

Information about Internationalization and Localization in XML

Expressing Internationalization and Localization information in XML

Felix Sasaki

Richard Ishida

World Wide Web Consortium

29th Internationalization and
Unicode Conference

1

San Francisco,
March 2006

This presentation describes the current status of work on the Internationalization Tag Set (ITS), which is being developed by the W3C i18n ITS Working Group.



Overview

- Background: ITS purposes / audiences
- ITS basic usage: data categories
- ITS extended usage:
 - positioning data categories
 - mapping data categories
- Open issues

We will first give an overview of the purpose and possible audiences for ITS. Then we will describe its basic and extended usage scenarios. Finally some current open issues are discussed.

Information about Internationalization and Localization in XML


Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage Open Issues

W3C[®]

Background


29th Internationalization and Unicode Conference 3 San Francisco, March 2006

Information about Internationalization and Localization in XML			
Background	Basic Usage	Extended Usage	Open Issues
<h2>Background</h2> <ul style="list-style-type: none">• ITS: "Internationalization Tag Set" http://www.w3.org/TR/its• Draft produced by the W3C i18n ITS Working Group http://www.w3.org/International/its			
29th Internationalization and Unicode Conference	4	San Francisco, March 2006	

ITS, the "internationalization tag set", is a working draft of the W3C i18n ITS Working Group.

Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage Open Issues



Background


- Target: XML Documents and Schemas
- Purpose: To express information ("data categories") for internationalization and localization of XML documents and schemas

29th Internationalization and Unicode Conference5San Francisco, March 2006

ITS targets XML documents and schemas (e.g. XML DTDs, W3C XML Schema or RELAX NG). The purpose of ITS is to express information (in the terminology of ITS so-called "data categories") for internationalization and localization of XML. Examples for data categories will be given later.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Audience / Usage Scenarios

- Content authoring
- Terminology creation and translation
- Software development


29th Internationalization and Unicode Conference6San Francisco, March 2006

The audience for ITS are

- content authors who need to mark up internationalization-related or localization-related information in an XML document.
- support of terminology creation and translation in the localization process, e.g. insertion of special markers for terms, and
- software development, where the software-related material (code and / or documentation) is stored in an XML based format).

Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage Open Issues



Requirements on ITS

29th Internationalization and Unicode Conference7San Francisco, March 2006

Besides the ITS tagset which is covered in this talk, the ITS working group is dealing with a lot of other topics. Some of them will be handled within the tagset specification. Others will be described as techniques for the localization and internationalization of XML in a separate document. The following slide provides a list of the most important topics the working group is working on.

Information about Internationalization and Localization in XML

Information about Internationalization and Localization in XML

- 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

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

Information about Internationalization and Localization in XML

Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage Open Issues

W3C[®]


Basic Usage

29th Internationalization and Unicode Conference 9 San Francisco, March 2006

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues

ITS Data Categories in Current Working Draft



- Translatability
- Localization Information
- Terminology
- Directionality
- Ruby


29th Internationalization and
Unicode Conference10San Francisco,
March 2006

The basic usage of ITS is to add information, so-called "data categories" for internationalization and localization, to an XML document. The data categories on this slide are part of the current working draft.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues

Sample Data category: Translatability



- Parts of a document should not be translated:

```
<book its:translate="yes" ...>...  
  <p>And he said: you need a new <quote  
    its:translate="no">motherboard</quote>  
  </p> ...  
</book>
```


29th Internationalization and
Unicode Conference11San Francisco,
March 2006

As an example we will introduce the "translatability" data category. It is used to express information about translatability of (parts of) an XML document. In the example, an attribute `@its:translate` with the values "yes" or "no" is used for this purpose. Attaching this attribute to the `<book>` elements means that the whole textual content of this element has to be translated. This includes child elements, but excludes attributes.

An exception to this statement about translatability of child elements can be made via a `@translate` attribute at a child element. For example, the `@translate` attribute with the value "no" at the `<quote>` element means that the content of this element should not be translated.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Why Data Categories?

Separation of


1. prose description of ITS categories ("data category")
2. implementation (schema language independent)
3. schema language specific declaration (XML DTDs, XML Schema, RELAX NG)

29th Internationalization and Unicode Conference12San Francisco, March 2006

Why does ITS define data categories and not markup directly? The benefit of data categories is the separation of the prose description of what the ITS information is about (the "data category"), the implementation on a schema language independent level, and the declaration which is specific to a schema language. In its current version, ITS provides declarations for three schema language: XML DTDs, XML Schema and RELAX NG.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Example: Translatability

1. Data category: Prose description

"Parts of a document should (not) be translated)."
2. Schema language independent implementation

`<book its:translate="yes"...>... </book>`
3. Schema language specific declaration

`<!ELEMENT book ... >`
`<!ATTLIST book its:translate (yes|no) #IMPLIED>`

29th Internationalization and Unicode Conference13San Francisco, March 2006

Again we give an example of the "translatability" data category. Its prose description is very simple: "Parts of a document should (not) be translated". On a schema language independent level, "translatability" is implemented via an attribute @its:translate with the two values "yes" or "no". In the schema language XML DTDs, the attribute is declared as an optional attribute on the book and other, possibly all elements of a schema.

Information about Internationalization and Localization in XML

Background Basic Usage **Extended Usage** Open Issues


W3C[®]

Extended Usage

29th Internationalization and Unicode Conference 14 San Francisco, March 2006

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Extended Usage of ITS

- Positioning ITS data categories
 - in an XML document ("local")
 - in a schema (XML Schema, RELAX NG)
 - global
- Mapping to existing markup


29th Internationalization and Unicode Conference15San Francisco, March 2006

In addition to the basic usage of ITS, there are two aspects of ITS extended usage: positioning of ITS information, and mapping to existing markup.

There are three positions for ITS data categories: in an XML document ("local" which we introduced before), in a schema, and global.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



ITS Data Categories in a Schema

XML Schema:


```
<xs:element name="p">
  <xs:annotation>
    <xs:appinfo>
      <its:schemaRule its:translate="yes"/>
    </xs:appinfo>
  </xs:annotation> ...
</xs:element>
```

29th Internationalization and
Unicode Conference16San Francisco,
March 2006

As for a schema, ITS defines schema language specific ways of adding the data categories. In an XML Schema, the `<xs:appinfo>` element contains the attributes for ITS data categories attached to an `<its:schemaRule>` element. In the example, the attribute `@its:translate` with the value "yes" expresses that all `<p>` elements should be translated.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



ITS Data Categories in a Schema

Relax NG:


```
<element name="p">  
  <its:schemaRule its:translate="yes"/> ...  
</element>
```

29th Internationalization and
Unicode Conference17San Francisco,
March 2006

As for RELAX NG, the `<its:schemaRule>` element appears as a child of the `<element>` element. Its function is identical to the example from XML Schema.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Global ITS Data Categories

Using XPath to select parts of an XML document:

```
<someDocument ...>
....
<its:documentRule its:translate="no"
  its:translateSelector="//p[@editor='john']"/>
<!-- This rule holds for p elements which are edited by John.
--> ...
</someDocument>
```

Other selector attributes: dirSelector, termSelector, ...


29th Internationalization and
Unicode Conference18San Francisco,
March 2006

ITS data categories can also appear in "global" positions, that means independent of a specific position in an XML document or schema. For this case, a `<its:documentRule>` element is used. Additional attributes express via XPath to which information the data category should be applied to. In the example, the `@its:translateSelector` attribute (in combination with the `@its:translate` attribute) expresses that the `<p>` element with the attribute `@editor="john"` should not be translated.

For other data categories, similar selector attributes are defined like `@dirSelector` or `@termSelector`.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Combination of Positions

Global and local:


```
<book ...>
<its:documentRule its:translate="yes"
its:translateSelector="/book/body//*>
<body> ...
  <p>And he said: you need a new <quote
its:translate="no">motherboard</quote>
  </p> ... </body>
</book>
```

29th Internationalization and
Unicode Conference19San Francisco,
March 2006

It is also possible to combine the positions of ITS data categories. In the example, the content of the `<body>` element should be translated. This is expressed via the two attributes `@its:translate` and `@its:translateSelector` at the `<its:documentRule>` element. As an exception to this general specification of translatability, the content of the `<quote>` element should not be translated. This is expressed via the `@its:translate` attribute at the `<quote>` element.

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues



Mapping to Existing Markup

Mapping to DITA and other markup:

```
<book ...>
<its:documentRule
its:translate="no"
its:translateSelector="//*[@dita:translate='no']"
its:term="yes" its:termSelector="//quote"/>
<body> ...
  <p>And he said: you need a new
  <quote dita:translate="no">motherboard</quote>
  </p> ... </body>
</book>
```

29th Internationalization and Unicode Conference20San Francisco, March 2006

The selector attributes can also be used to map ITS data categories to existing markup. In the example, the attribute `@its:translate="no"` is mapped to the attribute `@dita:translate="no"` which belongs to the DITA namespace. In addition, the attribute `@its:term="yes"` which implements the terminology data category is mapped to all `<quote>` elements via the `@its:termSelector` attribute. This mapping expresses that all `<quote>` elements are interpreted as terms.

Example: Implementation in XSLT

- XML document (ITS+TEI)
- Processing with XSLT1
- Result: XSLT2
- XSLT2 is applied to (ITS+TEI) document
- Result: ITS information is visualized

Example: Implementation in XQuery

- XML document (ITS+XML Spec)
- Processing with its2xquery.xq
- Result: xmlspec1.xq
- xmlspec1.xq is applied to (ITS+XML Spec) document
- Result: ITS information is extracted

Information about Internationalization and Localization in XML

Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage **Open Issues**

W3C[®]


Open Issues

29th Internationalization and Unicode Conference 23 San Francisco, March 2006

Information about Internationalization and Localization in XML

BackgroundBasic UsageExtended UsageOpen Issues

Precedence Rules for Various Positions



Local has precedence over global:

```
<book its:translate="yes"
its:translateSelector="/book/body//*">
<body> ...
  <p>And he said: you need a new <quote
its:translate="no">motherboard</quote>
  </p> ... </body>
</book>
```

Precedence rules are a possible burden for implementers / users of ITS

29th Internationalization and Unicode Conference24San Francisco, March 2006

Open issues which the working group is discussing currently concern among others the precedence between multiple positions of the same data category. The slide shows the example from before which works only if there are rules about the precedence between local and global data categories. However, such rules produces a possible burden for implementers / users of ITS.

Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage Open Issues

W3C[®]

Complexity of ITS


- Need of various positions for ITS?
- Need to do mapping to existing markup?
- How to serve multiple audiences at the same time?

29th Internationalization and Unicode Conference 25 San Francisco, March 2006

In general the question is how much need there is for the extended usage of ITS, or whether the main application scenario will be the basic usage. The task of the working group to produce a standard for a variety of audiences underlies this question.

Information about Internationalization and Localization in XML

Background Basic Usage Extended Usage Open Issues



Modularizations

- Creating fixed modularizations for ITS, e.g. ITS+HTML, ITS+DITA, ITS+DocBook, ...
- This should encourage tool vendors to adopt ITS

29th Internationalization and Unicode Conference26San Francisco, March 2006

Finally, the working group is creating fixed modularizations of ITS with widely adopted markup vocabularies like HTML, DITA, DocBook or others. These modularizations should encourage vendors of e.g. translation tools to adopt ITS.

Information about Internationalization and Localization in XML

How to Express Information about Internationalization and Localization in XML Documents and Schemata

Richard Ishida
Felix Sasaki
World Wide Web Consortium

29th Internationalization and
Unicode Conference

27

San Francisco,
March 2006