Status of Linux Developments for Tamil

Tamil-Linux Developers Group
July 26,2000

The Tamil-Linux Developers Group is pleased to present this status report to the Tamilnadu IT Task Force, towards the 'Tamil and Linux' workshop to be held on July 28, 2000.

1. Introduction

Since 1997, few developers have been enthusiastically working on Linux, contributing to the growth and development of Tamil, through their various applications and software. These developers are part of the Tamil Standard Code for Information Interchange (TSCII) group and hence their contributions are presently TSCII related.

2. Past Work

Till date, the following tasks have been completed, tested and are found to work seamlessly in all major flavors of Linux like RedHat, Slackware, Mandrake and others.

2.1 Tamil Fonts

Various styles of Tamil fonts, freely available from the TSCII group have been converted to Unix style bitmapped and postscript fonts. This enabled Linux users to read Tamil web pages and documents using Netscape web browser. After the Government of Tamilnadu issued the TAM and TAB standards, more fonts were made for Linux.

For some of the Tamil news magazines on the web, whose fonts were not available for Unix, their fonts for other operating systems were converted to Linux. Also, alternative means that use font servers were identified for reading these web pages. Thus, a Linux user will be able to easily read most of the online Tamil news magazines.

2.2 Tamil Editor

Few years back, a powerful Tamil editor was developed using Tcl/Tk (version 8.0 and above) and this remains to be the single editor available for typing Tamil in Linux. Since Tcl/Tk is available for all operating systems, this editor can be used across all operating systems, which is highly advantageous. Some of the salient features of this editor are its options to save documents in text and rich-text formats, cut and paste operations using clipboard, icon toolbar and text formatting.

2.3 Tamil FAQ

A detailed FAQ on using Tamil on Linux and Unix platforms is now available. This document explains in easy steps how to install Tamil fonts, input/output documents. It is constantly updated on the recent developments made.

2.4 Console Fonts

Recently, a console font was developed which has been tested to work. This enables users to read Tamil pages from the console using text based web browsers like Lynx and w3m. Linux console uses Unicode in all its internal processing.   But the console tools do not have enough support for handling complex scripts like Tamil.

3. Present Work

The enthusiastic and consistent work by the developers often meet their goals and this forces them to set more new goals. The following section explains some of their present work.

3.1 Tamilinix

Realizing the need for information exchange between the developers, a group named 'Tamilinix' is in the process of being formed. Any Linux/Unix enthusiast will be able to join the group. The formation of this group will help to identify the talents and interests of developers throughout the world and thus create a wide resource pool for Tamil development work on Linux.

3.2 Tamil Locale

From the man pages of locale (7):
> "A locale is a set of language and cultural rules. These cover aspects such as language for messages, character sets, lexigraphic conventions, etc. A program needs to be able to determine its locale and act accordingly to be portable to different cultures."

Besides the available console font, a locale has been setup for Tamil. The work still remaining in this is that the files need to be extracted and a document be made explaining how to use. Tests have shown that this currently displays the calendar/date commands using Tamil. This is still in development and is not yet released.

3.3 Keyboard Driver for KDE

KDE International Keyboard (kikbd) is a project whose aim is to enable the input of other symbols/characters, other than the default Latin symbols available under X11.

A pseudo-phonetic keyboard for KDE based on kikbd is currently under development. This input method is slightly cumbersome due to the heavy use of Alt and Shift key combinations to type Tamil characters. But this driver is compatible with the old phonetic style keyboard.

In this regard, a HOWTO for KDE Tamil translation has been prepared. Presently, a KDE keyboard driver based on romanized keyboard is available.

3.4 Tamil RPM for KDE

Developers at Mandrake have put forth a RPM for KDE-Tamil files. But, this package contains no fonts but has only the translations. This RPM should work on RedHat systems also, although it is not yet tested. Most of the KDE efforts are based on TSCII. The KDE team is in the process of switching over to Unicode.

3.5 UTF8-TSCII Codec

An UTF8-TSCII Codec is available at the website by Hans Peter Bieker. Based on comments from its source code, this seems to have been based on Dr. K. Kalyanasundaram's conversion table. This needs to be validated since the original conversion table initially had few problems.

4. Future Work

The growing popularity of Linux, ease of use and free availability are some of the key features, which hint that Linux is emerging as a major operating system of the future. Hence, any development effort spent on the applications for Tamil will certainly benefit the future, large user community.

4.1 Tamil support for Pango

Pango () is a high quality text-processing engine based on Unicode, aimed at rendering internationalized text. It is a general-purpose engine, which for time being is supported in GTK+. The current development version of GTK+ 1.3.1 already includes Pango.

A Tamil module for Pango, which renders Unicode using a TSCII font, has been developed. This can be modified to use a monolingual font for better quality text. This module is still under development and so requires more work.

4.2 Optical Character Recognition (OCR)

Few years back, efforts were taken to adapt a freely available OCR software for English to Tamil. But, unfortunately this work did not continue for long (lack of font standards being one reason). More work is required in this area, as OCR is a highly valuable tool in digitizing Tamil documents, manuscripts and hence will be of immense help in building an electronic Tamil library.

4.3 Others

Following are the recommended tasks that should be given more attention in the near future.
- Enable Linux console tools to handle Tamil and other Indic scripts (http://lct.sourceforge.net).
- Create a console keyboard driver.
- Enable Yudit (Unicode editor) to handle Tamil.
- Create a generic Indic engine for Pango to handle multiple fonts..
- Develop XIM based keyboard drivers for Tamil99 and other keyboards
- Translate GNOME, KDE and other packages

5. Recommendations

Many of the Linux based projects are now switching over to Unicode. So any step taken now, should take Unicode also into account. For this, an agreement is to be made on how to handle the Unicode Tamil segment. Sivaraj had suggested the INFITT Unicode working group to propose a Ligature segment for Tamil. About 117 ligatures are needed, besides those preset in the current Unicode Tamil segment, to render all Tamil characters properly. If this does not materialize, a decision can be made on using a common segment in private use area for this purpose, after discussing with other Linux teams that use the private use area. H. Peter Anvin (hpa@zytor.com), of the Linux Kernel team used to maintain an allocation for this segment in the kernel source. But this file seems to be very out dated.

It is urged that a forum be setup to bring all these efforts under one umbrella. While developers abroad are in touch with each other, those in Tamilnadu are not known due to lack of communication. Hence, setting

up this forum will present an opportunity to interact among developers and so promote Tamil development on Linux.

6. References

TSCII Homepage
> http://www.tamil.net/tscii

TSCII Tools
> http://www.tamil.net/tscii/tools.html

Tamil-Linux and other FAQs
> http://www.tamil.net/tscii/faq.html

HOWTO for KDE Tamil translation
> http://tamil.htmlplanet.com

KDE Translation Project for Tamil
> http://members.home.net/gomathiss/kde.html

Tamil-Linux Developers Group

D. Sivaraj sivaraj@tamil.net
Chinnasamy Nagarajan nagu@tinet.ie
Anbumani Subramanian anbumani@webname.com
R.R.Subramanian rrs.manian@wipro.com
Thuraiappah Vaseeharan vasee@ee.umanitoba.ca
Sivakumar Shanmugasundaram sshanmu@yahoo.com