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

www.idrc.ca

Published by

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

Copyrights © International Development Research Center, Canada

Printed by Walayatsons, Pakistan

ISBN: 969-8961-00-3

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada, administered through the Centre for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences (NUCES), Pakistan.

Farsi (Persian)

Farsi belongs to Iranian branch of Indo-European language family. Persian or Western Farsi is spoken by about 22 million people residing in Iran. A close variant, Dari (Eastern Farsi), is spoken by seven million people in Afghanistan, Iran and Pakistan. Both varieties are also spoken in some other countries as well [1, 2]. Figure 1 shows the language family tree of Farsi.

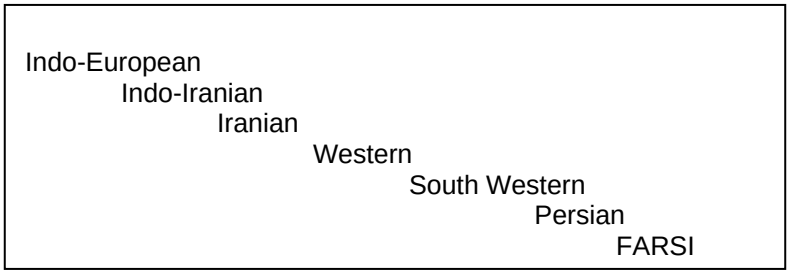


Figure 1: Language Tree Farsi of Language [1]

Farsi has been written with a number of different scripts, including Old Persian Cuneiform, Pahlavi, Aramaic, and Avestan. However after 642 AD Arabic script has been used for writing Farsi [2]. Nastalique style for Arabic script is used for writing Farsi.

Character Set and Encoding

Unicode Arabic script block from 0600-06FF is the standard character set encoding used for Farsi. A national standard based on relevant Unicode character subset within Arabic script block is also defined by Institute of Standards and Industrial Research in Iran (ISIRI) [3]. Earlier popular Farsi character set used for encoding was “Iran System”. The figure below shows this character set encoding.

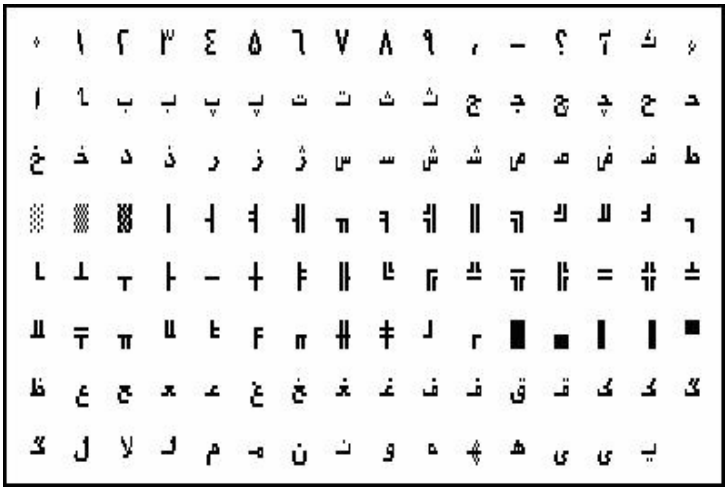


Figure 2: Farsi Character Set Encoding for “Iran System”

Fonts and Rendering

Microsoft Platform

Microsoft Windows fonts Tahoma and Microsoft Sans Serif can be used for typing Farsi text. In addition to this there are other Unicode Farsi fonts available, some of which have been shown in the figure below. All these fonts follow the Naskh style of Arabic script. No Nastalique style font is available by Microsoft. Nastalique Open Type fonts are available from other organizations, which can be used for Persian, e.g. Nafees Nastalique [4].

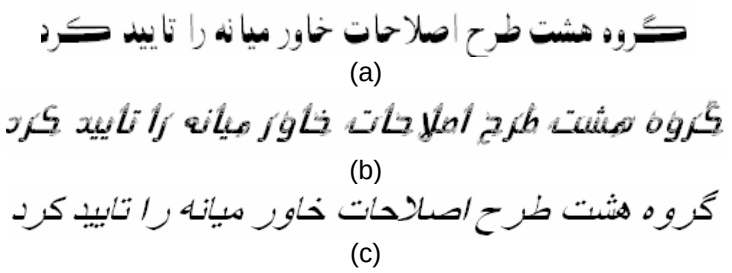


Figure 3: Farsi Text Written in Arash, Sorkhpust and Times New Roman Fonts

Linux Platform

Open Type support for Nastalique style is not available on Linux platform as shown in Figure 4, which shows three fonts rendered on GNOME.

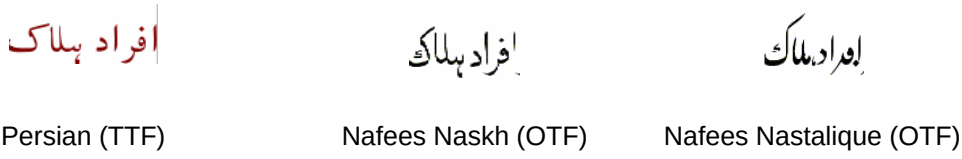


Figure 4: Farsi Fonts on GNOME

KDE only displays True Type fonts properly and does not display Open Type fonts, like Nafees Nastalique as shown in Figure 5.

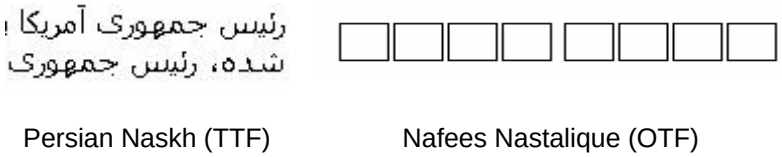


Figure 5: Farsi Fonts on KDE

Persian live CD Shabdix provides localized distribution of Persian on KDE 3.1.2. It has many built in Persian fonts but the rendering results are similar to those obtained on any KDE distribution [5].

Keyboard

ISIRI has published keyboard standard ISIRI 2901:1994 [6], shown in Figure 6.

Figure 6: Iranian Standard ISIRI 2901 for Keyboard Layout [6]

Microsoft Platform

Microsoft Windows XP provides a Farsi keyboard, slightly different from ISIRI 2901. This can be used with all Unicode compatible applications. Figure 7 given below shows Farsi on-screen keyboard layout on Microsoft platform (see [7]). ISIRI 2901 compatible keyboards for Microsoft platform are also available from other organizations, e.g. see [8].

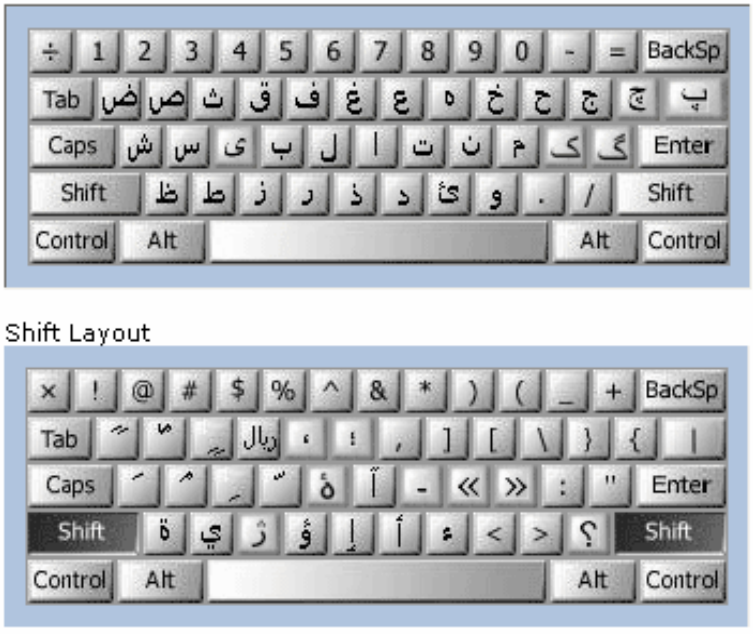


Figure 7: Microsoft Onscreen Keyboard Layout for Farsi [7]

Linux Platform

Red Hat 9 provides a Farsi keyboard layout which can be used both with KDE and GNOME. Keyboard support is part of input locale and runs at XWindows in Linux. Therefore, once enabled it works for all Unicode based technologies like Open Office and Mozilla. GNOME and KDE editors also support Farsi text. Farsi live distribution Shabdix also provides Farsi keyboard [5].

Collation

Collation sequence for Farsi has been defined for GNU C Library, but it has not been standardized.

Farsi collation is supported by Microsoft through Farsi Language Interface Pack (LIP) [9]. Farsi collation is also supported on Linux platform.

Locale

Locale for Persian (fa_IR) is defined in CLDR 1.3.

Microsoft Platform

Microsoft XP provides Farsi locale enabling appropriate changes, as shown in Figure 8.

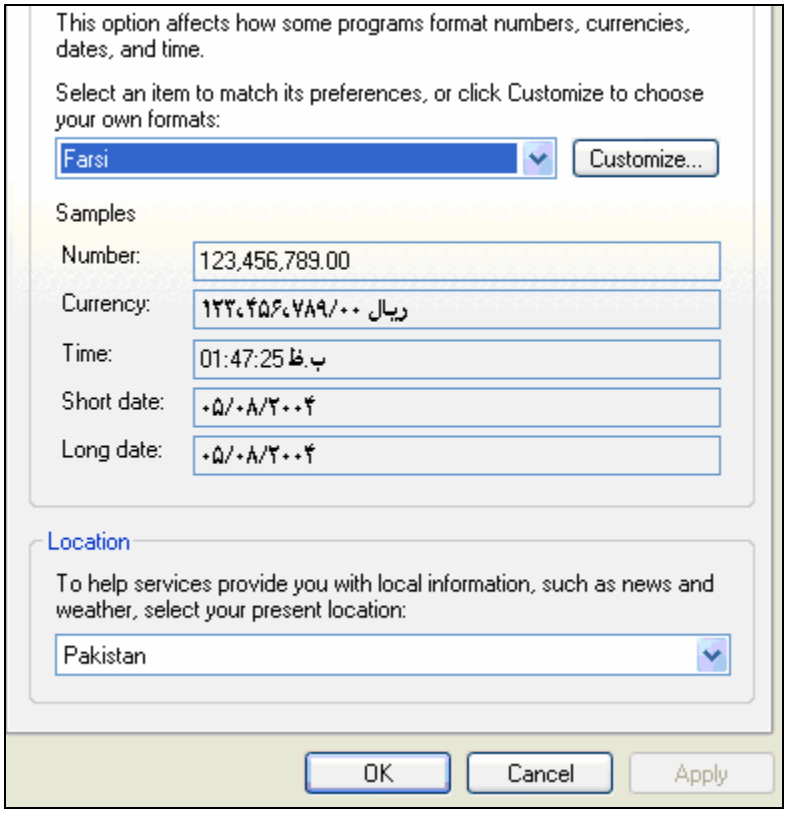


Figure 8: Farsi Locale in Windows XP

Linux Platform

Locale definition for Farsi has been defined in Red Hat Linux 9. Farsi locale is enabled by default in Shabdix. Figure 9 shows KDE module localized in Shabdix.

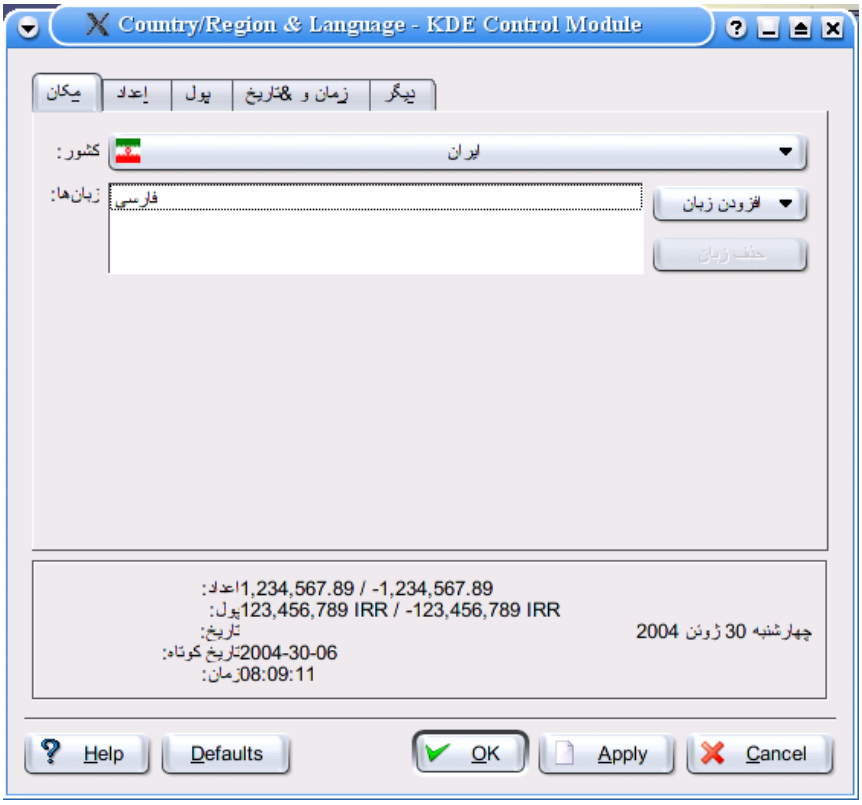


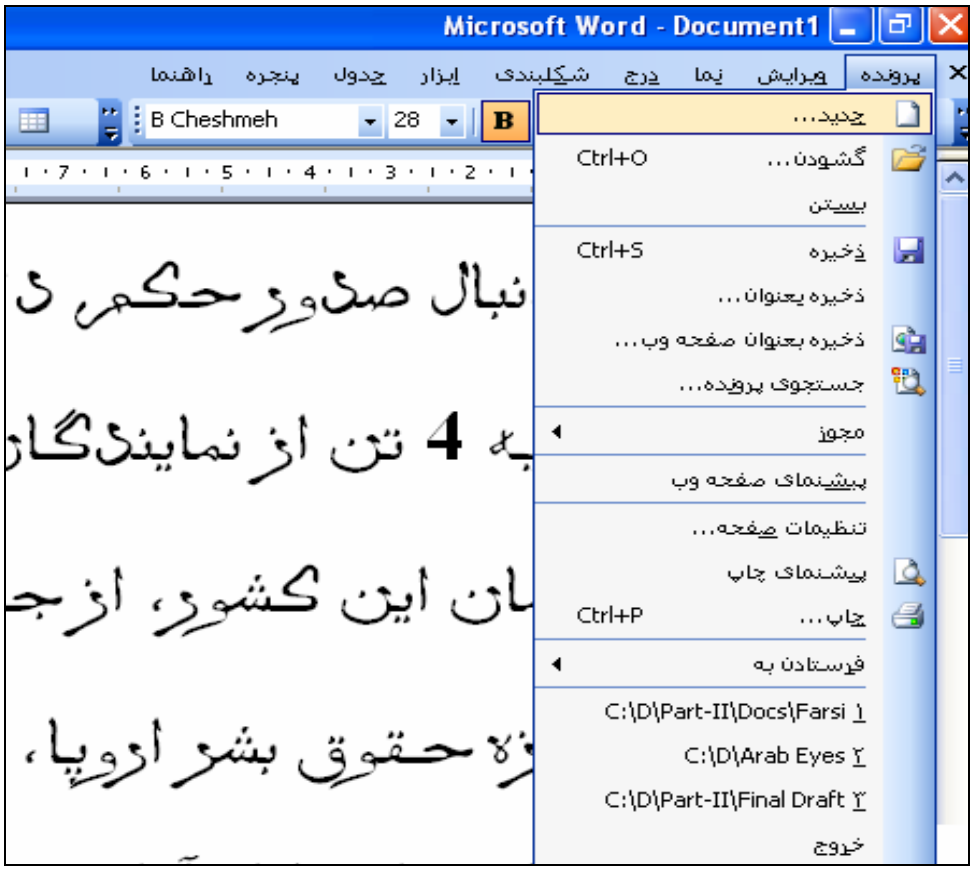
Figure 9: Farsi Date and Time Settings in Shabdix [4]

Interface Terminology Translation

Interface terminology for computers exists in Farsi, but has not been standardized.

Microsoft Platform

Microsoft provides complete Farsi interface through Farsi LIP (except Microsoft Office help files) for Windows 2000 and XP and Office 2003 [10]. Figure 10 shows the localized interface of Microsoft Word and Outlook in Farsi.



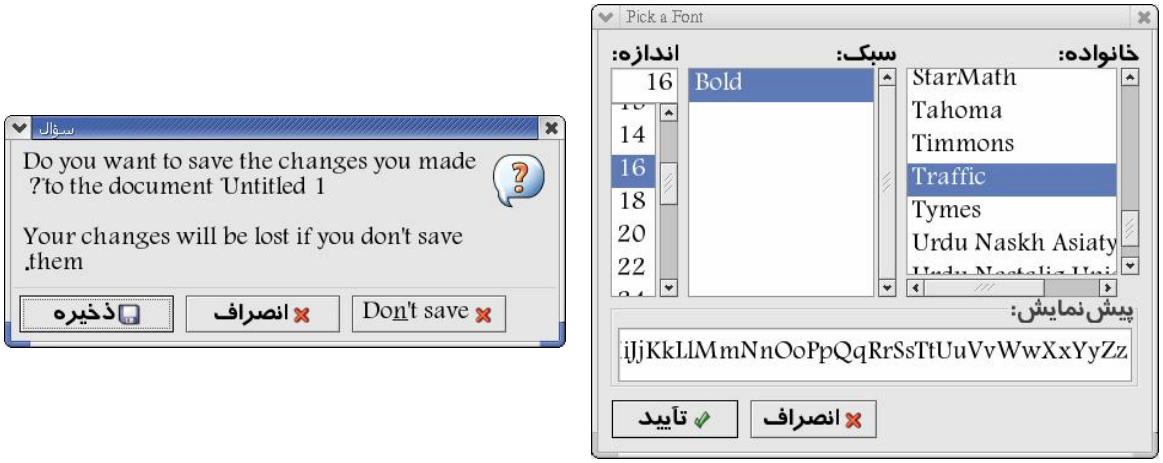
(a)

(b)

Figure 10: Localized Microsoft (a) Word, and (b) Outlook in Farsi

Linux Platform

Red Hat does not provide a localized Farsi interface for GNOME or KDE, but efforts are underway for the translation. GNOME 2.12 has been translated up to 45.89%, GNOME 2.10 is about 48% complete, nd KDE has been done up to 20%. Figure 11 below shows the Farsi translation in KDE on Red Hat distribution.



(a) (b)

(c)

Figure 11: Partially Localized Red Hat 9 (a) File Save Dialogue, (b) Font Dialogue Box, and (c) Start-Up Menus

Shabdix provides a partial Farsi graphical user interface. Only the base KDE desktop menu files, an editor and Konqueror web browser are fully localized. Figure 12 shows the screen shots of Shabdix distribution. Farsi interface, desktop menus, editor menus all are designed to move from right-to-left, which is conventional for Arabic script based languages. This release is based on KNOPPIX 3.4 and includes an updated and modified Farsi KDE [5].

(a)

(b)

Figure 12: (a) Start-Up Menu, and (b) Control Panel in Shabdix KDE

Status of Advanced Applications

Work is under progress to develop Farsi text-to-speech systems, lexicon, spell checker and thesaurus. However, these products are not commercially available.

References

[1] http://www.ethnologue.com/14/show_language.asp?code=PRS
[2] http://www.omniglot.com/writing/persian.htm
[3] "Information Technology – Persian Information Interchange and Display Mechanism, using Unicode." http://www.isiri.org/std/6219.htm
[4] http://www.crulp.org
[5] http://shabdix.berlios.de/
[6] "Keyboard Layout for Farsi." http://www.isiri.org/std/2901.htm
[7] http://www.microsoft.com/globaldev/keyboards/kbdfa.htm
[8] http://www.farsiweb.info/howto/win2keyb/
[9] http://www.microsoft.com/middleeast/arabicdev/farsi/wPaper.asp
[10] http://www.microsoft.com/middleeast/arabicdev/farsi/wPaper.aspx