

中国哈萨克阿拉伯文与哈萨克斯拉夫文文本转换

伊力亚尔·加尔木哈买提，古丽拉·阿东别克

(新疆，乌鲁木齐，新疆大学信息科学与工程学院计算机系，830046)

摘要：全世界的哈萨克族都使用同一种语言---哈萨克语，但是由于历史的原因，形成了基于同一语言的两种文字形式这种特殊情况。这两种文字形式分别是以阿拉伯文字母为基础的中国哈萨克文字和以斯拉夫文字母为基础的哈萨克斯坦文字。这种特例，为哈萨克族科学技术的发展和文化的交流带来了不便，因此开发和完善哈萨克语这两种文字形式之间的计算机转换系统具有重要意义。本文将介绍用 VC++6.0 编写的哈萨克语的两种文字形式间的计算机自动转换，采用机器翻译的方法并在 Windows 环境下实现哈萨克语的两种文字形式间的智能转换。

关键字：哈萨克斯拉夫文；哈萨克阿拉伯文；智能转换

Conversion procedure of Kazakh letter of different form

Ylyar.Jarmakamet, Gulia.Altenbek

(School of Information Technology and Engineering, Xinjiang University, Urumchi, 830046)

Abstract: All the Kazakh people in the world use the same language---Kazakh, but because of the historical reason, this become special fact that the same language is based on two different writings. The two different writings is Arabic based letter and Slavic based letter for Kazakh. This special case bring inconvenience for the development of science and intercourse of culture, so it is significant to develop and perfect conversion system in two different writings of Kazakh. The text is the introduction of automatic conversion program in two different writings of Kazakh, the program make aptitude conversion of two different writings in the condition of windows and automatic translation will be used.

Keywords: Kazakh in Arabic; Kazakh in Slavic; automatic conversion

6 引言

哈萨克族作为跨境民族居住在不同的国家和地区，因此跨境文化交流的程度和效率对哈萨克族的发展状况具有重大影响。由于历史原因，同一哈萨克族语言发展出两种不同的文字形式---以阿拉伯文字母为基础的哈萨克文字和以斯拉夫文字母为基础的哈萨克斯坦文字。随着全球经济一体化发展，中国和哈萨克斯坦等国的经济贸易往来日益频繁，哈萨克语的使用交流范围扩大。哈萨克语成为我国哈萨克族同其他国家地区的哈萨克族经济文化

基金资助：教育部语信司（MZ115-022）

作者简介：伊力亚尔·加尔木哈买提（1978—），男，新疆，助教，硕士 ylye1212@eyou.com
古丽拉·阿东别克，新疆大学，教授

交流的重要载体。然而由于双方文字形式的不同，需要首先翻译对方使用的书籍、文件、合同、书信等文字资料，最初的传统做法是人工翻译，这种翻译方式不适用于跨国民族交流的深入快速发展，而且耗费人力、物力、财力，阻碍了经济文化交流的程度。随着科学技术的快速发展，计算机翻译应用技术日益推广，哈萨克语文字形式之间的转换实现了基于 DOC 环境的机器翻译，大大简化了两种文字形式间的转换程序。相对 DOC 环境，Windows 以其方便快捷和友好的界面而倍受推崇，为了能够在 Windows 环境下哈萨克语两种文字形式间的计算机自动转换，我们提出了实现哈萨克语的两种文字间转换的应用程序。

7 系统设计

2.1 基本字母的转换设计（一一对应的字母设计）

由附录的字母表可以看出，哈萨克斯拉夫字母和哈萨克阿拉伯字母这两种文字的绝大多数字母是一一对应的，设计采用字母查表一一对应的方法。但在设计时，还要注意由于哈萨克斯拉夫字母有大小写的区分，而哈萨克阿拉伯字母则无大小写的区别，所以同一个哈萨克文阿拉伯字母对应两个哈萨克斯拉夫字母。

2.2 非对应字母的转化设计（一对多或多对一）

3.2.3 2.2.1 哈萨克斯拉夫文转换成哈萨克阿拉伯文

哈萨克斯拉夫字母中“Е”、“Ю”、“Я”这三个字母与哈萨克阿拉伯字母没有相应的一一对应的字母。如：“Е”→“ي”+“و”=“يو”；“Ю”→“ي”+“ؤ”=“يؤ”；“Я”→“ي”+“ا”=“يا”。在程序中采用当读到这三个字母中的任何一个，就在这一字母所在文章中的相应位置连续写入两个哈萨克文阿拉伯字母，写入的这两个字母是与这一个哈萨克斯拉夫字母对应的两个哈萨克阿拉伯字母。

...

```
if(indexofUniTable==33){
    strTmp+=strkazakhLettersList.GetAt(33);
    strTmp+=strkazakhLettersList.GetAt(19);}
else{
    if(indexofUniTable==34){
        strTmp+=strkazakhLettersList.GetAt(34);
        strTmp+=strkazakhLettersList.GetAt(0);}
    else{
        if(indexofUniTable==36){
            strTmp+=strkazakhLettersList.GetAt(17);
            strTmp+=strkazakhLettersList.GetAt(17);}
        else
            strTmp+=strkazakhLettersList.GetAt(indexofUniTable);
    }
}
```

...

在哈萨克斯拉夫文中，有用于俄文借词的两个字母，一个是软音符“Ь”，另一个是硬音符“Ъ”。虽然在哈萨克阿拉伯文中也存在软音符“ё”，这一软音符与哈萨克斯拉夫文的这两个音符并不对应。所以在程序设计中采用读到这两个哈萨克斯拉夫字母时，跳过不写这两个字母的方法。

3.2.4 2.2.2 哈萨克阿拉伯文转换成哈萨克斯拉夫文

在哈萨克斯拉夫文到哈萨克阿拉伯文的转化中由于哈萨克斯拉夫文采用的是怎么读就怎么写，所以不存在字

母的音变现象。但在哈萨克阿拉伯文中的一个单词中，如果有“گ” □ “ق” □ “ك” 三个字母在单词中的出现，使得本应读为“ف” □ “و” □ “و” □ “ى” 的四个字母写作对应的“ف”、“و” □ “و” □ “ى” 这四个字母。值得注意的是，这一规则在外来语借词中（如借自俄文、拉丁文、英文）则不适用。

例如：ف → و : گول، → ‘Г□Л’（花） □ → ف : انه، → ‘□НЕ’（在那里）

و → و : ورمەكشى، → ‘□РМЕКШЫ’（蜘蛛） □ → و : بىلگەك، → ‘□ЛГЕК’（书）

2.2.2.2 在哈萨克斯拉夫文中存在“И”、“Й”两个字母，而在哈萨克阿拉伯文中这两个字母都用同一个“ي”字母代替，这就产生了到底在什么时候用 ي → И □ 什么时候用 ي → Й 的问题。在解决这部分问题时，因为在哈萨克斯拉夫文中“И”字母在大多数情况都是在外来借词中使用，所以采用建立词库的方法，将所有带“И”字母的外来词制成一个表，然后再通过读到“И”时到表中查找的方法解决。

2.3 标点符号的转换

表 1 标点符号的转换

Table 1 Conversion of interpunction

	标点符号	标点符号	标点符号	标点符号
哈萨克斯拉夫文	,	;	?	.
Unicode 编码	002C	003B	003F	002E
哈萨克阿拉伯文	،	؛	؟	•
Unicode 编码	060C	061B	061F	0660

在哈萨克阿拉伯文中有如上表所列的四个标点符号采用的是阿拉伯文的标点符号记法，而哈萨克斯拉夫文中采用的是与英文字母相同的标点符号。在两种文字的相互转换中这四个标点符号互相对应。

2.4 显示的两种文字所采用的编码

现在国际上应用范围最广的编码，是由 ISO/IEC 10646-1: 2000/Unicode 4 中对应的字符集及其编码。哈萨克斯拉夫文采用 Unicode4 中对应的 04 区的斯拉夫文字母。同时由新疆维吾尔自治区科学技术厅、新疆维吾尔自治区信息化办公室提出以 GB/T2311 体系的字符编码标准，规定了信息交换用的维吾尔文、哈萨克文、柯尔克孜文的基本图形字符及其二进制编码的 16 进制表示标准，规定了维吾尔文、哈萨克文、柯尔克孜文（以下简称“维、哈、柯文”）信息交换用的字符集、基本字符集和扩展字符集在 ISO/IEC 10646-1: 2000/Unicode 4 中对应的字符集及其编码和位置。以上的这些标准为此程序的开发及应用提供了参考。

2.5 程序基本的操作界面

因为要实现哈萨克文在斯拉夫和阿拉伯这两种文字间的转换，所以采用了在窗体上部用 GridCtrl 控件来显示，发音相同的字母组成的一一对应的表格。在窗体下部则是两个多段的文本框来实现两种文字转换前后的显示。同时还有相应的打开及保存功能。下面是经过程序运行后的一部电视剧片断

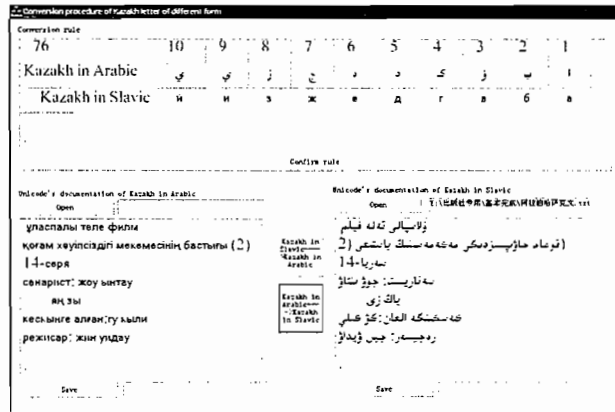


图 1 操作界面

Fig.1 operation interface

8 总结

本文主要介绍了哈萨克语的两种文字-斯拉夫字母和阿拉伯字母的相互转换。采用在 Unicode 代码系统中，斯拉夫字母和阿拉伯字母对应的方法。这一程序为在两种文字的不同文献中进行转换提供方便。

参考文献:

- [8] 阮宇. 基础教程系列Visual C++6.0基础教程[M], 北京: 清华大学出版社, 2005.7
- [9] Charles Wright. Visual C++程序员实用大全(精华版)[M], 北京: 中国水利水电出版社, 2005.5
- [10] 严蔚敏, 吴伟民.数据结构(C语言版)[M].北京: 清华大学出版社, 1996
- [11] 《哈萨克文正词法》, 新疆语委会, 1985
- [12] 古丽拉.阿东别克, 涂序彦, 哈萨克老文字与斯拉夫文间智能转换研究, 全国第九届计算机学会年会论文集, 1996. 5