PAN
Localization

Survey of Language Computing in Asia
2005

Sarmad Hussain
Nadir Durrani
Sana Gul

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences

www.nu.edu.pk

IDRC ✳ CRDI

Canadä

www.idrc.ca

Mongolian

Mongolian is an Altaic language spoken by approximately 2 million speakers in Mongolia [1]. Khalkha Mongolian (the most widely used dialect for Mongolian language) is the national language of Mongolia. It is also spoken in some parts of China, Afghanistan and Russia [1]. The figure given below shows the language family tree of Mongolian language.
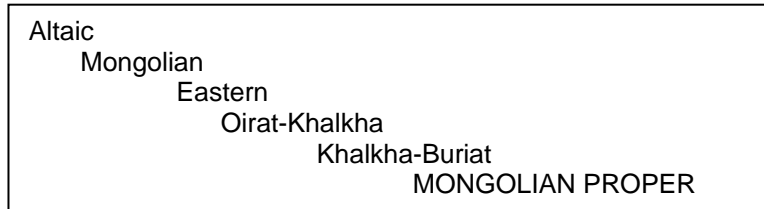
```
Altaic
      Mongolian
               Eastern
                     Oirat-Khalkha
                                Khalkha-Buriat
                                          MONGOLIAN PROPER
```

Figure 1: Language Family Tree for Mongolian [1]

Traditional Mongolian alphabet was derived from Uighur alphabet in the 12th century, which in turn evolved from Sogdian alphabet, which ultimately came from Aramaic. Mongolian has also been written in Chinese, Arabic and Tibetan alphabet. Eventually, due to Soviet Union, Latin alphabet was adopted in 1931 and eventually changed to Cyrillic alphabet, which is still widely used. In 1941, Mongolian government abolished Mongolian alphabet. However, since 1994 there is move to revive the Mongolian alphabet [2, 3]. Traditionally, Mongolian script is written from top to bottom, left to right [2, 20].

Character Set and Encoding

Mongolian, written in the Mongolian script uses Unicode code block 1800-18AF as the standard encoding [4]. Mongolian written in Cyrillic script uses the Unicode Cyrillic code page 0400-04FF as the standard encoding.

For the Mongolian character set based on Cyrillic script adopted, Microsoft code page 866 and Windows code page 1251 are adopted as national standards [5]. Some guidelines to convert between these encodings are available at [9]. Other encodings include ISO 8859-5 and KO18 [14].

The Mongolian national standards development body, National Center for Standardization and Metrology (MNCSM) has recently adopted 31 ISO standards on IT terminology as their Mongolian national standards. For Mongolian script, the ISO/IEC 10646 (Unicode) character encoding standard has been adopted as the national standard [4].

Fonts and Rendering

Cyrillic script fonts are available through vendors. Mongolian script fonts are fewer and not as readily available. Both are rendered well on various platforms (see [7, 13] for a list of (mostly) Cyrillic fonts based on various encoding schemes).

Microsoft Platform

Cyrillic script is supported in the Microsoft platform. Many Mongolian language Cyrillic fonts have also been developed by other research groups. Figure 2 shows few of the Cyrillic script fonts for Mongolian on Microsoft platform.

1. MgCourierReg

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

2. MgCourierBold

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

3. MgHelReg

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

4. MgHelItalic

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

5. MgHelBold

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

6. MgHelBoldItalic

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

7. MgKharkhorn

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

8. MgTimesRegular

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

9. MgTimesItalic

Ззурган дээрх хүүхэдийн царайнд сэргэлэн цовоо дүр терх нь тодрох мөчийг тусгасан байх шаардлагатай.

Figure 2: Unicode Cyrillic Script Fonts for Mongolian [7]

Mongolian script fonts are presently not provided by Microsoft, but Microsoft does provide some guidelines to develop these fonts [8]. Mongolian script Unicode based fonts have been developed by other vendors. Figure 3 shows a sample of Mongolian font rendered on MS platform.
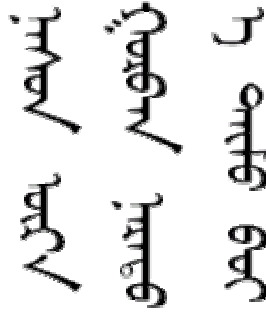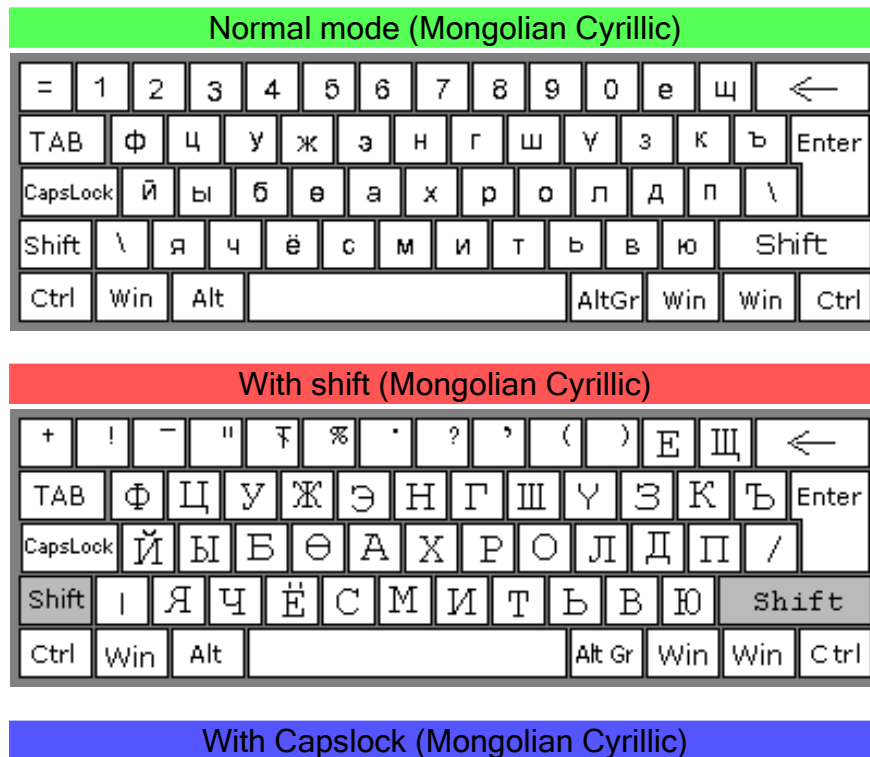
Figure 3: Unicode Mongolian Script Fonts [10]

Linux Platform

Mongolian Cyrillic script fonts are available and supported on Linux platform. However, there is no support for Mongolian script fonts. See [11, 12] for details.

Keyboard

Cyrillic keyboard standard is available. In addition, phonetic based and other layouts are also used, e.g. see [15, 16, 17]. These keyboards are available for all the different encoding used for Cyrillic. Keyboard standard for Mongolian script is not available. However, ad hoc layouts are available, e.g. the one shown in Figure 4.

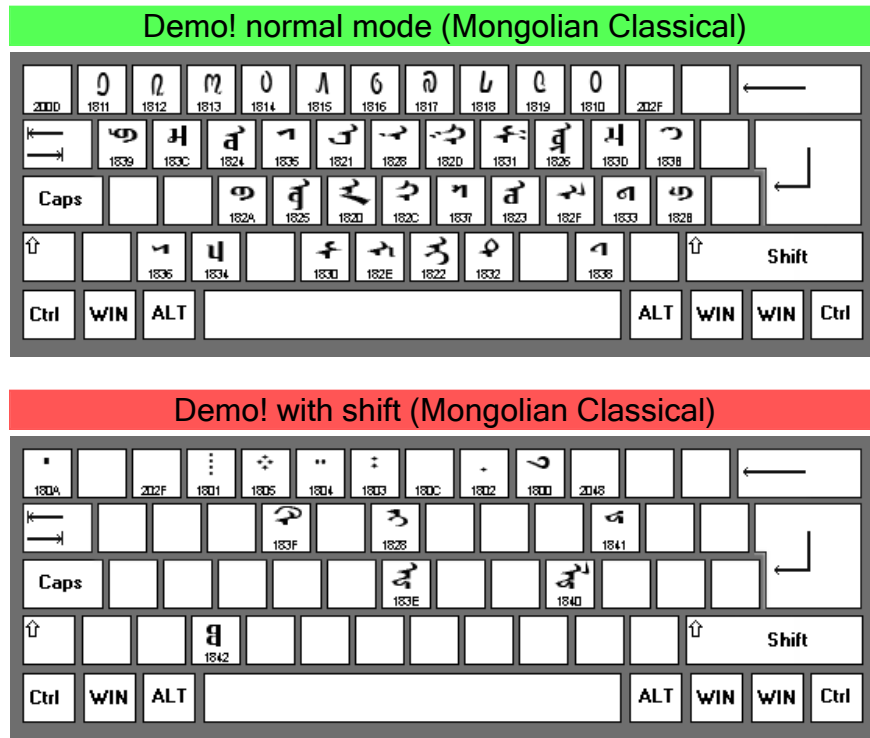| №	1	2	3	4	5	6	7	8	9	0	е	щ	←
TAB	Ф	Ц	У	Ж	Э	Н	Г	Ш	Ү	З	К	Ъ	Enter
CapsLock	Й	Ы	Б	Ө	А	Х	Р	О	Л	Д	П	/	
Shift	\	Я	Ч	Ё	С	М	И	Т	Ь	В	Ю	Shift	
Ctrl	Win	Alt					Alt Gr	Win	Win	Ctrl			

Figure 4: Cyrillic and Mongolian Script Keyboard Layouts for Mongolian [19]

Microsoft Platform

Microsoft Windows XP provides Cyrillic based in-built support for Mongolian keyboard. Once this keyboard layout is enabled it will facilitate Mongolian (Cyrillic) text input. Figure 5 below shows Mongolian keyboard layout (normal state and shift state) on Microsoft platform (also see [18]).

(a)

(b)

Figure 5: (a) Normal and (b) Shift State for Mongolian (Cyrillic) Keyboard on Microsoft Platform

Linux Platform

There are multiple keyboards available for different encodings for Cyrillic on Linux platform (e.g. see [17]). However Mongolian script keyboards are not available.

Collation

Collation for Cyrillic has been defined.   However, there is no standard for Mongolian script. Cyrillic is supported by Microsoft and Linux platforms.

Locale

Locale for Mongolian has not been defined in IBM ICU [21]. CLDR 1.3 includes a locale for Mongolian, which is partially defined. The locale for Mongolian is mn_MN.

Microsoft Platform

Microsoft Windows XP provides support for Mongolian (Cyrillic) locale.   If system locale is switched to Mongolian, changes may be monitored in the date, time, and currency symbol within all application of Microsoft Windows.   Figure 6 shows an example of setting user locale to Mongolian for Microsoft Windows XP. Mongolian script based locale is not available.
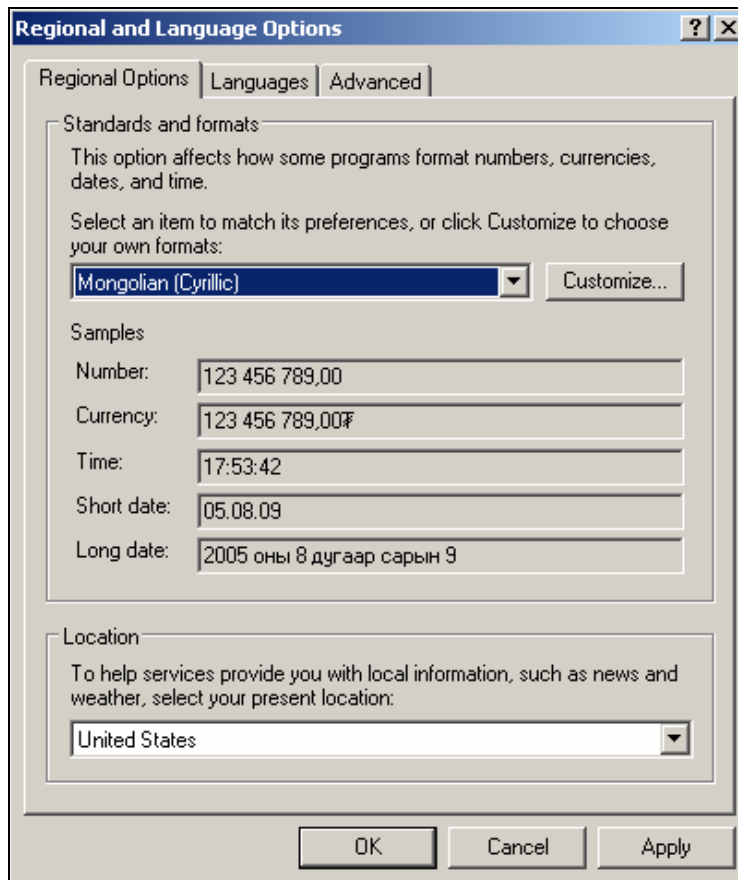
Figure 6: Mongolian (Cyrillic) Locale on Microsoft Platform

Linux Platform

Cyrillic based locale is also defined for Linux platform for Mongolian language, e.g. see [22].

Interface Terminology Translation

Mongolian interface is not available on Microsoft platform [23, 24]. Majority of localization work done on Linux platform has been done for Cyrillic. A complete Mongolian Linux distribution, Soyombo, with completely localized Mozilla has been developed by the OpenMN research Group for Cyrillic [12]. In addition the localization teams for Mongolian have also registered for translating GNOME 2.2 where about 99% of Developers.lib and 74% of Desktop strings have been translated [25]. Cyrillic glossary translation for KDE [26] and Open Office [27] in Mongolian is also under progress.

Status of Advanced Applications

There are a few Mongolian-English dictionaries available on the Internet. Some links are listed in [2]. Though most of them are in Cyrillic, some are also in Mongolian script [28]. However there is very limited work on Mongolian within Mongolia on other language technologies. Spell checkers are available [29]. Some work has been initiated on Mongolian text-to-speech system based on Cyrillic input [30]. Work is also being done on Cyrillic based OCR.

References

[1] http://www.ethnologue.com/show_language.asp?code=mvf
[2] http://www.omniglot.com/writing/mongolian.htm
[3] http://en.wikipedia.org/wiki/Mongolian_language
[4] http://www.unicode.org/charts/PDF/U1800.pdf
[5] Bayarmagnai, N. "Current Status of Information Technology Standardization and its Issues in Mongolia," in Proceedings of 13th AFIT, Myanmar, 1999. Available at http://www.cicc.or.jp/english/hyoujyunka/af13/13-11.html
[6] http://www.paratype.ru/default.asp?page=/library/languag?langCode=51
[7] http://www.salika.co.jp/emgfont.html
[8] http://www.microsoft.com/typography/otfntdev/mongolot/default.htm
[9] http://asuult.net/badaa/docs.php?p=trans_table
[10] http://www.travelphrases.info/gallery/fonts_Mongolian.html
[11] http://www.mnbsd.org/
[12] http://openmn.sourceforge.net/
[13] http://cgm.cs.mcgill.ca/~luc/mongolian.html
[14] http://ourworld.compuserve.com/homepages/PaulGor/fonts_e.htm
[15] http://ourworld.compuserve.com/homepages/PaulGor/screen_e.htm
[16] http://www.amherst.edu/it/software/languages/cyrillic/keyboard98.html
[17] http://asuult.net/badaa/pros.php?p=offline
[18] http://www.microsoft.com/globaldev/reference/keyboards.mspx
[19] http://asuult.net/badaa/docs.php?p=mongol_layout
[20] http://www.unicode.org/versions/Unicode4.0.0/ch12.pdf
[21] http://www-950.ibm.com/software/globalization/icu/demo/locales
[22] http://lists.debian.org/debian-devel/2003/02/msg01965.html
[23] http://www.microsoft.com/globaldev/DrIntl/faqs/lipfaq.mspx#EWB
[24] http://www.microsoft.com/office/editions/prodinfo/language/availability.mspx
[25] http://l10n-status.GNOME.org/GNOME-2.8/mn/index.html
[26] http://l10n.openoffice.org/languages.html#content
[27] http://i18n.KDE.org/teams/index.php?a=i&t=mn
[28] http://laurencio.webz.cz/mongolxel/classical/
[29] http://www.magicnet.mn/modules.php?name=News&file=print&sid=177
[30] http://www.apdip.net/projects/ictrnd/2005/l13-mn/