Global Information Systems:

Localization and Internationalization (5)

Prof. Dr. Jan M. Pawlowski
Autumn 2013
Contents

- Introduction
- Definitions and Terms
- Design approaches
- Summary
The Open Unified Process – Disciplines

- Architecture
  - Architecture Notebook
- Configuration and Change Management
- Development
  - Design
  - Build
  - Developer Test
  - Implementation
- Project Management
  - Iteration Plan
  - Project Plan
  - Work Items List
  - Risk List

- Requirements
  - Supporting Requirements Specification
  - Vision
  - Use Case
  - Glossary
  - Use-Case Model
- Test
  - Test Case
  - Test Log
  - Test Script

Roles
Artefacts / Support

[Source: http://www.epfwiki.net/wikis/openup/]

سلاً، يتم استخدام الإنترنت في مجموعة واسعة من الأغراض. من المفيد استخدام车辆 للبحث عن المعلومات وتوفير روابط للمواقع المفيدة. يمكن استخدام هذه المواقع للتعلم والبحث. أعتقد أن الإنترنت يستخدم بشكل كبير في العالم اليوم.
Samples
Definitions

- **Internationalization (I18N)** is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for redesign. Internationalization takes place at the level of program design and document development (W3C, 2007).

- **Localization (L10N)** is the process of taking a product and making it linguistically and culturally appropriate to a given target locale (country/region and language) where it will be used (W3C, 2007).
Definitions

- **Globalization (G11N)** defines a business strategy and business activities to act on a global market.

- A **Locale** is a geographic location and a language of a region (e.g., Germany, French-speaking Quebec, Central Finland) – classes based on a locale are locale-sensitive.
Types of internationalization

- Application development (business logic)
- User interface design (presentation logic)
- Time
  - Run-time
  - Compile-time
  - Design-time
- Aspects
  - Software
  - Documentation (process documentation, help, manual)
  - Web pages
  - Learning materials
  - Knowledge & experiences
Types of internationalization

Abstract GUI

GUI for culture X

Culture X Locale

GUI for culture Y

Culture Y Locale

[Adapted from Kersten, 2002]
Types of internationalization

Production

Surface Culture X
Deep Culture X
Surface Culture Y
Deep Culture Y

Product

GUI X
Core Application X
GUI Y
Core Application X

Deployment

Surface Culture X
Deep Culture X
Surface Culture Y
Deep Culture Y
Challenges in Localization

- Text string expansion
- Character sets and encoding
- Bidirectional text and vertical display
- Keyboard character layout, shortcuts
- Fonts
- Sorting order
- Placeholders
- Abbreviations
- Terminology
- And many more
Aspects

- Formats
  - Date
  - Time
  - Currency
  - Addresses, Postal codes
- Symbols, icons, graphics, colors
- Language
  - Translation
  - Writing system
  - Characters
- Other
  - Contents…
  - Sounds
  - Messages
  - Measurements / Units
Format samples

- Dates:
  - 31.10.2007, 13:15:26 CET
  - 10-31-2007, 01.15.26 am CET
  - 31 OCT 2007, 13 h 15 CET
  - ...

- Numbers
  - 1 234 567,89
  - 1.234.567,89
  - 1,234,567,89
  - 1,234,567,89

- Additionally: Other calendars, holidays
- Separate representation and presentation – using identifiers, string indexing
Localization by country

ISO 3166 Country Codes

<table>
<thead>
<tr>
<th>Official country names used by the ISO 3166/MA</th>
<th>Numeric</th>
<th>Alpha-3</th>
<th>Alpha-2</th>
<th>Local ISO codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>004</td>
<td>AFG</td>
<td>AF</td>
<td>ISO 3166-2:AF</td>
</tr>
<tr>
<td>Åland Islands</td>
<td>248</td>
<td>ALA</td>
<td>AX</td>
<td>ISO 3166-2:AX</td>
</tr>
<tr>
<td>Albania</td>
<td>008</td>
<td>ALB</td>
<td>AL</td>
<td>ISO 3166-2:AL</td>
</tr>
<tr>
<td>Algeria</td>
<td>012</td>
<td>DZA</td>
<td>DZ</td>
<td>ISO 3166-2:DZ</td>
</tr>
<tr>
<td>American Samoa</td>
<td>016</td>
<td>ASM</td>
<td>AS</td>
<td>ISO 3166-2:AS</td>
</tr>
<tr>
<td>Andorra</td>
<td>020</td>
<td>AND</td>
<td>AD</td>
<td>ISO 3166-2:AD</td>
</tr>
<tr>
<td>Angola</td>
<td>024</td>
<td>AGO</td>
<td>AO</td>
<td>ISO 3166-2:AO</td>
</tr>
<tr>
<td>Anguilla</td>
<td>660</td>
<td>AIA</td>
<td>AI</td>
<td>ISO 3166-2:AI</td>
</tr>
<tr>
<td>Antarctica</td>
<td>010</td>
<td>ATA</td>
<td>AQ</td>
<td>ISO 3166-2:AQ</td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>028</td>
<td>ATG</td>
<td>AG</td>
<td>ISO 3166-2:AG</td>
</tr>
<tr>
<td>Argentina</td>
<td>032</td>
<td>ARG</td>
<td>AR</td>
<td>ISO 3166-2:AR</td>
</tr>
<tr>
<td>Armenia</td>
<td>051</td>
<td>ARM</td>
<td>AM</td>
<td>ISO 3166-2:AM</td>
</tr>
<tr>
<td>Aruba</td>
<td>533</td>
<td>ABW</td>
<td>AW</td>
<td>ISO 3166-2:AW</td>
</tr>
</tbody>
</table>

## Localization by language

### ISO 639 Language Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Language Name</th>
<th>ISO 639 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>arn</td>
<td>Mapudungun; Mapuche</td>
<td>mapudungun; mapuche; mapuche</td>
</tr>
<tr>
<td>arp</td>
<td>Arapaho</td>
<td>arapaho</td>
</tr>
<tr>
<td>art</td>
<td>Artificial (Other)</td>
<td>artificielles, autres langues</td>
</tr>
<tr>
<td>arw</td>
<td>Arawak</td>
<td>arawak</td>
</tr>
<tr>
<td>asm</td>
<td>Assamese</td>
<td>assamais</td>
</tr>
<tr>
<td>ast</td>
<td>Asturian; Bable</td>
<td>asturien; bable</td>
</tr>
<tr>
<td>ath</td>
<td>Athapascan languages</td>
<td>athapascanes, langues</td>
</tr>
<tr>
<td>aus</td>
<td>Australian languages</td>
<td>australiennes, langues</td>
</tr>
<tr>
<td>ava</td>
<td>Avaric</td>
<td>avar</td>
</tr>
<tr>
<td>ave</td>
<td>Avestan</td>
<td>avestique</td>
</tr>
<tr>
<td>awa</td>
<td>Awadhi</td>
<td>awadhi</td>
</tr>
<tr>
<td>aym</td>
<td>Aymara</td>
<td>aymara</td>
</tr>
<tr>
<td>aze</td>
<td>Azerbaijani</td>
<td>azéri</td>
</tr>
<tr>
<td>bad</td>
<td>Banda languages</td>
<td>banda, langues</td>
</tr>
<tr>
<td>bai</td>
<td>Bamileke languages</td>
<td>bamilékés, langues</td>
</tr>
</tbody>
</table>

But….the example of Khmer…

- Written from left to right, characters being placed also above and below the main line of writing.
- Words are not separated by spaces. A space in Khmer is a punctuation sign similar to a comma.
- A word is composed of clusters, syllemes. They are not a proper syllable, as syllables are a unit of consonants and vowels pronounced in one stroke of breath. Consonants pronounced after a vowel are part of the syllable, but not part of the cluster or syllemes.

Unicode is a universal character set, i.e. a standard that defines, in one place, all the characters needed for writing the majority of living languages in use on computers. It aims to be, and to a large extent already is, a superset of all other character sets that have been encoded.

A coded character set is a set of characters for which a unique number has been assigned to each character. Units of a coded character set are known as code points. For example, the code point for the letter à in the Unicode coded character set is 225 in decimal, or E1 in hexadecimal notation. (Note that hexadecimal notation is commonly used for identifying such characters, and will be used here.)

The character encoding reflects the way these abstract characters are mapped to bytes for manipulation in a computer. (W3C, 2007)
Formats

- **Character**: The smallest component of written language that has semantic value; refers to the abstract meaning and/or shape (Unicode Glossary, 2007)

- **Visual rendering** introduces the notion of a glyph. Glyphs are defined by ISO/IEC 9541-1 [ISO/IEC 9541-1] as "a recognizable abstract graphic symbol which is independent of a specific design". There is not a one-to-one correspondence between characters and glyphs. (W3C, 2005)
Specifications, software and content MUST NOT require or depend on a one-to-one correspondence between characters and the sounds of a language.

Specifications, software and content MUST NOT require or depend on a one-to-one mapping between characters and units of displayed text.

Protocols, data formats and APIs MUST store, interchange or process text data in logical order.

Independent of whether some implementation uses logical selection or visual selection, characters selected MUST be kept in logical order in storage.

Specifications of protocols and APIs that involve selection of ranges SHOULD provide for discontiguous logical selections, at least to the extent necessary to support implementation of visual selection on screen on top of those protocols and APIs.
Specifications and software MUST NOT require nor depend on a single keystroke resulting in a single character, nor that a single character be input with a single keystroke (even with modifiers), nor that keyboards are the same all over the world.

Software that sorts or searches text for users SHOULD do so on the basis of appropriate collation units and ordering rules for the relevant language and/or application.

Specifications, software and content MUST NOT require or depend on a one-to-one relationship between characters and units of physical storage.

More on characters and encoding: [http://www.w3.org/TR/charmod](http://www.w3.org/TR/charmod)
## Formats

- Different encodings for character sets
  - ISO 8859-1
  - Unicode

<table>
<thead>
<tr>
<th>Code point</th>
<th>UTF-8</th>
<th>UTF-16</th>
<th>UTF-32</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>U+0041</td>
<td>00 41</td>
<td>00 00 41</td>
</tr>
<tr>
<td>נ</td>
<td>U+05D0</td>
<td>05 D0</td>
<td>00 00 05 D0</td>
</tr>
<tr>
<td>好</td>
<td>U+597D</td>
<td>59 7D</td>
<td>00 00 59 7D</td>
</tr>
<tr>
<td>不</td>
<td>U+233B4</td>
<td>F0 A3 8E B4</td>
<td>D8 4C DF B4</td>
</tr>
</tbody>
</table>
Recommendation samples

- Internationalisation Tag Set (W3C)
  - Used to develop localizable schemata
  - Identifying translation needs
  - Elements: Translate, localization note, terminology, directionality, language information, elements within text
Recommendation samples

Internationalisation Tag Set (W3C)

```xml
<Res
  xmlns:its="http://www.w3.org/2005/11/its"
  its:version="1.0">
  <prolog
    its:translate="no">
    <revision>Sep-07-2006</revision>
    <its:rules version="1.0">
      <its:translateRule selector="/msg/notes" translate="no"/>
      <its:locNoteRule locNoteType="description" selector="/msg/data">
        <its:locNote>The variable \{0\} is the name of the host.</its:locNote>
      </its:locNoteRule>
    </its:rules>
  </prolog>
  <body>
    <msg id="HostNotFound">
      <data>Host \{0\} cannot be found.</data>
    </msg>
    <msg id="HostDisconnected">
      <data>The connection with \{0\} has been lost.</data>
    </msg>
    <msg id="FileNotFoundException">
      <data>
        its:locNote="\{0\} is a filename">\{0\} not found.</data>
      </msg>
  </body>
</Res>
```

[Source: http://www.w3.org/TR/2007/REC-its-20070403]
“Culturalization” of applications

- Culture awareness
- Adapting business logic
- Adapting contents
- Adapting user interfaces

- Samples for culturally adapted interfaces
Types of internationalization

[Adapted from Kersten, 2002]
Culture-aware internationalization

[Adapted from Kersten, 2002]
At the end of this phase, the following results should be ready:

- Strategy for internationalization & localization
  - Design planning
  - Architecture refinement
  - Standards, guidelines
Summary

- There is no one-fits-all strategy for internationalization and localization.
- Standards should be considered.
- Based on a culture analysis, (internal) guidelines should be developed.
- Prototyping and participating is essential.
- Other individualization / personalization strategies should be considered.
Questions

- Describe the differences of globalization, internationalization, localization and adaptation.
- Which aspects should be considered when designing and developing international solutions?
- Which guidelines can be applied for designing a website for a Finnish university?
- Which steps are necessary to develop an Asian marketing site for JYU?
References


Contact Information ITRI

Prof. Dr. Jan M. Pawlowski
jan.pawlowski@titu.jyu.fi
Skype: jan_m_pawlowski

Office:
Telephone +358 14 260 2596
Fax +358 14 260 2544
http://users.jyu.fi/~japawlow