



Internationalization



A Tag Set for the Support of Internationalization and Localization

on this page: [front matter](#) - [its working group](#) - [its requirements](#) - [significant issues](#) - [further reading](#)

front matter

Intended audience

This presentation was developed by Felix Sasaki and Richard Ishida for the 28th International Unicode Conference. It provides a snapshot of ongoing work at the W3C in the Internationalization Tag Set Working Group.

Why should you read this?

The mission of the ITS Working Group is to develop a set of elements and attributes that can be used with new DTDs/Schemas to support the internationalization and localization of documents; and provide best practice techniques for developers of DTDs/Schemas that show how to enable internationalization of their documents.

This presentation will introduce the general goals of the W3C Internationalization Tag Set ("ITS") Working Group. We will describe the basic concepts of ITS, its application scenarios, and the main features and questions which are arising during its development. A key issue is the integration of ITS into existing, "given as is" schemas, which are possibly relying on different schema languages (DTDs, XML Schema, Relax NG).

Outline

The tutorial covers the following topics:

- The ITS Working Group
- ITS requirements
- Significant issues

Please send any comments to fsasaki@w3.org.

the ITS working group

Goal

The goal of the W3C Internationalization Tag Set (ITS) Working Group is the development of a set of elements and attributes that can be integrated into schemas for XML documents to support the internationalization and localization of documents. In addition, the Working Group will produce guidelines for developers of schemas that show how to enable internationalization.

The ITS tag set will provide a standard vocabulary for internationalization and localization needs. This vocabulary can be integrated into existing schemas such as DocBook or RSS, or into newly developed schemas. It will enable worldwide use of documents based on the schema, and will also support the needs of localizers who need to make the localization process more efficient

ITS Working Group

- ▶ Goal of Internationalization Tag Set Working Group
- ◆ Development of
 - ◆ a tag set for the purposes of Internationalization and Localization
 - ◆ guidelines for developers of schemas that show how to enable internationalization
- ◆ ITS: elements and attributes to be integrated into new or existing schemas:
 - ◆ DocBook
 - ◆ RSS
 - ◆ ...

W3C

Scope

IXML documents are the scope of ITS. Nevertheless, there are extreme differences between XML documents with respect to the nature of the data. Formats like OpenOffice concentrate on text. XUL (XML User Interface Language) is used for the description of user interfaces. I.e., it applies XML as a software format. The DocBook markup scheme provides tags for text AND software code. It is used mainly for software documentation. Finally, XHTML includes presentational and text structural aspects.

ITS Working Group

▶ Scope 5

- XML documents that
 - centre on text: OpenOffice
 - focus on code: e.g. XUL
 - mix prose and code: e.g. DocBook
 - include presentational aspects: e.g. XHTML

W3C

ITS requirements

Bidirectional text support

Let's have a look at some example requirements for ITS.

Generally the Unicode bidirectional algorithm will cause text in scripts such as Arabic and Hebrew to render mixed script text in an appropriate order. Sometimes, however, additional help is needed.

For example, in the phrase on the slide, the 'W3C' and the comma should appear to the left side of the quotation. This cannot be achieved using the bidirectional algorithm alone.

The desired effect can be achieved using Unicode control characters, but this is not recommended. Markup is needed to establish the default directionality of a document, and to change that if appropriate by creating nested embedding levels. In the example, a element with an attribute @dir="rtl" fulfils this purpose.

The slide is titled "Bidirectional Text support" and is part of a presentation on "ITS Requirements". It illustrates how to handle mixed-script text like "W3C, פעילות הבינאום" (W3C, binational activity) in Hebrew. It shows three scenarios:

- Wrong display without markup:** The text "פעילות הבינאום, W3C" is displayed, where the Hebrew text is on the left and the English text is on the right.
- With markup:** The HTML code is shown: `"W3C ,פעילות הבינאום"`. The `dir="rtl"` attribute is highlighted in red.
- Appropriate display:** The text "W3C ,פעילות הבינאום" is displayed, where the English text and comma are on the left and the Hebrew text is on the right.

The W3C logo is visible in the bottom right corner of the slide.

Ruby markup

Ruby is used especially in Japanese or Chinese texts to add additional information e.g. about pronunciation to a text. Depending on the layout of the text (top to bottom or left to right), Ruby can be placed at various positions. The slide shows an example where Ruby is placed above the characters it refers to.

There is already a W3C specification, [Ruby Annotation](#), which describes tags for ruby markup. ITS could deploy this specification as part of its tag set.

In the example, a `<p>` element contains a `<ruby>` element. The base text is inside a `<rb>` element, the ruby annotation is inside the `<rt>` element.

The slide is titled "Ruby markup" and is part of a presentation on "ITS Requirements". It shows an example of Japanese text with ruby annotations. The text is "これは紙芝居です。" with "かみしばい" written above "紙芝居". Below the text, the HTML markup is shown: `<p>これは<ruby><rb>紙芝居</rb><rt>かみしばい</rt></ruby>です。</p>`. The W3C logo is visible in the bottom right corner.

Span-like element

For general i18n or l10n related information, a `` like element would be useful.

In many markup schemes like XHTML, such an element is already available. The example shows a fragment from an HTML document. It contains a `<code>` element with textual content. Part of its content is in Japanese. This is expressed by using the XHTML `` element.



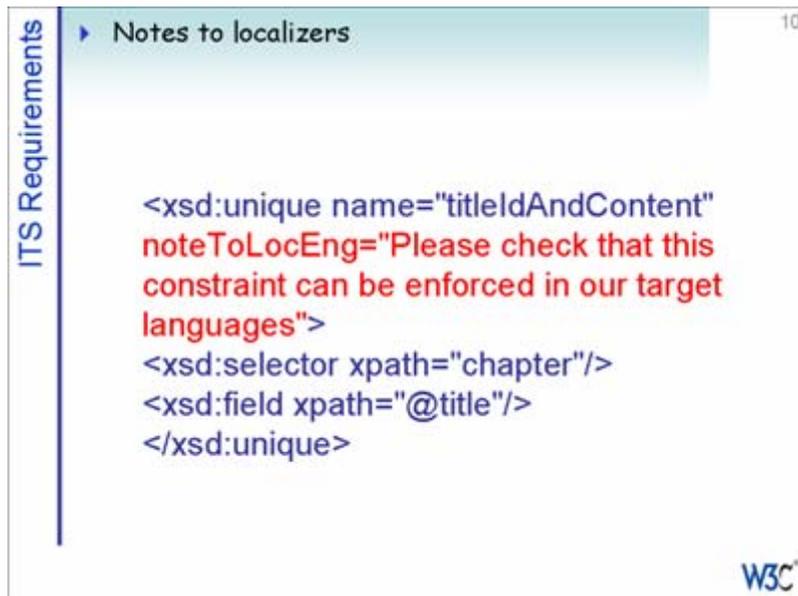
The image is a screenshot of a presentation slide. On the left side, there is a vertical blue bar with the text 'ITS Requirements' written vertically. The slide title is 'Span-like element' in a light blue box at the top. The main content is a code snippet: `<code>System.out.println("国際化活動 W3C");</code>`. The code is color-coded: the opening `<span` tag is red, the attribute `xml:lang="ja"` is red, the Japanese text `国際化活動` is black, the closing `` tag is red, and the text `W3C` is black. In the bottom right corner of the slide, there is a blue 'W3C' logo.

Notes to localizers

Another requirement for the ITS tag set which is useful for the localization process is to provide notes to localizers.

In the example, uniqueness of values should be assured for the combination of the <chapter> element and its @title attribute.

The slides shows an XML Schema definition which assures this constraint. We will not go into detail with XML Schema here. Important is the attribute @its:noteToLocEng, which conveys information about the uniqueness constraint to the localization engineer.



ITS Requirements

Notes to localizers 10

```
<xsd:unique name="titleIdAndContent"
noteToLocEng="Please check that this
constraint can be enforced in our target
languages">
<xsd:selector xpath="chapter"/>
<xsd:field xpath="@title"/>
</xsd:unique>
```

W3C

Indicator of translatability

Here we see an example for the indication of translatability.

One aspect which has to be taken into account here is the scope of such an indication. For example, in the SVG (Standardized Vector Graphics) format, the general indication is that nothing should be translated. Nevertheless, in certain contexts, for example in an SVG `<text>` element, translation is necessary.

Generally speaking, it must be possible for ITS to specify inheritance and exceptions for translatability.

The image is a screenshot of a presentation slide. On the left side, there is a vertical blue bar with the text 'ITS Requirements' written vertically. The slide title is 'Indicator of translatability' with a small blue arrow icon to its left. The slide number '11' is in the top right corner. The main content of the slide is XML code for SVG elements. The code is: `<svg ... translate="no">` on the first line, `...` on the second line, `<text translate="yes">...</text>` on the third line, and `</svg>` on the fourth line. The words 'no' and 'yes' in the code are highlighted in red. In the bottom right corner of the slide, there is a blue logo that says 'W3C'.

Guidelines on attributes

Some aspects of ITS cannot be represented as an element tag or an attribute. An example is information about attributes.

The slides shows a XUL document which describes properties of a window. The problem is here how to express that the content of the @onclick attribute should not be translated.

Due to such issues which arise with attributes, ITS proposes the guideline not to use attributes for translatable content.

The slide is titled "Guidelines on attributes" and is part of a presentation on "ITS Requirements". It displays a snippet of XUL code. The code is as follows:

```
<window ...>
  <box align="center">
    <button label="hello xFly"
      onclick="alert('Hello World');"/>
  </box>
</window>
```

The `onclick="alert('Hello World');"/>` line is highlighted in red, indicating a problem with its translatability. The slide number "12" is in the top right corner, and the "W3C" logo is in the bottom right corner.

ITS requirements

For various kinds of XML documents, the Working Group is producing these requirements which the ITS tag set and guidelines should meet.

The requirements on this slide are "work in progress". A First Working Draft was recently published, entitled [Internationalization and Localization Markup Requirements](http://www.w3.org/TR/itsreq/) <http://www.w3.org/TR/itsreq/>. A complete list of current requirements is available at the [ITS home page](http://www.w3.org/International/its/requirements) <http://www.w3.org/International/its/requirements>.

ITS Requirements 13

- [Scenario - Authoring Content](#)
- [Scenario - Terminology Creation and Translation](#)
- [Scenario - Software Resources](#)
- [Indicator of Constraints](#)
- [Handling entities](#)
- [Cultural aspects of the content](#)
- [Purpose specification/mapping](#)
- [Span-like elements](#)
- [Unique identifier](#)
- [Locale/language identification](#)
- [Term identification](#)
- [Indicator of translatability](#)
- [Limited impact](#)
- [CDATA section](#)
- [Links to internal/external text](#)
- [Bidirectional text support](#)
- [Indicator for metrics](#)
- [Attribute and translatable text](#)
- [Naming scheme](#)
- [Localization Notes](#)
- [Handling of white-spaces](#)
- [Multilingual Documents](#)
- [Annotation Markup](#)
- [Identifying Date and Time](#)
- [Scenario - Server Software in Multilingual Environment](#)

Complete list at
<http://www.w3.org/International/its/requirements/>

W3C

significant issues

Impact on existing markup

There are some issues which have to be solved during the development of ITS. One issue is the impact of ITS on existing markup.

ITS markup can conflict with this markup. In the example, the `<its:span>` element is part of an SVG `<text>` element.

XML processing like validation against the SVG schema or transformation via XSLT might break. In the example, the processing of the XSLT element `<xsl:value-of select="text"/>` should return the textual content of the `<text>` element. But since the textual content is encapsulated in the `<its:span>` element, `<xsl:value-of>` will return no value.

Significant issues

▶ Impact on existing markup 15

Potential conflict with existing markup

```
<svg
...
<text><its:span>...</its:span></text>
</svg>
```

Potential for affecting processing

```
<xsl:value-of select="text"/>
```

W3C

ITS and markup schemes

The last example depicts a general technical issue which has to be solved by the Working Group for a successful, widespread adoption of ITS. ITS needs mechanisms to solve problems like the conflict between ITS and existing markup, or the integration into different versions of a markup scheme.

Another issue to be discussed is the role of schema languages. Should ITS be defined in terms of XML DTDs, XML Schema or RELAX NG?

The Working Group has not made a decision yet whether ITS will use one of these schema languages, all of them, or a different approach.

Significant issues

- ▶ ITS and markup schemes 16
- ◆ Conflict
- ◆ Integration
- ◆ Which schema language?
 - ◆ XML DTDs
 - ◆ XML Schema
 - ◆ Relax NG
 - ◆ ...

W3C

ITS will produce...

An example for a different approach might be the definition of ITS as abstract data categories. These then could be mapped to various schema languages and existing markup schemes.

ITS might also encompass a processing model, in order to avoid the conflict with existing processing steps like the XSLT transformation mentioned before.

In this area, the Working Group has not devoted itself to a specific solution. The Working Group does not intent to develop such solutions for themselves. The goal is to review mechanisms which are currently under development at W3C or related bodies and to adopt some of them.

Significant issues

▶ ITS will produce ... 17

- ◆ Set of tags?
- ◆ Abstract description of data properties?
- ◆ Description of relationship with existing markup schemes?
- ◆ Processing model for markup?
- ◆ Something else?

W3C

further reading

- [The ITS Working Group Home Page](http://www.w3.org/International/its/) *http://www.w3.org/International/its/*
- [Internationalization and Localization Markup Requirements Working Draft](http://www.w3.org/TR/itsreq/)
http://www.w3.org/TR/itsreq/

Author: Felix Sasaki, W3C.

Content created 14 August 2005. Last update 2005-08-14 08:33 GMT

Copyright © 2005 W3C[®] (MIT, ERCIM, Keio), All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#) and [software licensing](#) rules apply. Your interactions with this site are in accordance with our [public](#) and [Member](#) privacy statements.