

汉字笔顺识别的算法与实现

李正华

(湖南城市学院信息与计算机科学系, 湖南 益阳 413000)

摘要:提出了一种点阵汉字的笔顺字模的描述方法以及基于这种描述方法的汉字笔顺识别的算法和实现。

关键词:汉字笔顺识别; 识别算法; 点阵汉字

中图分类号: TP317.2

文献标识码: A

文章编号: 1008-1666(2004)02-0039-02

Algorithm and implementation in Chinese Character's Order of Strokes Recognition

LI Zheng-hua

(Department of Information & Computer Science, Hunan City University, Yiyang, 413000, China)

Abstract: On the basis of the coding principle about Chinese characters' order of strokes, the algorithm and implementation in Chinese characters' order of strokes recognition suggested in the paper.

Key words: Chinese characters' order of strokes recognition; algorithm of recognition; coding principle

1 引言

汉字笔顺是汉字书写规范化的重要环节。以显示汉字笔顺为主要功能的汉字笔顺字典软件的关键技术在于生成汉字笔顺字库。所谓汉字笔顺字典就是以动画形式模拟汉字书写,同时还显示每一笔的状态和该汉字的笔画数,这种字典无论用于教学还是作个人查阅都非常实用。通过对汉字结构的分析和笔顺显示的要求,提出一种汉字笔顺的识别算法。该算法只需少许人工干预即可提取汉字的笔顺信息,且适用于任何大小任何字体的点阵汉字。

2 问题描述

国标点阵字模用一个二进制位表示一个点的显示与否。一个 24×24 点阵的汉字字模需要 $3 \times 24 = 72$ 个字节来表示,因为每行 $24/8 = 3$ 个字节,共24行,其特点是比特位所在位置即为点的位置。显示扫描时只要将对应字模按从左到右、从上到下顺序进行即可,也就是说字模的存储位置包含了点的显示位置信息。这种字模信息的描述方式,对点阵汉字的显示而言是精简而完善的,但是它没有也不可能包含汉字的笔顺信息。

要利用点阵字模模拟笔顺写字,只要将相应笔画的显示点按笔顺顺序显示即可。因此,笔顺点阵字模必须包含两种信息,即显示点的位置信息和它的笔画顺序信息。

3 算法描述

显然,根据上面叙述的笔顺点阵字模要求,完全可以按照已有的点阵字模用人工编码实现笔顺点阵字模,但工作量大、容易出错。下面先说明笔顺字模的描述规则,然后给出从点阵字模生成笔顺字模的汉字笔顺识别算法。

3.1 笔顺字模的描述

总的原则是将点阵字模一个比特位表示一个点扩充为一个字节表示一个点。原因有二:其一,必须用一个字节表示一个点才能包含该点的笔顺信息;其二,笔顺字模在字库中的索引方法上与点阵字模的基本一致,只是偏移量扩大8倍。类似于点阵字模,该笔顺字模中点的存储位置决定了点的显示位置,点的字节值(ASCII字符的值)则包含点的笔顺信息。点的字节值的确定规则:

(1)不显示点即空白点用ASCII字符的空格表示。

(2)笔画点用从ASCII字符的'A'开始依次表示,而每一笔的起点用从ASCII字符的'S'开始依次表示。例如,

