A Micro Crowdsourcing Architecture to Localize Web Content for Less-Resourced Languages

Asanka Wasala  Chris Exton  Ruvan Weerasinghe  Reinhard Schäler
Adapted from: http://w3techs.com/technologies/overview/content_language/all
Adapted from: http://w3techs.com/technologies/overview/content_language/all
Scenario

You’ve come across a Sinhala web site..
ඒහල අද අද
ද
<table>
<thead>
<tr>
<th>Detect language</th>
<th>Chinese</th>
<th>Georgian</th>
<th>Italian</th>
<th>Persian</th>
<th>Tamil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afrikaans</td>
<td>Croatian</td>
<td>German</td>
<td>Japanese</td>
<td>Polish</td>
<td>Telugu</td>
</tr>
<tr>
<td>Albanian</td>
<td>Czech</td>
<td>Greek</td>
<td>Kannada</td>
<td>Portuguese</td>
<td>Thai</td>
</tr>
<tr>
<td>Arabic</td>
<td>Danish</td>
<td>Gujarai</td>
<td>Korean</td>
<td>Romanian</td>
<td>Turkish</td>
</tr>
<tr>
<td>Armenian</td>
<td>Dutch</td>
<td>Haitian Creole</td>
<td>Latin</td>
<td>Russian</td>
<td></td>
</tr>
<tr>
<td>Azerbaijani</td>
<td>English</td>
<td>Hebrew</td>
<td>Latvian</td>
<td>Serbian</td>
<td></td>
</tr>
<tr>
<td>Basque</td>
<td>Estonian</td>
<td>Hindi</td>
<td>Lithuanian</td>
<td>Slovak</td>
<td></td>
</tr>
<tr>
<td>Belarusian</td>
<td>Filipino</td>
<td>Hungarian</td>
<td>Macedonian</td>
<td>Slovenian</td>
<td></td>
</tr>
<tr>
<td>Bengali</td>
<td>Finnish</td>
<td>Icelandic</td>
<td>Malay</td>
<td>Spanish</td>
<td></td>
</tr>
<tr>
<td>Bulgarian</td>
<td>French</td>
<td>Indonesian</td>
<td>Maltese</td>
<td>Swahili</td>
<td></td>
</tr>
<tr>
<td>Catalan</td>
<td>Galician</td>
<td>Irish</td>
<td>Norwegian</td>
<td>Swedish</td>
<td></td>
</tr>
</tbody>
</table>

Type text or a website address or URL to translate.
Web site Localisation.

Web site l10n

Automatic (MT)  Manual
Internet
English Web
“English dominated web pages on the Internet have resulted in businesses losing sales revenues of $10 million per year”

-Harvard-based Forrester Research
English
Sinhalese New Year

From Wikipedia, the free encyclopedia

This article includes a list of references, related reading or external links, but its accuracy remain unclear because it lacks inline citations. Please improve this article by introducing more citations where necessary. (April 2011)

The timing of the Sinhala New Year coincides with the new year of many Southeast calendars of South and Southeast Asia. The festival has close parallels to the Thai New year, Bengali New Year, and Sankranthi in India [citation needed].

The Sinhalese anthropological history of the 'New Year' which is celebrated on month of April, goes back to an ancient period in Sri Lankan history. Various legends, perhaps those associated with the harvest, gave birth to many customs, and ceremonies connected with the New Year. The 3rd month of Buddhism in the third

<table>
<thead>
<tr>
<th>Sinhala new year</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Official name</strong></td>
</tr>
<tr>
<td><strong>Observed by</strong></td>
</tr>
<tr>
<td><strong>Type</strong></td>
</tr>
<tr>
<td><strong>Date</strong></td>
</tr>
<tr>
<td><strong>Celebrations</strong></td>
</tr>
<tr>
<td><strong>Observances</strong></td>
</tr>
</tbody>
</table>
Real Time Localisation of Software

Real Time Web Content Localisation

Real Time Localisation of Software
Web site Localisation..

Web site l1on

Automatic (MT) + Manual

+ TM
What is interoperability? Why is it important? How do we achieve it?

A few years ago, I converted my garage into a music studio to support my musical hobbies and interests. The initial investment I made in time and money has paid off in that I have spent countless hours of creative bliss experimenting with music on my own, and although I enjoy my hobbies immensely, I always come to the same conclusion after each session—music is social; I need other musicians to help me fully experience the phenomenon of creative expression through music. I need a group. I need a band.
Let's Make Music Together!

What is interoperability? Why is it important? How do we achieve it?

A few years ago, I converted my garage into a music studio to support my musical hobbies and interests. The initial investment I made in time and money has paid off in that I have spent countless hours of creative bliss experimenting with music on my own, but although I enjoy my hobbies immensely, I always come to the same conclusion after each session—music is social; I need other musicians to help me fully experience the phenomenon of creative expression through music. I need a group. I need a band.
<xliff version='1.2' xmlns='urn:oasis:names:tc:xliff:document:1.2'>
  <file original='site/e1/1223' source-language='en' target-language='si' datatype='plaintext'>
    <body>
      <trans-unit id='124i8j'>
        <source>Let’s Make Music Together!</source>
        <target>සැෙවම එක පයෝ මගින් යනු</target>
        <context-group>
          <context context-type="element">
            xpath:html/body/div#ja-a-content/div#ja-clearfix/a
          </context>
        </context-group>
      </trans-unit>
    </body>
  </file>
</xliff>
A few years ago, I converted my garage into a music studio to support my musical hobbies and interests. The initial investment I made in time and money has paid off in experimenting with music on my own, but although I enjoy my hobbies immensely, I always come to the same conclusion after each session—music is social; I need a group. I need a band.
Issues

- Legal issues
- L10n of non-textual content
- Constant updates
- Style/Formatting issues
- Translation voting issues
- Deployment issues
Benefits

- A practical web translation method for less-resourced languages
- Web site independent framework
- Non-intrusive mechanism for local language users to benefit from the vast knowledge embodied in the Internet
- A method to facilitate l10n of web pages (in early stage of l10n cycle) e.g. in context l10n./;
- TM based – more accurate translations
- Helps building of resources such as parallel corpora
Standardization considerations

- Different extension mechanisms
Standardization considerations

- Different rendering mechanisms/engines (esp. mobile)
Standardization Considerations

- Browser Extensions
- Rendering Engines and Fonts
Summery

- Multilingual Web is not just the top10 languages
- Browser extensions
  - Enable a majority to obtain the most out of the vast body of knowledge encompassed in the largely English
- Novel TM based web site localisation methodology
- Standardization considerations
  - Browser Extensions
  - Rendering Engines
Thank you