

แนะนำ libthai

เทพพิทักษ์ การุญบุญญานันท์

thep@links.nectec.or.th

กลุ่มทำงานลินุกซ์ไทย (Thai Linux Working Group)

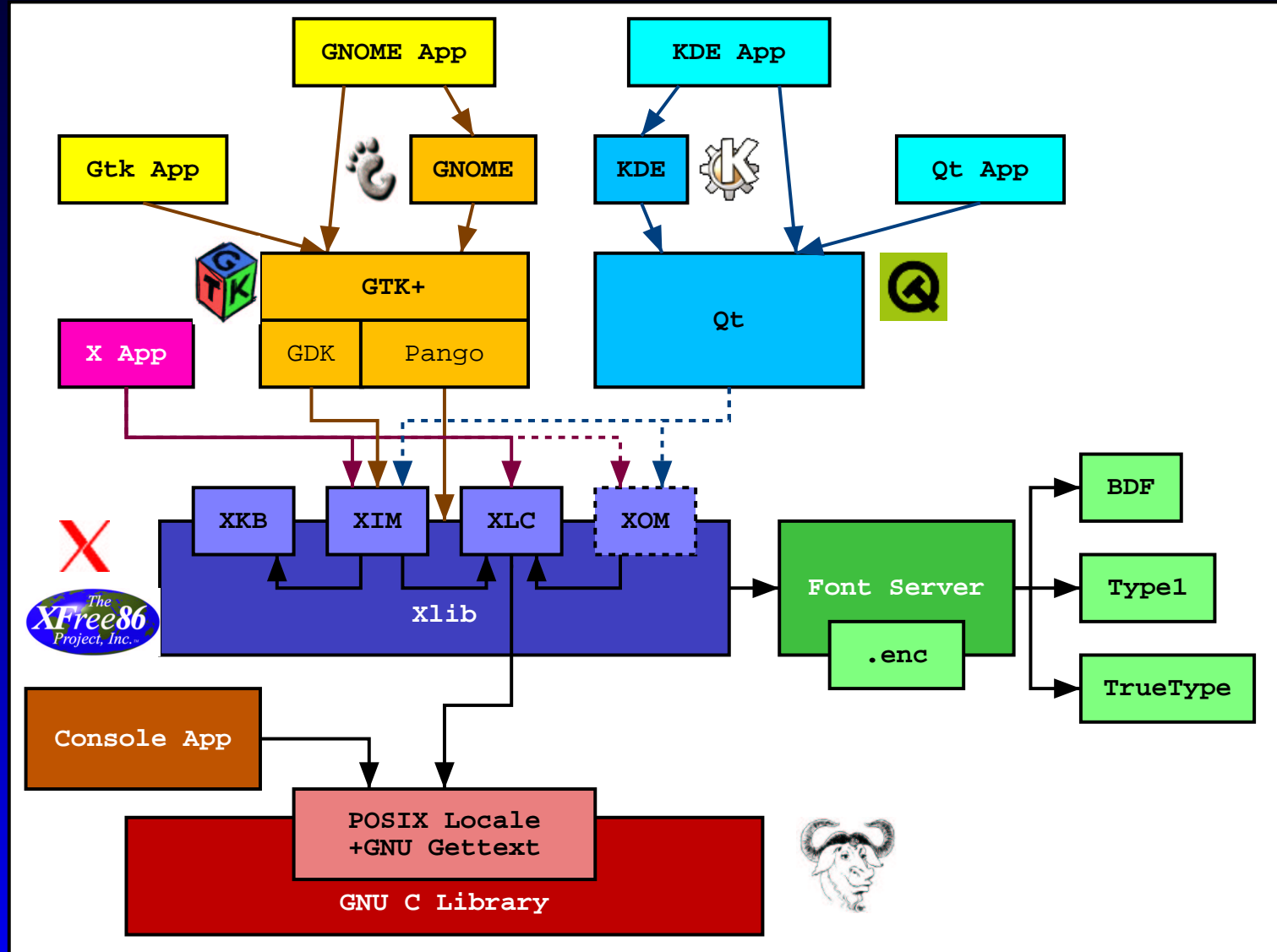
<http://linux.thai.net/>

ภูมิหลัง

- การพัฒนาระบบภาษาไทยบนลินุกซ์ที่ผ่านมา
- ความต้องการ Thai Library
- ความพยายามในอดีต

การพัฒนาาระบบภาษาไทยบนลินุกซ์ที่ ผ่านมา

ภาพรวม



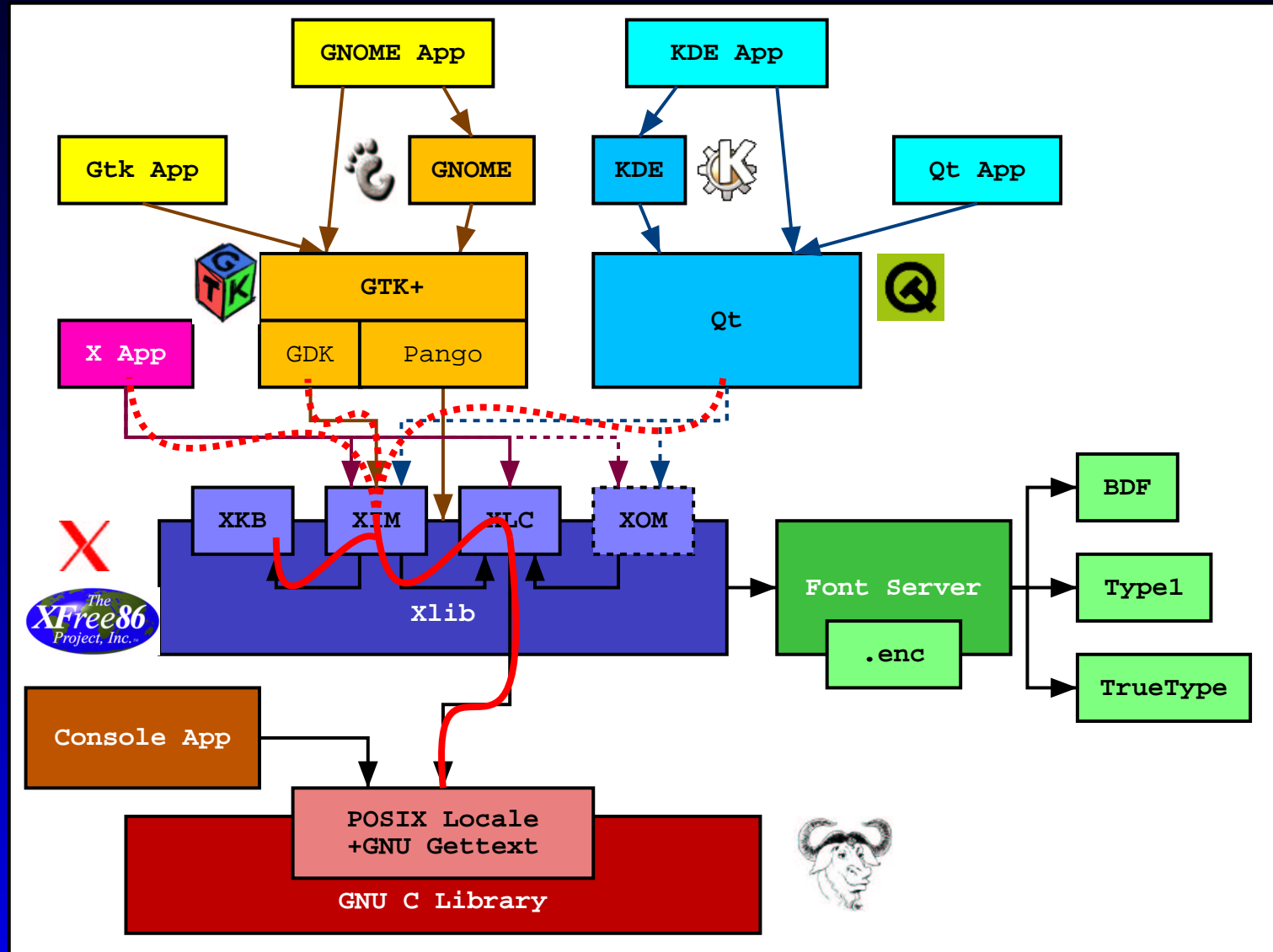
GNU C Library

- Thai Sorting Algorithms
(<http://www.links.nectec.or.th/~thep/tsort.html>)
- The Thai Locale Project (อ้างอิง ISO/IEC 14652)
(<http://software.thai.net/locale/>)
- Thai-English Bilingual Sorting (กลุ่มเฉพาะกิจ LC_COLLATE)
(<http://www.links.nectec.or.th/~thep/blsort.html>)
- Thai Locale (อ้างอิง POSIX)
(<http://www.links.nectec.or.th/~thep/th-locale/>)

GNU C Library

- glibc 2.1.1
 - TIS-620 charmap/repertoiremap
 - th_TH.TIS-620 POSIX Locale (ครบทุกหัวข้อ)
- glibc 2.2.2
 - พ.ศ. ใน LC_TIME
 - หัวข้อโลแคเลอื่นๆ ของ ISO/IEC 14652

Input Method



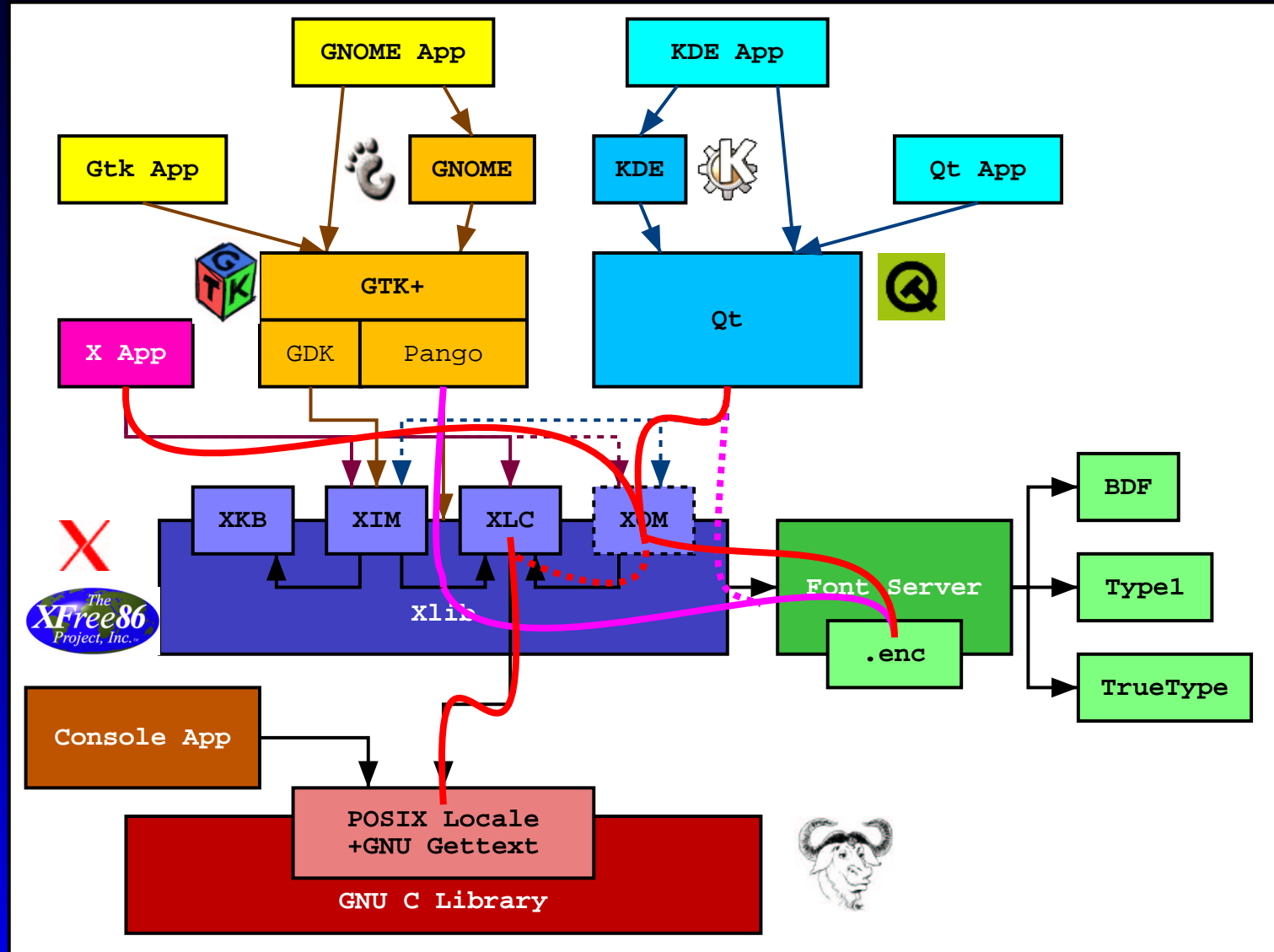
Input Method

- keyboard symbol ไทยใน XKB (Pablo Saratxaga)
- แก้ bug ใน XIM ไทยเสร็จในเวอร์ชัน 4.0.3
- ทดลอง implement เรื่อง context-sensitive Thai XIM ใน 4.1.0
- สนับสนุน callback ใน xiterm+thai 1.04

Input Method

- GTK+ 1.2 สนับสนุน XIM ใน GDK
- GTK+ 2.0
 - เพิ่ม API สำหรับ context-sensitive XIM (รวม input sequence correction ด้วย!)
 - เลื่อนเคอร์เซอร์ทีละเซลล์ได้
- Qt 2.3.0 สนับสนุน XIM แต่มีปัญหาเรื่องการเซ็ทไฟกัสคีย์บอร์ด ทำให้ XIM ไทยไม่ทำงาน
- งานในอนาคต
 - XIM ภายใต้อัปเดต UTF-8 locale
 - สนับสนุน input sequence correction ใน XIM

Output Method



Output Method

- ผลิตภัณฑ์ข้อตกลง tis620-[0|1|2]
- tis620-0 fontset ใน XFree86 4.1.0
- tis620-2 fontset check-in สำหรับ XFree86 4.2.0
- XOM: รอ XomCTL จาก li18nux ใน X11R6.6
- GTK+ 2.0: Pango:
 - cell clustering ตาม วรรท 2.0
 - shaping ตามข้อตกลง tis620-[0|1|2]
 - ขาด word break implementation
- gnome-print 0.34: TrueType support
- Qt3: ???

ความต้องการใช้ Thai Library

- โปรแกรมที่ใช้ GNU/XFree86 I18N framework
 - อาศัย glibc locale: รหัสอักขระ, เรียงลำดับคำ, จัดรูปแบบวัน-เวลา ตัวเลข และหน่วยเงิน, แปลข้อความ
 - อาศัย X11 locale: input method (XIM), font set (tis620)
 - ปัญหา: I18N framework ของ XFree86 ยังไม่ทันเหตุการณ์ในบางเรื่อง (เช่น XOM)
 - ต้องการ Thai library: ตัดคำ, cell clustering (เลื่อนเคอร์เซอร์+แสดงผล), shaping (ในระหว่างที่ XomCTL ยังไม่เข้า XFree86)

ความต้องการใช้ Thai Library

- โปรแกรมที่อาศัย toolkit
 - อาศัย glibc locale และ X11 locale ตามปกติ
 - อาศัย toolkit (GTK+/Qt): cell clustering, shaping
 - อาศัย Thai library: ตัดคำ (จนกว่าจะ check-in)
- โปรแกรมที่ cross-platform
 - GTK+/Qt: อาศัย Thai library แค่เรื่องตัดคำ
 - อาศัย Thai library ทุกเรื่อง

ความพยายามในอดีต

- LibInThai
 - C++ library ภายใต้ MPL license
 - สนับสนุน thctype, shaping, word iterator
 - มีความพยายามจะ merge กับ thctype และ Thai sorting
 - ระบบ build ที่ซับซ้อน และ mozilla-oriented
 - predefined macro ที่ทำให้มี C/C++ syntax ที่เป็นแบบฉบับของตัวเอง
 - GNOME, GTK+ และ XFree86 ไม่ใช่ C++

ความพยายามในอดีต

- ELEFANT
 - Extensible Language Extension Framework for Application supporting Native Thai
 - C++ framework สำหรับ pluggable implementation
 - ริเริ่มโดยเจ้าของ libinthai เพื่อขยายงานต่อ โดยหารือกับนักพัฒนาในเน็ต
 - ขณะเดียวกัน มีการผลักดันโครงการนี้ในเนคเทค แต่..
 - การทำงานหยุดชะงักเพราะต้องแยกย้ายมาทำ Linux TLE และ Open Office ในเวลาต่อมา
 - ขาดผู้ร่วมพัฒนา

LibThai

- การพัฒนาระบบภาษาไทยมาถึงขั้นที่ต้องการ library สนับสนุนภาษาไทยอย่างจริงจัง
- รื้อฟื้นความคิดเรื่อง Thai library

แนวคิด

- implement ด้วยภาษา C เพื่อความสะดวกในการ check-in
- เน้นความง่าย ไม่สร้าง syntax ใหม่
- ใช้ระบบ build ของ GNU เป็นหลัก
- ใช้ LGPL
- อิงมาตรฐาน วทท 2.0, มอก. 620-2533 และ ISO/IEC 10646-1
- พัฒนาโดย TLWG

โครงสร้างของ libthai

- Character Set Supports
 - มอก. 620-2533 (TIS-620)
 - ISO/IEC 10646-1 (UCS-2, UCS-4, UTF-8)
- Cell Clustering ตาม วทท. 2.0
 - ใช้โดย Input Method ส่วนเลื่อนเคอร์เซอร์ และการแก้ลำดับ input
 - ใช้โดย Output Method เพื่อ render อักขระผิดลำดับ

โครงสร้างของ libthai

- Input Method
 - Input Sequence Check ตาม วทท. 2.0:
 - Level 0: Passthrough
 - Level 1: BasicCheck
 - Level 2: Strict
 - Input Sequence Correction
- Output Method
 - Cell-based rendering
 - Shaping

โครงสร้างของ libthai

- Sorting

- สนับสนุน interface ในลักษณะ strcoll() และ strxfrm() ของ ISO C
- อ้างอิงงานจาก LC_COLLATE special group (<http://www.links.nectec.or.th/~thep/blsort.html>)
- เรียงตามหลักพจนานุกรม ฉบับราชบัณฑิตยสถาน และส่วนขยายเรื่องเครื่องหมายวรรคตอน ตัวเลข และการเข้าลำดับกับภาษาอังกฤษ

โครงสร้างของ libthai

- Word Break
 - ปัจจุบันใช้โค้ดจาก ctex
 - มีโครงการจะปรับปรุงในอนาคตเพื่อให้ใช้ทรัพยากรน้อยลง
- อื่นๆ
 - Spelling Check
 - Soundex
 - ฯลฯ

เว็บไซต์ของโครงการ

- <http://linux.thai.net/libthai>
- <http://libthai.sourceforge.net>