Thai & Hindi Support in Sun's Java™ Runtime Environment

Naoto Sato

Member of Technical Staff
Sun Microsystems, Inc.

Outline

- Thai & Hindi Overview
- Text Rendering - Glyphs
- Text Processing - Characters
- Text Input - Keystrokes
- Non-Text Issues
- Localization
- Demo
- Future work

Hindi Overview

- J2RE 1.4 supports Hindi/Devanagari
  - Left to right
  - 48 letters - 13 vowels and 35 consonants
    - Consonants have inherent vowels (usually short vowel “a”)
    - Shaping, reordering
  - Very large number of glyph forms in use
    - Minimally requires a few hundred
- Hindu Calendar
  - Lunisolar

Hindi Example

- च्‌छ - ca + Virama + cha
- च्छ - ccha
- च्छ - ccha conjunct
- इच्छा - "wish" using the ccha conjunct
- इच्‌छा - "wish" without the ccha conjunct

Thai Overview

- Derived from South Indian through Khmer
  - Left to Right - no word breaks
  - 44 consonants with inherent vowels
  - Diacritics for 18 vowels and 6 diphthongs. Placed around corresponding consonant
  - Tone indication using 4 tone marks. Placed on top of consonant and vowel
  - 4 stacking levels
- Buddhist Calendar
  - Essentially Gregorian with a 543 year offset

Thai Example

ไทย - ai th y
- thai(y)

น้ำส้มคั้น - n am so m ku n
- fresh orange juice

หยุด - (silent) y oo d
- stop

Complex Text

- Requires more effort than Latin based scripts
- N:M mapping of characters to glyphs
- May have the following properties
  - Bidirectionality
  - Shaping
  - Ligatures
  - Positioning
  - Reordering
  - Split characters

Complex Text in J2RE

- Layout engines in J2RE 1.2 and 1.3
  - Capable just for Arabic and Hebrew
  - Feature includes Bidi, shaping, ligatures
- New rendering engine in J2RE 1.4
  - OpenType for Devanagari
  - Windows glyph set for Thai
  - Possible to add support for other glyph sets
- More glyphs in the J2RE supplied font
  - Lucida Sans font includes additional 625 glyphs

Text Processing

- Text Breaking
  - Finding of boundaries in text
    - Character, Word, Sentence, and Line
  - Used for text layout, selecting or highlighting a word, or caret movement
- Collation
  - Locale sensitive sorting
    - 4 strength levels (locale specific)
    - Extensible plain text rule based

Text Breaking

- Demarcation rule based text breaking prior to J2RE 1.4.
- Hindi required extra rules to support consonant clusters and other punctuation
- Thai has no demarcation between words
  - Required dictionary based text breaking
  - Maximal matching with backtracking

Collation

- Hindi uses the same mechanism with other scripts
- Thai requires extra processing for leading vowels
  - Vowels 0E40 through 0E44 must be swapped with following consonant (0E01 - 0E2E)

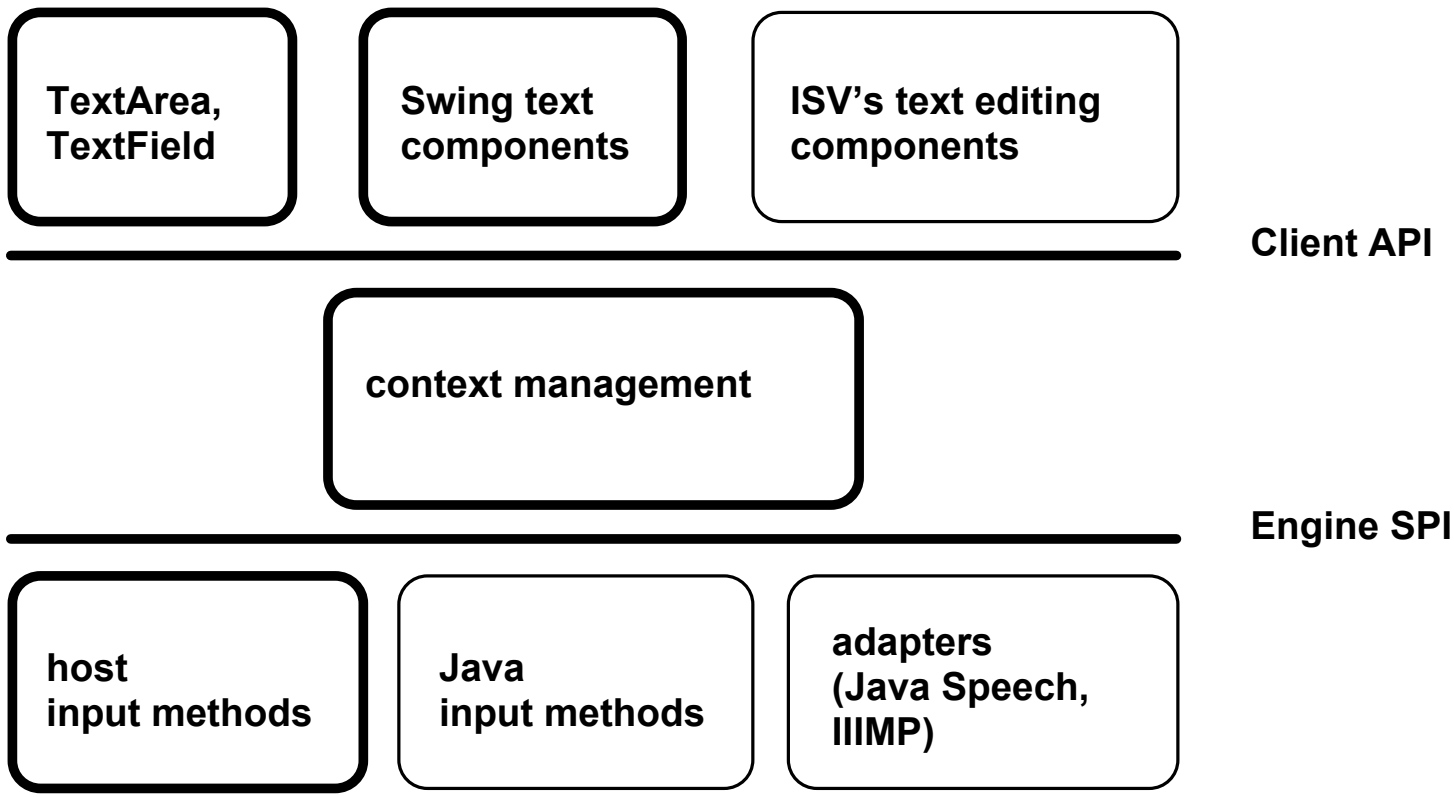
กเขฃ 0E01 **0E40** 0E02 0E03

กขเฃ 0E01 0E02 **0E40** 0E03

Text Input

- Thai and Hindi text input is made possible through the Input Method Framework
- Both host input methods and Java input methods are available for inputting Thai and Hindi

Input Method Framework



Java input method

- Easily pluggable
  - Devanagari input method in 1.4.1
  - Thai input method in the near future

- Thai input method will support Thai input sequence checking (WTT 2.0)

Typing ท ◌่ ◌ี makes ที่

Typing ท ◌ี ◌่ does not make ที่

Non-Text Issues

- Calendar
  - Thai Buddhist calendar support is provided
  - Hindi lunisolar calendar is not
- Character Conversion
  - Standard converters provided as part of the J2RE
  - Thai - TIS620
  - Hindi - ISCII

Localization

- Date/Time & Number formatting
- Localized names for languages/countries
- Support for 3 different Thai locales supporting different combinations of calendar and formatting
  - th - Gregorian calendar with Western digits
  - th_TH - Thai calendar with Western digits
  - th_TH_TH - Thai calendar with Thai digits

Demonstration

- Thai and Hindi in J2RE demo

Future Work

- Pluggable locale framework
  - Adding favorite locales
  - Replace existing locale elements
- More locales
  - E.g. th_TH_TH in release 1.4.1
- Improved calendar support
  - Support for non-Gregorian calendars
- Text rendering optimization
  - Eliminate synchronization cost

Information Pointers

- Java 2 SDK documentation

  http://java.sun.com/j2se/1.4/docs

- Supported locales

  http://java.sun.com/j2se/1.4/docs/guide/intl/locale.doc.html

- Java internationalization tutorial

  http://java.sun.com/docs/books/tutorial/i18n/

- Please send us feedback:

  java-intl@java.sun.com