

蒙古文单词音节自动拆分软件的设计

菊花

(内蒙古师范大学 传媒学院,内蒙古 呼和浩特 010022)

**摘 要:** 阐述了蒙古文单词音节自动拆分软件的设计过程. 该软件能够自动拆分一个蒙古文词的音节构成, 并能判断构成单词的音节个数及音节类型.  
**关键词:** 蒙古文单词; 音节; 软件设计  
**中图分类号:** TP 391. 2    **文献标识码:** A    **文章编号:** 1001-8735(2006)04-0436-03

蒙古文在长期的实践中逐渐形成了音节文字的特点,因此在蒙古语文教学中学习拆分音节是必不可少的一环. 本文介绍了蒙古文单词音节自动拆分软件的设计过程,该软件能自动拆分蒙古文单词的音节构成,并能判断构成单词的音节个数及音节类型. 学生在自主学习的过程中,可以使用该软件分析一个蒙古文单词音节的构成情况.

1 拆分音节模型

蒙古文音节是人们在讲话中自然分解出的发音单位<sup>[1]</sup>. 一个音节一般由几个音素组成,有时一个音素也可以组成一个音节. 一个蒙古文词的构成如图 1 所示.

蒙古文单词的音节是以元音为中心构成的<sup>[1]</sup>,单词的音节有 Y, YF, YFF, FY, FYF, FYFF 等 6 种类型,其中 Y 代表元音, F 代表辅音. 6 种类型的音节分布范围不完全一样,单音节词和多音音节词的第 1 音节可以是 6 种类型中的任一种,但是,多音音节词的后续音节只能是 FY, FYF, FYFF 等 3 种类型中的任一种.

拆分音节的模型可看作 4 元组(Q, , F)<sup>[3]</sup>,其中: Q 是组合音节过程中的各个状态集; 是蒙古文音素字母表,包括所有元音和辅音,并且要区分词首、词中和词尾音素; 是生成音节的转移函数; F 为接受状态,输入一个蒙古文音素时,若 Q 能通过转移函数 达到 F,则能构成一个音节.

根据上述音节的分类,组合音素构成音节,其表示如图 2 所示,其中 Y1 表示词首元音, Y2 表示词中元音, Y3 表示词尾元音, F1 表示词首辅音, F2 表示词中辅音, F3 表示词尾辅音. 拆分音节模型中 4 元组之间的关系如图 3 所示.

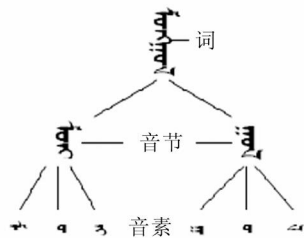


图 1 词的组成

	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
Y <sub>1</sub>	N	Y	Y	N	Y	Y
Y <sub>2</sub>	Y	Y	Y	Y	Y	Y
Y <sub>3</sub>	Y	Y	Y	Y	Y	Y
F <sub>1</sub>	N	Y	Y	N	Y	Y
F <sub>2</sub>	Y	Y	Y	Y	Y	Y
F <sub>3</sub>	Y	Y	N	Y	Y	N

图 2 的表示

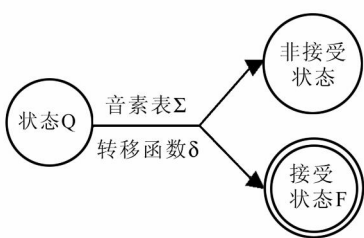


图 3 4 元组的关系

2 自动拆分音节过程描述

2.1 自动拆分音节的步骤 输入或选择一个蒙古文单词,然后单击拆分音节的按钮,可以调用拆分音节过

收稿日期: 2006 - 05 - 13  
基金项目: 国家语委科研项目“民文资源库建设工具软件”之子课题“蒙古文资源库建设工具软件”(MZ115 - 022)  
作者简介: 菊花(1974 - ),女(蒙古族),内蒙古通辽市人,内蒙古师范大学讲师,主要从事计算机教育应用研究.

程. 这时对应的输出框中显示该词的音节构成、音节类型和音节个数. 自动拆分一个单词音节的步骤为:

输入/选择单词——拆分词的全部音素——判断组合音节——输出各音节、音节类型、音节个数.

实现第2步“拆分词的全部音素”时,可利用输入框的KeyPress事件来捕获用户按键的KeyAscii值.用该值作为判断一些变形元音的依据,并使用Chr函数把该值转换成对应的字符.

实现第3步“判断组合音节”时,主要由以下两步实现:

(1) 统计构成所拆分词的全部音素中元音的个数.依据蒙古文元音表判断组成蒙古文单词的元音部分时应注意:组成蒙古文词的音素个数均不相同,也就是保存这些音素的数组的大小是不定的,所以需要定义动态数组来调整数组的长度;用明安途输入法输入蒙古文单词时,最后按的空格键也被保留到数组中,所以循环次数等于数组长度减一;由于所选蒙古文输入法的特殊性,单独输入词首元音时后面需要按“\*”键,输入词中元音时前后需要按“\*”键,输入词尾元音时前面需要按“\*”键,因此列写判断元音语句Case的条件表达式时,应使用Left函数和Right函数去掉前后多余的符号“\*”;判断连接到特殊辅音后有形变的元音时不能用上述Case语句,只能使用它的编码来判断.

(2) 依据所统计的元音个数,组合该词的音节部分.组合音节时应考虑以下情况:元音个数numYuan=0时,给用户出错提示;元音个数numYuan=1时,生词由一个音节组成,直接输出原词;元音个数numYuan=2时,判断两个元音是否连续,连续的情况下生词由一个音节组成,直接输出原词,否则调用组合音节的过程,判断拆分两个音节及音节类型;元音个数numYuan>2时,直接调用组合音节的过程,判断拆分各音节、音节类型及音节个数.

判断组合音节时应注意:For循环的次数等于元音的个数numYuan加辅音的个数numFu.第2步的判断中除了元音,其余字符都看成辅音来处理;判断及连接一个元音后连续出现的辅音;组合音节的过程中,已使用过的音素的标志位要清零;正确输出每个音节并判断音节的类型和个数等.

2.2 拆分音节流程图 拆分音节的流程图如图4所示,该图说明了自动拆分音节的过程.

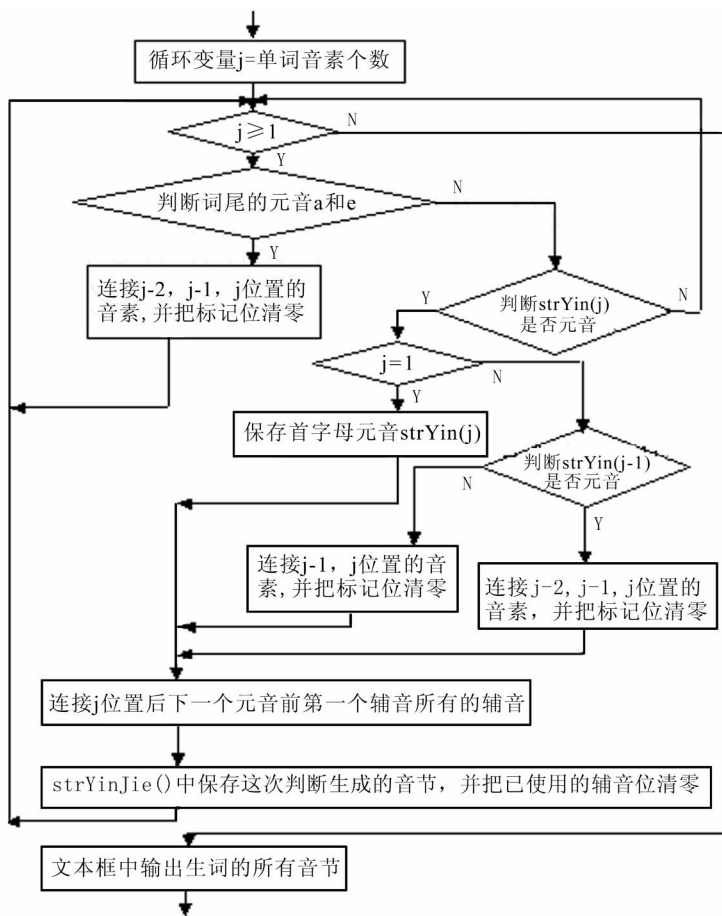



图4 拆分音节流程图

### 3 蒙古文单词音节拆分软件的应用

#### 3.1 软件应用举例

蒙古文单词音节自动拆分软件界面如图5所示.用户在输入框中输入所分析的单词,单击“音节”按钮,输出框中显示所分析词的音节、音节类型和音节个数.同样,用户可以从输入框下显示的文本中选择一个词查看它的音节构成.例如,用户输入或选择“”后,分析结果如图6所示.另外,可单击帮助按钮获得有关使用该软件的帮助信息及蒙古文单词的音节知识信息.

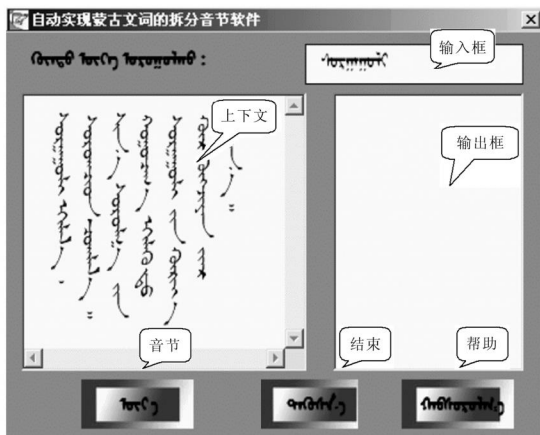


图 5 拆分软件界面图

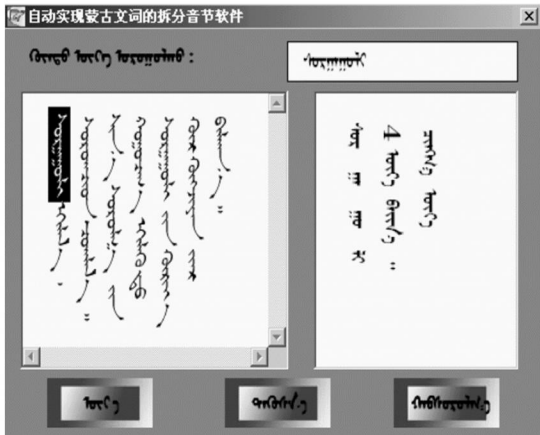


图 6 拆分出的音节

3.2 软件测试

常用的软件测试方法有白盒测试和黑盒测试方法. 本设计的重点是软件的功能测试, 所以采用了黑盒测试方法. 测试用例选择了 200 个蒙古文单词进行小样本测试, 选择样本时尽量平衡选择了由不同蒙古文元音和辅音构成的单词. 对 200 个样本的音节分析结果见表 1. 测试中影响正确率的原因主要有: 对所输入的蒙古文单词, 该软件无法判断它的正确性; 用户输入蒙古文单词时, 由于不熟悉蒙古文键盘表而误输入同形不同音的语音音素, 所以判断音节类型的准确率较低.

表 1 200 个样本的音节分析结果

测试项	正确数	错误数	正确率
拆分音节	181	19	90.5 %
判断个数	181	19	90.5 %
判断类型	169	31	84.5 %

总之, 蒙古文单词音节自动拆分软件能够自动分析一个蒙古文单词的音节构成, 而自动拆分一个单词的音节是自动拆分一个单词的词干和词根的基础. 该软件可以作为蒙古语文教师的工具软件, 同时也能对学生的自主学习起到较好的帮助作用.

参考文献:

[1] 清格尔泰. 蒙古语语法 [M]. 呼和浩特: 内蒙古人民出版社, 1991.  
[2] 郑阿奇. Visual Basic 实用教程 [M]. 北京: 电子工业出版社, 2003.  
[3] Michael Sipser. Introduction to the Theory of Computation [M]. China Machine Press, 2002.  
[4] 那顺乌日图. 蒙古文词根、词干、词尾的自动切分系统 [J]. 内蒙古大学学报: 自然科学版, 1997(2): 53 - 57.

Design of Mongolian Syllables Dividing Software

J Ühua

(College of Media, Inner Mongolia Normal University, Huhhot 010022, China)

**Abstract:** This paper introduced the process of designing a Mongolian syllables dividing software. The software can automatically divide the syllables of a Mongolian character and can judge count and kind of the syllables.

**Key words:** a Mongolian word; syllables; software design

【责任编辑 陈汉忠】