



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

Canada

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.

Indonesian

Indonesian or Bhasha Indonesia is similar to Malay language. It is the lingua franca and official language of Indonesia [1, 3]. Indonesian is spoken by 30 million people in Indonesia and other countries [2]. Indonesian vocabulary constitutes many foreign words from Arabic, English, Portuguese, etc. The language family tree of Indonesian is shown below.



Figure 1: Language Family Tree for Indonesian [1]

Indonesian is written in Latin script, though sometimes Arabic script is also used [1, 2]. Latin script was first introduced during the period of Dutch colonization, and later after independence in 1930, the script was formally adopted for writing Indonesian [2].

Character Set and Encoding

Unicode code values for Latin script character set are also used for Indonesian. Eight-bit Latin standard ISO 8859 is also used.

Fonts and Rendering

Usual Latin fonts and rendering is used for Indonesian.

Keyboard

Usual Latin characters based keyboard layout is used for Indonesian.

Collation

English collation order is used for Indonesian, which is available on most platforms.

Locale

Standard Locale for Indonesian has been defined in IBM ICU (id_ID). This locale definition has localized settings [4]. CLDR 1.3 also includes a locale for Indonesian.

Microsoft Windows XP provides support for Indonesian locale, as shown in Figure 2 below. The definition is also available in the Language Interface Pack for Indonesian [5]. Locale is also being incorporated in the Linux distributions.

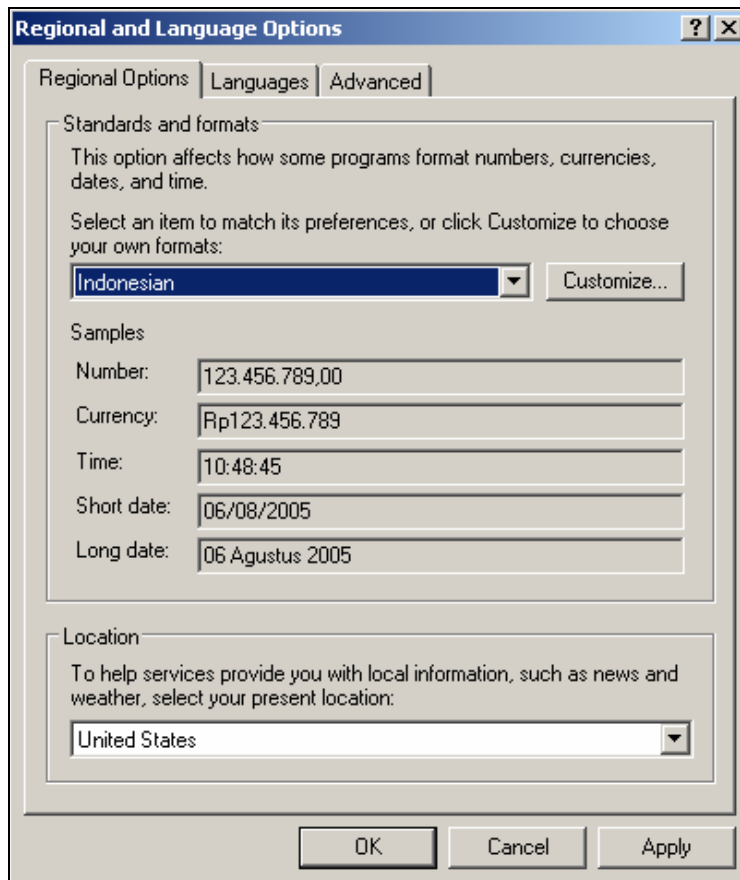


Figure 2: Indonesian Locale on Microsoft Platform

Interface Terminology Translation

Microsoft provides LIP for Indonesian which includes local language interface [5,6]. A complete Linux distribution in Indonesian is also available [11]. A team working on localizing GNOME in Indonesian language is currently active and GNOME glossary translation work has almost completed. Almost 87% of developers.lib and 84% of desktop strings have been translated [7]. In addition there are localized open source tools available through web portals in Indonesian (e.g. [8, 9, 10]).

The government of Indonesia has launched "Indonesia Go Open Source" program in 2004. This project is collaboration between Ministry of Research and Technology, Ministry of Communication and Information and the Indonesian Institute of Sciences. Through this project it is aimed to strengthen national information technology system and develop and freely disseminate OSS in Indonesian [12]. Following the objectives of the project it is believed that very soon a number of localized OSS applications will be available in Indonesian.

Status of Advanced Applications

Work is being done in language processing of Indonesian. There are many Indonesian and Indonesian-English dictionaries available online (e.g. see [13]). Spell checkers and morphological parsers for Indonesian are also available. There is very limited work on corpus [14, 15]. Indonesian text-to-speech system based on MBROLA was released in 2003 [16] and work has started on speech recognition [17] and machine translation with other languages.

References

- [1] http://www.ethnologue.com/show_language.asp?code=ind
- [2] <http://www.omniglot.com/writing/indonesian.htm>
- [3] http://en.wikipedia.org/wiki/Indonesian_language
- [4] http://www-950.ibm.com/software/globalization/icu/demo/locales/en/?_gu&d=en&l=id
- [5] <http://www.microsoft.com/downloads/details.aspx?displaylang=id&FamilyID=0db2e8f9-79c4-4625-a07a-0cc1b341be7c>
- [6] <http://www.microsoft.com/office/editions/prodinfo/language/availability.mspx>
- [7] <http://110n-status.GNOME.org/GNOME-2.8/id/index.html>
- [8] <http://opensource-indonesia.com/kioss.php/index.php>
- [9] <http://www.pandu.org/>
- [10] <http://www.gerbanglinux.com/mod.php?mod=katalog&op=NewLinks&menu=26>
- [11] <http://www.linux.or.id/>
- [12] <http://www.igos.web.id/english/english.htm>
- [13] <http://indonesian.dictionary.kamous.com/translator/reference.asp>
- [14] <http://torvald.aksis.uib.no/corpora/2004-1/0231.html>
- [15] Nazief, B. "Development of Computational Linguistic Research: A Challenge for Indonesia," available at <http://acl.ldc.upenn.edu/P/P00/P00-1075.pdf>
- [16] <http://lss.ee.itb.ac.id/~aa/indotts/>
- [17] Martin, T., Svendsen, T., Sridharan, S. (2003): "Cross-lingual pronunciation modeling for Indonesian speech recognition", In EUROSPEECH-2003, 3125-3128.