

## Internationalization Tools: ITS

Richard Ishida  
W3C Internationalization Lead

Overview 

General aims  
Using ITS  
Supported concepts

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 2

I18n Overview: Document formats  
Script-specific markup

W3C®

Characters as ordered in memory:  
The title says "<span>תוליע פואניבה , W3C</span>" in Hebrew.



The title says "W3C , פעילות הבינאום" in Hebrew.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 3

Certain types of markup that are needed to support the use of non-Latin scripts on the Web. One important example is markup to support bidirectional text in languages based on Arabic or Hebrew scripts.

The ITS (International Tag Set) Working Group at the W3C is currently specifying markup of this kind that can be used by any schema developer to support international use of documents, and also efficient localization of documents.

For example, this slide shows what you would expect to see displayed, given the sequence of characters in memory shown near the top of the slide.

I18n Overview: Document formats  
Script-specific markup 

Characters as ordered in memory:  
The title says "<span>פ תוליע םואניבה , W3C</span>" in Hebrew.

The title says "W3C , פעילות הבינאום" in Hebrew. 

 Using the bidi algorithm only  
The title says "פעילות הבינאום , W3C" in Hebrew.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 4

If we rely solely on the Unicode bidirectional algorithm we will not get what we expected.

I18n Overview: Document formats  
Script-specific markup

W3C®

Characters as ordered in memory:  
The title says "<span dir="rtl">פועליות הבינאום, W3C</span>" in Hebrew.

The title says "פועליות הבינאום, W3C" in Hebrew. ✓

✗

Using the bidi algorithm only  
The title says "פועליות הבינאום, W3C" in Hebrew.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 5

This slide shows markup that can be used to indicate the desired directional behavior over and above what comes with the Unicode bidirectional algorithm.

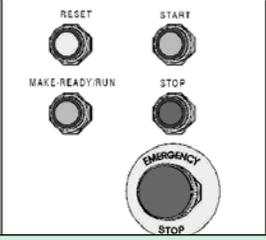
W3C<sup>®</sup>
I18n Overview: Document formats  
Markup to support localization

At the VISTA console, submit a job to print. (Refer to “Submitting a Job” in Chapter 5.)

At the operator control panel, make sure the printing system is in Make-Ready mode. The MAKE-READY/RUN indicator should not be lit.

→ Press the START button to sound the horn. The MAKE READY / RUN indicator flashes.

At the third beep, press the START button again. The START indicator remains lit and paper



```

<para>
  Press the
  <uixtext>START</uixtext>
  button to sound the horn. The
  <uixtext>MAKE-READY/ RUN</uixtext>
  indicator flashes.
</para>

```

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 6

Another area the ITS Working Group is addressing relates to improvements for translation and localization.

An example of markup that can help make translation more efficient is the provision of a flag to indicate whether or not text should be translated. This can be used by translation tools to screen text from translators or machine translation systems where necessary.

In this example of product documentation, 'START' and 'MAKE-READY/RUN' appear on a hard panel that will not be translated. This is not obvious from the markup shown above, but additional markup could be used to indicate that (see the next slide).

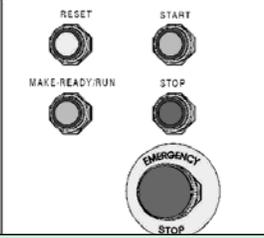
W3C<sup>®</sup>
I18n Overview: Document formats  
Markup to support localization

At the VISTA console, submit a job to print. (Refer to “Submitting a Job” in Chapter 5.)

At the operator control panel, make sure the printing system is in Make-Ready mode. The MAKE-READY/RUN indicator should not be lit.

→ Press the START button to sound the horn. The MAKE READY / RUN indicator flashes.

At the third beep, press the START button again. The START indicator remains lit and beeps.



```

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

```

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 7

Addition of a translate attribute in this case provides significant clarification for human translators and (in particular) automatic translation tools.

In actuality, the ITS group will come up with more than one way of implementing a translate flag. In some cases these may be used by content authors, in other cases they may be applied via rules. For more detail, follow the development of the working draft at <http://www.w3.org/TR/its/> .

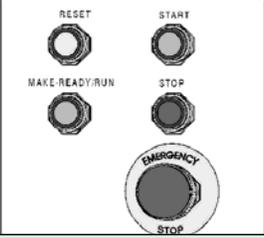
W3C<sup>®</sup>
I18n Overview: Document formats  
Markup to support localization

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>
  <uixtext translate="no">START</uixtext>
  drücken, so dass die Hupe ertönt und die Anzeige
  <uixtext translate="no">MAKE-READY/ RUN</uixtext>
  blinkt.
</para>
            
```

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 8

This shows the translated text, with hard panel citations preserved in English.

W3C<sup>®</sup>
I18n Overview: Document formats

Avoid text in attributes, and other such useful advice

Volcanic eruptions have literally devastated large inhabited areas. During the 1914 eruption of Sakurajima in Kyushu, 687 houses in Kurokami were buried in hot ash. What remained of this shrine gate, previously five meters tall, was left as a reminder.



*Kurokami maibutsu gate*  
(腹五社神社黒神埋没鳥居),  
*Sakurajima Island.*

```
<image src="kk-torii.jpg" height="180" width="240"
caption="Kurokami maibutsu gate (腹五社神社黒神埋没鳥居), Sakurajima Island."/>
```

Can't mark up for language, bidirectional markup, abbreviation, styling, etc.

Copyright © 2005 W3C (MIT, ERCIM, Keio)
slide 9

In some cases, an approach to schema design is important, rather than specific tags. For example, the Japanese text in an attribute value shown here cannot be marked up for language, directionality, abbreviation, styling, etc, since it is part of the attribute text.

W3C<sup>®</sup>

## I18n Overview: Document formats

Avoid text in attributes, and other such useful advice

Volcanic eruptions have literally devastated large inhabited areas. During the 1914 eruption of Sakurajima in Kyushu, 687 houses in Kurokami were buried in hot ash. What remained of this shrine gate, previously five meters tall, was left as a reminder.



*Kurokami maibutsu gate*  
(腹五社神社黒神埋没鳥居),  
*Sakurajima Island.*

```

<image src="kk-torii.jpg" height="180" width="240">
  <caption>Kurokami maibutsu gate
  (<span xml:lang="ja">腹五社神社黒神埋没鳥居</span>),
  Sakurajima Island.</caption>
</image>

```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 10

It would have made more sense to use an element for the caption.

The ITS Working Group will also provide advice of this kind to schema developers.

The I18n Core Working Group has also discussed concepts such as this with other W3C working groups. For example, XHTML 2 will hopefully address a number of situations in HTML where text cannot be marked up appropriately.

Overview 

General aims  
Using ITS  
Supported concepts

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 11

Local use 

```
<para>  
  Press the  
  <uixtext>START</uixtext>  
  button to sound the horn. The  
  <uixtext>MAKE-READY/ RUN</uixtext>  
  indicator flashes.  
</para>
```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 12

The concepts described in the ITS specification can be implemented locally or globally.

To illustrate this we will take an example introduced earlier – the need to indicate that the source language should not be translated. This is particularly useful for automated translation tools, but also for human translators.

Local use 

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

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 13

The content author could use an `its:translate` attribute on local markup to indicate the text that must be left in the source language. This is a local application of an ITS concept.

Global use
W3C<sup>®</sup>

```

<its:rules ... its:version="1.0">
  <its:translateRule selector="//uitext" translate="no"/>
</its:rules>

<para>
  Press the
  <uitext>START</uitext>
  button to sound the horn. The
  <uitext>MAKE-READY/ RUN</uitext>
  indicator flashes.
</para>

```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 14

Alternatively, a global rule could be defined to indicate that the content of all `<uitext>` elements should be left in the source language.

The global rule would most likely appear in the document head section, and would identify the concept involved, the location we are talking about (using XPath expressions), and indicating the semantics to be applied (`translate='no'`).

This approach may remove work from the content author where there is a clear separation of translatable and non-translatable text.

Global use
W3C<sup>®</sup>

```

<its:rules ... its:version="1.0">
  <its:translateRule selector="//*[@change='false']" translate="no"/>
</its:rules>

<para>
  Press the
  <uixtext change="false">START</uixtext>
  button to sound the horn. The
  <uixtext change="false">MAKE-READY/ RUN</uixtext>
  indicator flashes.
</para>

```

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 15

An alternative use for the global rules is to indicate that an existing markup construct in the document structure fulfills exactly the same role as concepts defined by the ITS specification.

In this case there is an existing attribute in the document format called 'change' which has values of true and false. The global rule shown states that any element where the change attribute is set to false should be left in the source language.

These are fairly simple examples. The XPath expressions can be made much more complex and sophisticated, if the content warrants it. The key benefit is that ITS allows you to describe the intended semantics of a document in such a way that a processor or translation tool can know what to do with the content.

Overview 

General aims  
Using ITS  
Supported concepts

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 16



## Data categories

- **Translability**
  - Should content be translated?
- **Localization note**
  - Communicate information to localizers about an item of content.
- **Terminology**
  - Mark terms, and optionally point to information about them.
- **Directionality**
  - Values matching those of dir in XHTML 2.
- **Ruby**
  - Markup to support ruby text.
- **Language information**
  - Identify markup that does the same as xml:lang.
- **Elements within text**
  - Identify elements that should/shouldn't make text segmentation boundaries.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 17

The concepts described by ITS are referred to in the specification as data categories.

This slide describes the data categories addressed by this version of the specification. In addition, a span-like element is proposed.



Thank you  
<http://www.w3.org/International/>

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 18

The ITS specification is currently in Working Draft, but has recently been through the Last Call process. It is hoped that the specification will progress to Candidate Recommendation status very soon.