

W3C i18n Workshop

Discussion Topic

Richard Ishida
Global Design Consultant
Xerox



International DTDs

Outline

- ◆ Standard tags for localisers
- ◆ Guidelines for DTD design
- ◆ And more...

- Allow for identification of non-translatable text.

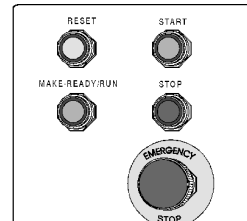
Von der VISTA-Konsole aus einem Druckauftrag übermitteln. (Siehe hierzu "Auftrag übergeben" in Kapitel 5.)

Am Steuerpult prüfen, ob der Make-Ready-Modus aktiv ist. (Die Anzeige MAKE-READY/RUN darf nicht leuchten).



START drücken, so dass die Hupe ertönt und die Anzeige MAKE READY / RUN blinkt.

Beim dritten Ton erneut START drücken. Die Anzeige START leuchtet konstant und der



```
<para>
  Press the
  <screen-text translate="no">START</screen-text>
  button to sound the horn. The
  <screen-text translate="no">MAKE-READY/ RUN</screen-text>
  indicator flashes.
</para>
```

- Allow for developers to send notes to localisers.

When the prompt reappears, type the date as follows: Date (Year Month)

'Year' is a 4-digit number, and 'Month' is a number between 1 and 12.

```
<para translation-note="The user must input the text marked as non-translatable
in English, otherwise the parser will not recognise the command.">
  When the prompt reappears, type the date as follows:
  <screen-text translation-alert="There must be a space between the parenthesis
  and the words Year and Month in this example!">
    <span translate="no">Date</span>
    ( Year Month )
  </screen-text>
</para>
```

- Don't use translatable text in attribute values.

In Tamil some combining characters which are stored after a base character are displayed before, and adopt special shapes.

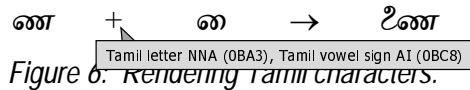


Figure 6: Rendering Tamil characters.

This illustrates how rendering rules may apply significant intelligence to the display of a sequence of code points.

```
<figure id="chdjachi" alt-text="Tamil letter NNA (0BA3), Tamil vowel sign AI (0BC8)">
  <graphic entity="graphic6" />
  <caption>Rendering Tamil characters.</caption>
</figure>
```

- Don't use translatable text in attribute values.

In Tamil some combining characters which are stored after a base character are displayed before, and adopt special shapes.

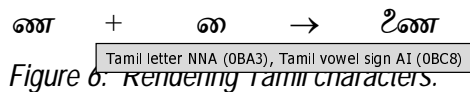


Figure 6: Rendering Tamil characters.

This illustrates how rendering rules may apply significant intelligence to the display of a sequence of code points.

```
<figure id="chdjachi">
  <graphic entity="graphic6" />
  <alt-text>Tamil letter NNA (0BA3), Tamil vowel sign AI (0BC8)</alt-text>
  <caption>Rendering Tamil characters.</caption>
</figure>
```

- Don't use translatable text in attribute values.

Partly as a result of these assumptions, it instructs the user to delete the current id values and replace them by republishing^T the document to the cross-reference database under the current publication id.

```
... replace them by republishing <index-entry term="document : republishing" /> the
document to the cross-reference database ...
```

- Don't use translatable text in attribute values.

Partly as a result of these assumptions, it instructs the user to delete the current id values and replace them by republishing^T the document to the cross-reference database under the current publication id.

```
... replace them by republishing
<index-entry>
  <index-primary>document</index-primary>
  <index-sub>republishing</index-sub>
</index-entry>
the document to the cross-reference database ...
```

- ▶ Don't use translatable text in attribute values.

Summary:

Allows application of meta-information such as translators notes, non-translate flags, etc.

Makes it easier to deal with text in a translation tool.

Note: Drives the author to a WYSIWYN environment for authoring, or one which allows for conditional visibility of elements.



- ▶ Make emphasis & style conventions logical, not presentational.

Formatting is not the same world-wide, eg:

- different ways of emphasising text in Japan (wakiten and amikake)
- problems of resolution to support bold and italics in small CJK characters on-screen
- no upper- vs. lower-case distinction in most non-Latin scripts
- no convention of distinguishing between proportional and mono-spaced fonts for some scripts



- ▶ Make emphasis & style conventions logical, not presentational.

When these elements are published to the cross-referencing database, the value of the display text is set based on the position of the element in the ***current document***.

When these elements are published to the cross-referencing database, the value of the display text is set based on the position of the element in the `<bold-italic>current document</bold-italic>`.

- ▶ Make emphasis & style conventions logical, not presentational.

When these elements are published to the cross-referencing database, the value of the display text is set based on the position of the element in the ***current document***.

When these elements are published to the cross-referencing database, the value of the display text is set based on the position of the element in the `<emph type="bold-italic">current document</emph>`.

- Make emphasis & style conventions logical, not presentational.

When these elements are published to the cross-referencing database, the value of the display text is set based on the position of the element in the **current document**.

importance
irony
contrast
stress
exasperation
etc.

When these elements are published to the cross-referencing database, the value of the display text is set based on the position of the element in the `<emph type="importance">current document</emph>`.

- Make emphasis & style conventions logical, not presentational.

Summary:

- Provide distinct elements for style and emphasis.
- Require authors to specify the intention of emphasis rather than the presentation.
- Apply the presentation in the stylesheet.
- Consider conditional clauses in the stylesheet based on language of the text.
- Note: provides more discipline, consistency and flexibility in author's use of emphasis & style conventions!

Semantic markup (terms, translation units, phrases...)

Extraction of localisable information from stylesheets

White space handling for world scripts

Markup to support international script features

Best practises for developing DTDs, stylesheet, content ...

Etc. ...

