Indica, an Indic preprocessor for \TeX
A Sinhalese \TeX \ System

Yannis Haralambous
Centre d'Études et de Recherche sur le Traitement Automatique des Langues
Institut National des Langues et Civilisations Orientales, Paris.
Private address: 187, rue Nationale, 59800 Lille, France.
Yannis.Haralambous@univ-lille1.fr

Abstract

In this paper a two-fold project is described: the first part is a generalized
preprocessor for Indic scripts (scripts of languages currently spoken in India—
except Urdu—, Sanskrit and Tibetan), with several kinds of input (\TeX\ commands,
7-bit ASCII, CSX, ISO 10646/UNICODE) and \TeX\ output. This utility is written
in standard Flex (the GNU version of Lex), and hence can be painlessly compiled on
any platform. The same input methods are used for all Indic languages, so that
the user does not need to memorize different conventions and commands for each
one of them. Moreover, the switch from one language to another can be done by
use of user-defineable preprocessor directives.

The second part is a complete \TeX\ typesetting system for Sinhalese. The
design of the fonts is described, and META\font-related features, such as
metaness and optical correction, are discussed.

At the end of the paper, the reader can find tables showing the different input
methods for the four Indic scripts currently implemented in Indica: Devanagari,
Tamil, Malayalam, Sinhalese. The author hopes to complete the implementation
of Indic languages into Indica soon; the results will appear in a forthcoming paper.

This paper will be published in the next issue of TUG\boat.