步骤后生成构件知识库，从而得到甲骨字的组成笔元之间的位置关系，产生甲骨字的构件描述。通过这个系统，可以统计出甲骨字构件的使用频率，也可产生甲骨字的数字表达式和甲骨字编码，方便互联网的传输。甲骨字自动构形流程图如图 3 所示。

4.1 甲骨文字形结构特征

在汉字的形成过程中，将部件的构成方式和规则称为汉字的结构，也称为汉字的字形结构。每一个汉字字符可以看成一系列部件在平面上的有序组合，每种组合方式可以当成一种拓扑结构^[10]。每个

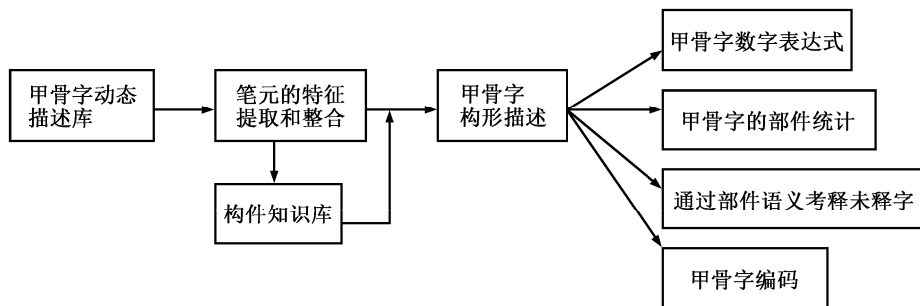


图 3 甲骨字构形系统

Fig. 3 Jiaguwen's configuration system

汉字的拓扑结构的形式由组成该字符的部件数和每个部件在平面上的位置决定,并且与部件的选择和分析方法直接相关^[11]。从数以千计的甲骨文字形拓印图中可以归纳出五大类基本结构:1)独体结构用“D”表示;2)左右结构,中又分为相接(LRj)和分离(LRf);3)上下结构,也分为相交(TBj)和分离(TBf);4)半包围结构(BBW);5)全包围结构(QBW)。因为甲骨字构造字时,有很多异体字,异体字大部分是构件位置的异体,如表4所示。

通过表4分析可知,如果构件的语义主体形状及个数相同,那么组合的方式并非是区分不同字的主要特征^[11]。

4.2 甲骨字的构件描述

以语义构件以及构件间的位置来表示甲骨字,实现了甲骨字的无字库输入和输出。甲骨字所对应的构件描述信息为该字所对应的构件编号,如“𠩺”对应的构件编号为018和005,为了反映构件间的位置关系,给出每一个构件的中心点018的中心点为(10,10);005中心点为(10,20)。

表4 结构位置形成的异体字

Table 4 Variant forms of Jiaguwen by positon structure

汉字	甲骨 文	异体 字 1	异体 字 2	异体 字 3	异体 字 4
马	𠩺	𠩺	𠩺	𠩺	𠩺
森	𠩺	𠩺			
雨	𠩺	𠩺	𠩺	𠩺	𠩺
名	𠩺	𠩺			

表5 甲骨字对应的构件描述

Table 5 Component description of Jiaguwen

甲骨字	构件编号	中心坐标
𠩺	018, 019	(10, 20); (10, 20)
𠩺	018, 009, 009	(10, 20); (10, 5); (10, 25)
𠩺	011, 009	(10, 10); (10, 20)
𠩺	004, 004	(5, 10); (15, 10)

4.3 实验

语义构件库中的每一个构件用特征点保存,现在已经总结出415个语义构件^[12],存储量为126KB,用500个甲骨字做实验。在存储甲骨字时只需给出构件名(即索引信息)3个字节,坐标参数就可以了。采用该方法组成的字库和使用Bezier曲线存储每一个甲骨字的存储量对比如图4所示。

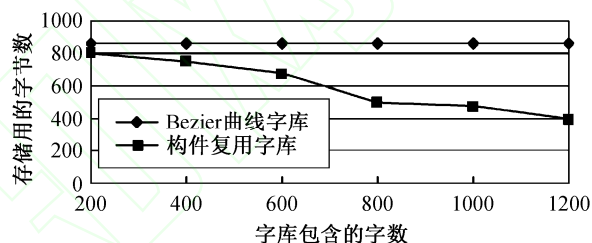


图4 平均存储量变化图

Fig. 4 Change graph of the average storage

构件笔元组成甲骨字的效果图如表6所示。

5 结束语

本文以已释的1024个甲骨字为研究对象,使用骨架信息的提取等基本操作,产生了甲骨字的笔元特征点。以完整的语义信息为基本构件,完成了

表6 由构件和笔元生成的甲骨字

Table 6 Jiagu characters were composed of component and character stock

𠩺	𠩺	𠩺	𠩺	𠩺	𠩺	𠩺
𠩺	𠩺	𠩺	𠩺	𠩺	𠩺	𠩺
𠩺	𠩺	𠩺	𠩺	𠩺	𠩺	𠩺
𠩺	𠩺	𠩺	𠩺	𠩺	𠩺	𠩺

笔元向构件的转化, 生成了甲骨字的构件知识库。最后通过仿射变换复用构件生成每个甲骨字形。采用此方法为以后的语义构件统计, 甲骨字形形成, 甲骨字的编码和甲骨字中未释字的考释打下基础。

虽然这种方法在运行时还有不少改进的地方, 但通过这种方法可以实现甲骨字分级字库的建立, 大大减少了人工的工作量, 使得字库的容量可以动态地增加和修改, 并且存储量小, 方便传播。

参考文献

- [1] 邓章应. 哥巴文造字机制研究. 中国文字研究: 第2辑. 郑州: 大象出版社, 2008
- [2] 王宁. 汉字构形学讲座. 上海: 上海教育出版社, 2002
- [3] 白小丽. 甲骨文与原始楔形文字新造合体字构成比较. 南昌大学学报: 人文社会科学版, 2011, 42(5): 137-142
- [4] 栗青生, 吴琴霞. 基于甲骨文字形动态描述库的甲骨文输入方法. 中文信息学报, 2012, 26(4): 28-33
- [5] 吴琴霞, 栗青生. 基于有向笔端甲骨文输入方法的设计与实现. 计算机应用, 2012, 32(8): 2374-2377
- [6] Li Qingshen, Yang Yuxing. A human-computer interactive dynamic description method for Jiaguwen Characters. Procedia Engineering, 2012: 1013-1017
- [7] 栗青生, 吴琴霞, 杨玉星. 甲骨文字形动态描述库及其字形生成技术研究. 北京大学学报: 自然科学版, 2013, 49(1): 61-67
- [8] 魏海涛. 计算机图形学. 北京: 电子工业出版社, 2003: 51-68
- [9] 冯万仁, 金连文. 基于部件复用的分级汉字字库的构想与实现. 计算机应用, 2006, 26(3): 714-716
- [10] 栗青生, 杨玉星. 甲骨文识别的图同构方法. 计算机工程与应用, 2011, 47(11): 112-114
- [11] 顾绍通, 马小虎, 杨亦鸣. 基于字形拓扑结构的甲骨文输入编码研究. 中文信息学报, 2008, 22(4): 123-128
- [12] 沈建华, 曹锦炎. 新编甲骨文字形总表. 香港中文大学出版社, 2001