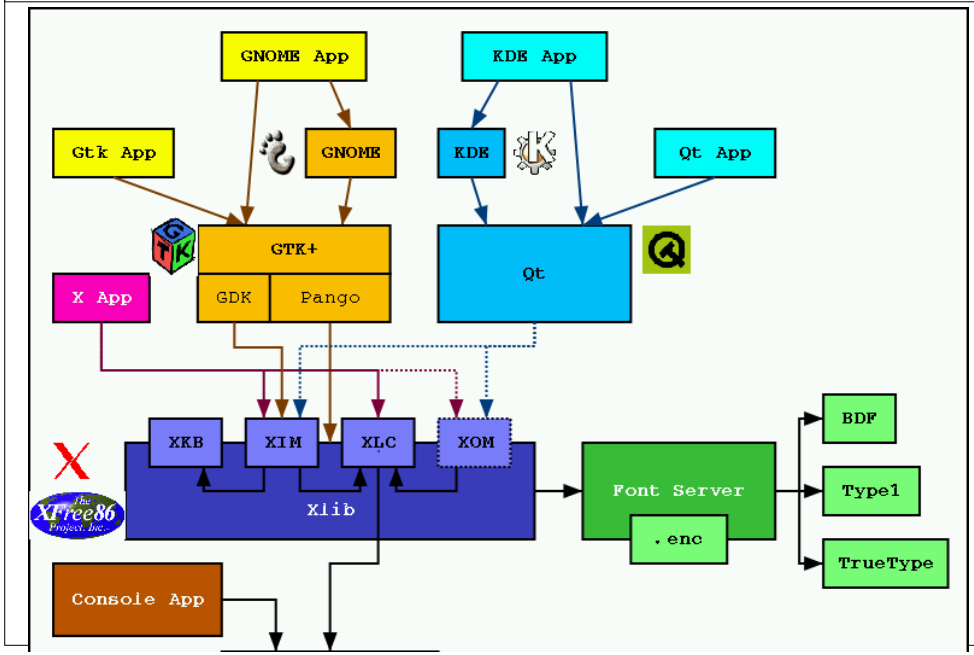
Status of Thai Support in GNU/Linux/X

Theppitak Karoonboonyanan
Thai Linux Working Group (TLWG)
thep@links.nectec.or.th
http://www.links.nectec.or.th/~thep/

Overview of the I18N Tasks



Thai Support Concerns

Contents

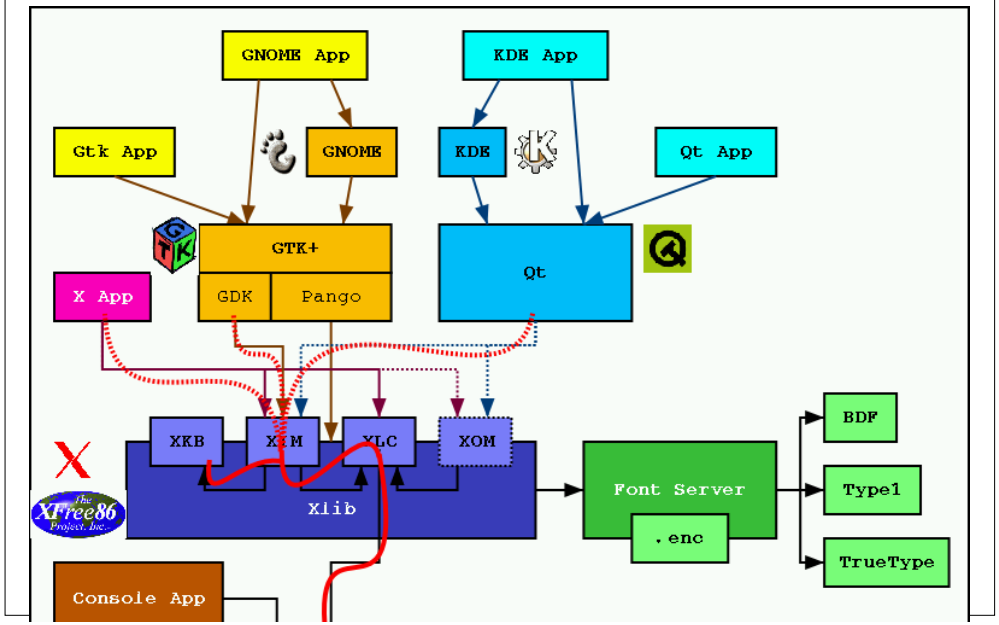
- Overview of the I18N Tasks
- General Cultural Conventions
- Input Method
  - Input Sequence Check/Correction
  - Future Works
- Output Method
  - Font Sets
  - Fonts
  - Rendering
  - Typesetting
  - Word Breaking

General Cultural Conventions

GNU C Library

- The Thai Locale Projects
  - http://www.links.nectec.or.th/~thep/tsort.html : Thai Sorting Algorithms
  - http://software.thai.net/locale/ : The Thai Locale Project (based on ISO/IEC 14652)
  - http://www.links.nectec.or.th/~thep/blsort.html : Thai-English Bilingual Sorting (**`LC_COLLATE`** special group)
  - http://www.links.nectec.or.th/~thep/th-locale/ : Thai locale development (POSIX-based)
- glibc 2.1.1
  - **`TIS-620`** charmap/repertoiremap
  - **`th_TH.TIS-620`** POSIX Locale (all categories)
- glibc 2.2.2
  - Buddhist Era in **`LC_TIME`**
  - (experimental) ISO/IEC 14652 locale categories

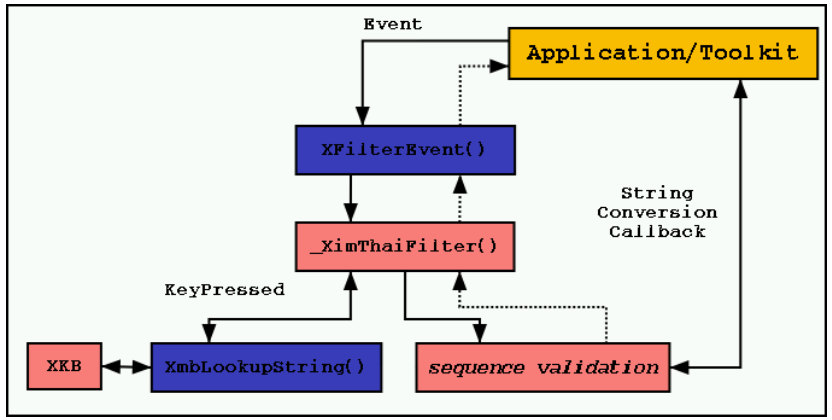
Input Method



Input Method

Input Sequence Check/Correction (cont.)

- Thai XIM structure



Input Method

Input Sequence Check/Correction

- X11R6
  - Library-model Thai XIM based on WTT 2.0
- X11R6.1
  - Thai XKB map with Latin hack
  - XIM defect in translating special keys
  - Workaround with **`LC_CTYPE`** disabling or by setting

    `XMODIFIERS="@im=none"`

Input Method

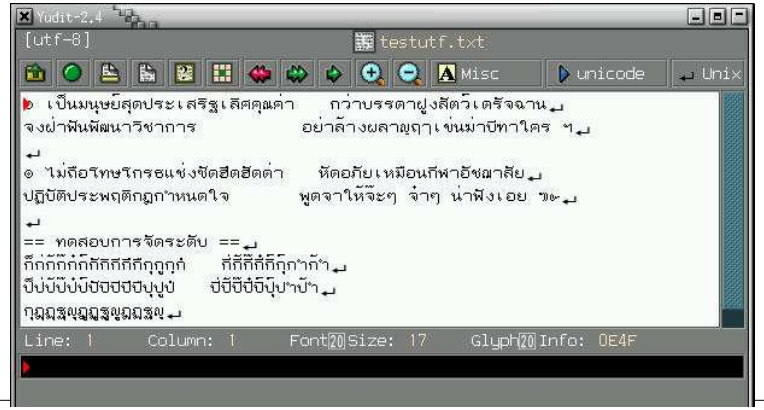
Input Sequence Check/Correction (cont.)

- XFree86 4.0.1d
  - New Thai XKB map in accordance with **`<X11/keysymdefs.h>`**
  - Thai XIM did not translate keys properly
  - Workaround with European XIM by creating a blank **`th_TH/Compose`** file
- XFree86 4.0.1g
  - Thai XIM keymap translation fixed
- XFree86 4.0.3
  - More XIM bug fix
- XFree86 4.0.99.1 (branch for 4.1.0)
  - XIM bug fix (same as 4.0.3)
  - Context-sensitive re-conversion introduced
- XFree86 4.0.99.2
  - The **`th_TH/Compose`** workaround removed

Input Method

Unicode Inputting

- xterm + **`th_TH.UTF-8`** locale
  - Local encoding (TIS-620) automatically converted to UTF-8
- yudit
  - Multilingual Unicode editor (Thai input method out of the box)



Input Method

Future Works (cont.)

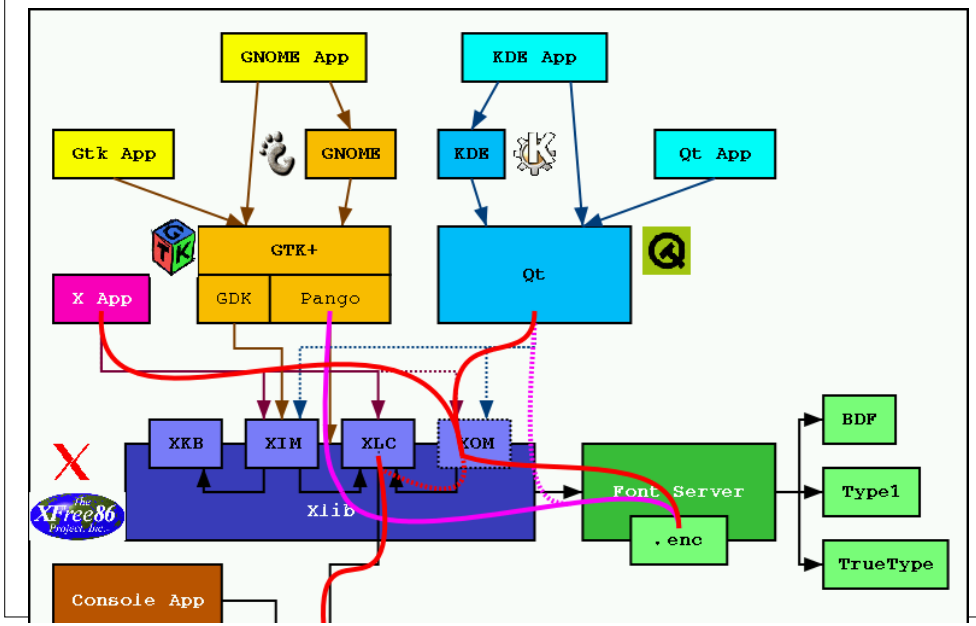
- Advanced Feature
  - Thai input sequence correction by means of the **`XIMStringConversionSubstitution`** operation of the callback
- Problems
  - Imprecise definition in the Xlib Programming Manual

Input Method

Future Works

- Concerns
  - Thai input sequence check is context-sensitive
  - Requires **`XNStringConversionCallback`** support by toolkits and applications
  - Minimal requirement: **`XIMStringConversionRetrieval`** operation with **`XIMBackwardChar`** direction
  - Currently falls back by buffering one previous character in the XIC, which will be outdated by cursor movements and backspaces
  - xiterm+thai 1.04pre2 as a first one to support the callback
  - GTK+
    - XIM already supported by GDK
    - Has a positive response in supporting the callback (after GTK+ 2.0)
    - Cell-based cursor movement (New)
  - Qt
    - XIM already supported in version 2.3.0
    - Keyboard focusing not complete, Thai XIM not active, however :-(

Output Method



Output Method

Font Set

- X11R6
  - **th_TH.TACTIS** locale
  - **TIS620.2533-1** encoding
- New Font Set Convention
  - **tis620-0** : Plain TIS-620
  - **tis620-1** : MacIntosh Extension to TIS-620
  - **tis620-2** : Windows Extension to TIS-620

Output Method

Font Set (cont.)

- GNOME 1.4
  - GTK+ 1.4
    - (GDK) Based on X
  - Gnome Print 1.4
    - Appears to have Unicode support, but no Thai fonts incorporated yet
- GNOME 2.0
  - GTK+ 2.0
    - (Pango) **tis620-0/1/2, tis620.2529-1, tis620.2533-0** recognized
- KDE 2
  - Qt 2.2.3
    - **tis620-0, iso8859-11** recognized

Output Method

Font Set (cont.)

- XFree86 4.0.99.1
  - **th_TH.TIS620** locale
  - **TIS620-0** encoding
- XFree86 4.0.99.1 'freetype' module & font server
  - Glyph map for **iso8859-11, tis620-0, tis620.2529-1, tis620.2533-1, tis620.2533-0** encodings
- XFree86 4.1.0
  - **tis620-2** is under submission (New)
  - **tis620-1** left to work out

Output Method

Available Thai Fonts (8 bit)

- TIS620 BDF fonts
  - Manop : monospace fonts with negative-offset glyphs
  - Phaisarn : proportional fonts, monospace fonts with negative-offset glyphs
  - Yenbut : proportional font, monospace fonts with negative-offset glyphs
  - ETL : true charcell fonts
  - NECTEC : monospace fonts with negative-offset glyphs
- Type1 fonts
  - DearBook : DB ThaiText (proportional)
  - Omega/NECTEC : Norasi (proportional)

Output Method

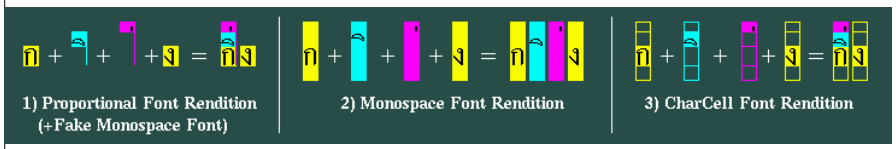
Available Thai Fonts (16 bit)

- ISO10646 BDF fonts
  - XFree86 : true charcell fonts (fixed), proportional fonts (ClearlyU)
- TrueType fonts
  - Omega/NECTEC : Norasi, Garuda (proportional)
  - Non-free : many Windows fonts available

Output Method

Thai Rendering Concerns

- Stacking of combining characters
  - Proportional fonts : use negative-offset glyphs
  - Monospace fonts : no stacking
  - CharCell fonts : special treatments on combining characters



- Cell clustering
  - The handling of invalid character sequences

Output Method

Thai Fonts Concerns

- Fake CharCell/Monospace fonts (with negative-offset glyphs) could be eliminated if CharCell rendering is properly supported by applications
- Discrimination between treatments of CharCell, Monospace and Proportional fonts in rendering engines

Output Method

Thai Rendering Concerns (cont.)

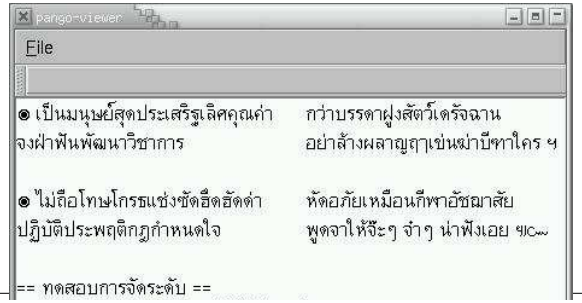
- Shaping
  - **`tis620-0`** fonts : do nothing
  - **`tis620-1`** fonts : use MacThai glyph variation
  - **`tis620-2`** fonts : use WindowsThai glyph variation

- Word Breaking
  - Thai is known to have no word delimitor
  - Word-breaking engine is necessary for line wrapping

Output Method

Status of Thai Rendering Supports

- X11/XFree86
  - XOM in X11R6.6 (XomCTL promised by li18nux)
- GTK+ 2.0 : Pango
  - WTT 2.0 clustering & stacking (proportional font assumed)
  - Normal rendering with **`iso10646`** fonts
  - Shaping for **`tis620-0, tis620-1, tis620-2`** fonts
  - Thai word breaking API defined with blank implementation



Output Method

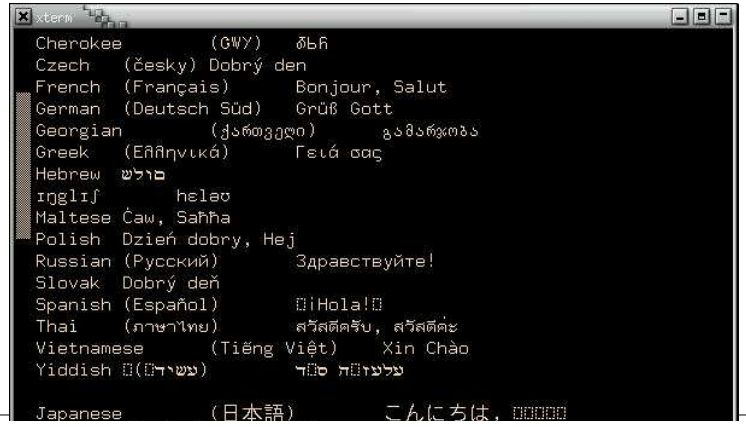
Status of Thai Rendering Supports (cont.)

- xiterm+thai
  - Stacking with fake monospace fonts (+negative-offset glyphs)
- GNU Emacs
  - Stacking with CharCell fonts
  - Shaping with dynamic glyph composition
  - Thai word breaking supported

Output Method

Status of Thai Rendering Supports (cont.)

- Qt 2.2.3
  - Stacking with proportional fonts (based on **`XFontSet`**)
- xterm
  - Stacking with CharCell fonts (Unicode supported)



Output Method

Typesetting

- Thai LaTeX support based on Babel
  - Stacking with proportional fonts
  - Shaping with TeX virtual fonts
  - Thai word breaking with filters
  - San Serif and Roman font families provided
  - Teletype font lacking
- Omega
  - At least as good as Babel, plus special kerning across word boundaries

Output Method

Word Breaking

- Thai Wordbreak filters
  - cttex (wordbreak filter + shaping for LaTeX)
  - swath (Smart Word Analysis for THai) (wordbreak filter for Babel-based Thai LaTeX, recently made to work with Omega)
- Thai Wordbreak library
  - libInThai (C++ Word Iterator)
- Thai Wordbreak support in applications
  - Mozilla (rule-based)
  - GNU Emacs (dictionary-based)

Thank you

http://linux.thai.net/

This slide is constantly updated at
ftp://ftp.nectec.or.th/pub/thailinux/cvs/docs/thaisupp/