# Thailand FOSS Localization Report

Theppitak Karoonboonyanan

Localization Workshop, Phenom Penh, Cambodia
25 October 2005

## Reporter's Backgrounds

First of all, the reporter would like to declare his backgrounds, so the reader can expect which aspects of the report should be limited.

- Author of `th_TH` locale for GNU C library
- XFree86 contributor (before X.Org 6.7.0 fork back)
- GTK+/Pango contributor
- Thai GNOME translation team coordinator
- Thai Linux Working Group (TLWG) volunteer
- Former Linux TLE team member
- *Limited participation with Qt, None with KDE*
- *Limited participation with Mozilla*
- *No participation with OpenOffice.org and OpenI18N so far*
- *No participation with SIPA*

## 1 Technical Works

In addition to text input/output methods, Thai language also needs word break algorithms for proper line wrapping. All these tasks can be shared among different desktops at lower levels, but must be done separately in the toolkit level and above. However, Thai developers of both KDE and GNOME camps try to share implementations as much as possible.

### 1.1 Input Method

X Input Method (XIM) is the common input method available to both KDE and GNOME. And it is still a best choice for Thai input on both desktops nowadays, although they benefit from XIM with different qualities.

IIIMF from OpenI18N also provides `thai-le` input method from Sun. But it is still rarely touched by local developers, however.

- XIM
  - Original X11R6: Thai XIM in Xlib (broken, though)
  - Fixed X11R6 Xlib source (via XFree86, now in X.Org)
  - Added Read/Write **StringConversionCallback** support (i.e. context-sensitive sequence check/correction)
- GTK+
  - Requested for context-sensitive IM API in GTK+ 1.3 (pre-2.0)
  - Implemented retrieve/delete-surrounding signals in im-xim module
  - Requested/Fixed the signal callbacks in many GTK+ apps (e.g. gnumeric, abiword, eel (nautilus), gal (evolution))
  - Proposed Thai-Lao GTK+ IM Module in gtk-im-extras
- Qt 3
  - Just relies on limited XIM support (without access to surrounding text)
  - Context-insensitive sequence check
  - No correction, of course
- IIIMF
  - thai-le from Sun
  - No context sensitivity yet

## 1.2 Rendering

X Output Method (XOM) seems to arrive too late, when toolkits already implement text rendering engines of their own. Thai text rendering is supported on KDE and GNOME with different qualities (again).

- Pango
  - WTT-based clustering
  - Substitution-based shaping for legacy Win & Mac fonts
  - OpenType-based glyph positioning
- Mozilla
  - Pango-lite CTL support
  - Full pango support build option
- Qt 3
  - Unicode-based class for glyph positioning
  - Poor-man typography
  - Bug-prone, Thai devs tired of fixing every new versions → dismissed, time to leave for Qt 4

## 1.3 Word Break

Word break has been a hot exercise for Thai developers to gain as high accuracy as possible. There are many implementations from academic and research institutes. Some of them are FOSS and are picked for plugging in GNU/Linux desktops.

- Utilities & Libraries
  - cttex: Thai word break filter for LaTeX, HTML
  - swath: Thai word break filter for LaTeX, HTML, RTF
  - libthai: basic routines for Thai language processing, with cttex-based word break functionality
  - ICU: now supports Thai word break (from Mozilla Thai project)
- Pango
  - `pango-libthai`: libthai-based language engine for word breaking
- Qt 3
  - Dynamically load `libthai.so` for word breaking

## 1.4 Cultural Conventions

Early works on Thai cultural conventions were just simple translation of months and days of week. But the first complete one started from a different angle: string collation. Then, other categories were not difficult to add.

- GNU C library: all ISO/IEC 14652 locale categories supported
- CLDR: being audited by Thai OO.o hackers, and soon by Thai GNOME hackers

## 1.5 OpenI18N-related Tasks

As the reporter was asked to classify tasks into OpenI18N-related and those that aren't, this section is just to say:

- No known activities further than experimenting yet
- Reasons, maybe:
  - Current solutions still work
  - thai-le IIIMF from Sun is not so bad
  - Lack of, or too complicated, documentation
  - People are just too busy with current jobs
- CLDR is catching interests, anyway

## 1.6 Other Applications

Some tools outside end-user desktops are also used by power users. And they are also localized for Thai support.

- ThaiLaTeX: babel-based Thai support for LaTeX
- Xpdf: with TIS-620 support in `pdftotext` utility
- console-tools: with locally-distributed console fonts and keymaps

## 1.7 Human Resources

Early GNU/Linux localization tasks were mostly done by Thai students in Japan and NECTEC researchers. These people together founded the Thai Linux Working Group (TLWG) as a community where major parts of Thai FOSS developments happenned.

NECTEC also continues its missions by maintaining a GNU/Linux distribution named "Linux TLE" (Thai Language Extension), in which Thai developers' works are collected and delivered to end-users. And NECTEC researchers still contribute development works to the distribution as well.

Recently, another effort from government has been formed as a project in the new agency named "Software Industry Promotion Agency" (SIPA). FOSS promotion is its main mission, and necessary developments to achieve that goal are funded.

- TLWG (Thai Linux Working Group)
  - Public community for user supports and developers' meeting place
  - $\approx$ 10 active developers working on:
    - Patching upstream GNOME, KDE sources
    - Joining message translation teams
    - `libthai` development
    - Fonts development
    - ThaiLaTeX development
    - Debian, Gentoo packages maintenance
- NECTEC (National Electronics and Computer Technology Center)
  - An R&D agency under M. of Science, Technology and Environment
  - With an open source department
  - $\approx$ 10 hired researchers working on:
    - Linux TLE distribution
    - GNOME, OpenOffice.org message translation
    - OpenOffice.org hacking
    - Fonts development
    - Miscelleneous promotion activities
- SIPA (Software Industry Promotion Agency)

  - A promotion agency under M. of ICT
  - With an open source project
  - Hired developers working on:
    * OpenOffice.org and Mozilla hacking
    * FOSS-for-Windows distribution CD
  - Subsidizing companies working on:
    * Debian-based GNU/Linux distribution for servers
    * Fonts development

# 2 Translation

KDE was the very first project known for its complete Thai translation, while Thai GNOME developers focused on language support infrastructure first, and translation later. The result was that KDE gained user awareness very quickly, supplemented by its rich features. However, the effort later dropped suddenly, when the core members got busy with something else.

OpenOffice.org is now another magnet project that has gained much interests, funding, and human resources (even from other projects).

## 2.1 KDE

- Totally done by unpaid volunteers
- Started around 1998
- Apparently defunct over a year
- Current status (2005-08-30)
  - stable = 24.40%
  - head = 17.96%

## 2.2 GNOME

- History
  - Initiated by Paisa Seeluangsawat, an individual volunteer, since 2.4
  - Contributed by Burapha U. students and NECTEC stuffs in 2.6
  - More volunteers and paid (by NECTEC, then by SIPA) translators in later versions
  - Substantial parts are maintained by unpaid participants, though
  - Current status (2005-09-07)
    * New supported language in GNOME 2.12
- Working process
  - A few GNOME CVS committers accept and verify PO files before committing

  - Tools to ensure consistency (facilitated by NECTEC)
    * Mailing list discussion
    * Glossary web (http://opentle.org/ossglossary)
    * Wiki page for recording discussion summaries and for collecting translations in organized sets (such as location names, colors, etc.) (http://l10n.opentle.org)

- Translated messages
  - Initial translators translate user-visible messages
  - Most packages are pursued by coordinating staffs for deeper messages, such as error messages and GConf schemas
  - Policy: try to achieve 100% for every translated package
  - Don't wait for string freeze. Keep catching new messages along the development phases.

### 2.3 OpenOffice.org

- Formerly two efforts
  - Algorithms Co.: Pladao Office
  - NECTEC: OfficeTLE
- Both merged upstream
- OpenOffice.org hacking now done by SIPA and NECTEC
- Translation by NECTEC (in progress)

## 3 Documents

- Currently no program document translation
- Some philosophical articles are already translated, e.g.
  - How To Become A Hacker
  - The Cathedral and the Bazaar
  - Homesteading the Noosphere

## 4 Fonts

Fonts are being developed by many individual groups. Known examples include NECTEC, TLWG, and SIPA.

### 4.1 NECTEC

- National Fonts
  - Initiated by NECTEC, with experts' contributions
  - Publishing: Recommendations for Thai font creation
  - 3 standard, free families (2 serif, 1 san serif)
- UI Font
  - Font with specialized metrics for on-screen GUI
  - 1 free family

### 4.2 TLWG

- National fonts maintenance
  - 2 out of 3 are maintained (1 dropped for license issue)
  - Fontforge as development tool
  - Spline clean-ups
  - Type 1 hints
  - OpenType tables
  - (Todo) Kerning pairs
- In-house created fonts
  - 3 families (1 Monospace, 1 Typewriter, 1 Handwriting)

### 4.3 SIPA

- Thai glyphs in Bitstream Vera fonts
- In progress

## 5 Testing

Most testings are community-based. So, it is just done as per typical FOSS models.

- Normal tests by individual developers
- Peer review among developers
- Bug reports from local distros
- Bug reports in the community

## 6 Packaging & Distributions

Since this report covers many groups, with different purposes and responsibilities, their packaging and distribution methods should be discussed separately.

### 6.1 TLWG

- Mainly working upstream, and develop own sources
- Distributed in many ways
  - Published patches in developer forum & upstream BTS
  - Source tarballs from project pages (http://linux.thai.net/)
  - Local APT archive for Debian (http://linux.thai.net/apt)
  - Official packages for Debian and Gentoo
  - Packaged by local distributions
- Distributed free of charge

### 6.2 Linux TLE

- Packaged and Distributed by NECTEC
- Formerly based on Mandrake, then Red Hat, Fedora
- The lastest version now Ubuntu-based
- Try to work with upstream sources (rather than base distros)
- Distributed free of charge

### 6.3 Others

- Burapha U.: Burapha Linux (Slackware-based)
- SIPA: Suriyan (Debian-based)
- Many others used to exist, now seem to fade out

## 7 Maintenance

TLWG, NECTEC and SIPA work on different sources. But a common goal the three groups try to achieve is to get the works incorporated in upstream source. When some patches miss the schedule of upstream projects, they are mostly deposited in local distributions, to let Thai users get the fixes or features and may provide some feedbacks. But being in local distributions is always supposed to be temporary, until upstream projects accept them.

That said, the three groups have different management natures.

### 7.1 Version Updating

- TLWG
  - Upstream CVS
  - Own CVS
  - Own official package sources (Debian, Gentoo)

- Linux TLE (NECTEC)
  - Base distribution sources upon major releases
  - In-house packages from TLWG sources
- OpenOffice.org (NECTEC, SIPA)
  - Upstream CVS
- Firefox (TLWG, NECTEC, SIPA)
  - Patched upon new upstream/distro releases

### 7.2 Special Intentions

- TLWG
  - As of individuals' interests
  - Mostly focusing on Thai supports in upstream sources
- Linux TLE (NECTEC)
  - Targets desktop users
- Suriyan-Chantra (SIPA)
  - New strategy to promote FOSS country-wide
  - GNU/Linux server and Windows workstations

### 7.3 Business Model

- TLWG
  - Server, bandwidth and domain name are donated
  - Volunteers are self-supported
  - FOSS hacking as personal hobbies
  - No enough interests in doing FOSS as business
- NECTEC
  - Government funded
- SIPA
  - Government funded

## 8 Migration

Some organizations have been active in migrating to FOSS. Two examples given here are two individual efforts. One is for the organization itself, the other is country-wide.

### 8.1 Electricity Generation Authority of Thailand (EGAT)

- Self-assembled PC, with GNU/Linux as alternative OS
- Slackware and FreeBSD as servers for long
- Pioneer organization in migration to GNU/Linux in desktop side
- Training courses instructed by NECTEC staffs
- In-house web services with LAMP

### 8.2 SIPA

- Chantra: FOSS-for-Windows CD
- Suriyan: GNU/Linux distro for server, based on Debian

## 9 Conclusions

- Thai input/output methods are mostly complete on GNOME, and sufficient on KDE
- Thai support in OpenOffice.org is mostly complete in local versions, and is being incorporated in the next upstream version
- Thai L10N are mostly hacker activities, many get direct association to upstream projects
- Governmental agencies play the major roles in creating public awareness and development facilitation
- No obvious business model for L10N yet, apart from governmental subsidiary (Most FOSS businesses are server-side service providing and system integration)