# HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

**Renat Bikmatov, Nathan Glenn, Serge Gladkoff, Alan Melby**

**Distribution: Public**

# Document Information

| Deliverable title: | HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS) |
|---|---|
| Contractual date of delivery: | 31$^{st}$ October 2013 |
| Actual date of delivery: | 30$^{th}$ October 2013 |
| Author(s): | Renat Bikmatov (Logrus, subcontractor), Nathan Glenn (LinguaTech), Serge Gladkoff (Logrus, subcontractor), Alan Melby (LinguaTech) |

# Revision History

| Revision | Date | Author | Organization | Description |
|---|---|---|---|---|
| 1 | 22/10/2013 | Renat Bikmatov | Logrus | Draft #1 |
| 2 | 25/10/2013 | Nathan Glenn | LinguaTech | Draft #2 |
| 3 | 30/10/2013 | Renat Bikmatov | Logrus | Revision #1 |
| 4 | 31/10/2013 | Renat Bikmatov | Logrus | Revision #2: the following link to the file format converter for hybrid previewing fixed: **File converter for hybrid previewing:** |
| | | | | |
| | | | | |

# CONTENTS

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

# 1. EXECUTIVE SUMMARY

ITS (Internationalization Tag Set) 2.0 specification (http://www.w3.org/TR/its20/) was introduced by W3C in 2013 as a complement to XML (http://www.w3.org/TR/xml/), HTML5 (http://www.w3.org/TR/html5/) and XLIFF (http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html) specifications. ITS markup delivers localization instructions and other context information via meta-data tags embedded into the content.

ITS 2.0 format can be used to exchange metadata between different data processing tools, and also to deliver metadata to linguists who are working on the content. The second task is the goal of this particular project.

Currently neither browsers, nor CAT tools display these meta-data to end users. Logrus has developed visualization tools to preview ITS 2.0 metadata embedded in HTML5, XML and XLIFF files. These tools enable translators and reviewers to refer to localization context visually presented in a web browser window while working on the content in their content editor, CAT or other content management tool. The core ITS metadata categories, such as Localization Note, Terminology, and Translate will help to build the bridge between localization instructions and the content, and also fill the gap between available limited functionality of current CAT tools and required rich and flexible context preview.

Previewing context-related metadata in Web browsers is a part of long-term initiative called "Work in Context System (WICS)". This initiative has been introduced by Logrus (www.logrus.ru) as an open project. The project final goal is to develop new and integrate existing technologies to deliver context-related information to translators, editors and reviewers, increase their productivity and enhance the quality of their output.

# 2. INTRODUCTION

Translation and editing in particular and localization in general still have two unresolved problems: 1) previewing (rendering) the source content in the final or publication format, and 2) supplying translation-related context information and instructions to language workers. With CMS implementation and mass transition to asynchronous update and fragmented translation of content "by bits and pieces", the problems are only getting more and more severe. Another problem is the linking of glossaries, translation instructions and style guides to the source content for automation and facilitation of translation.

The source content for translation is usually provided to a translator in various breeds of XML and XLIFF formats. The source content for preview is usually provided to translators in raw XML format without any support of its visualization. These formats are hard to read by humans, and context information and metadata are not usually provided. However, the translator (editor) needs to have visibility into not only full source text but also terminology, trademarks, client instructions, project-specific instructions, etc. Lack of context has been identified as major cause of disruption of human work.

Why can't you simply add any necessary metadata in ITS format to your source files to be localized? Direct integration of any metadata into the content going for localization through real-life production environments is hampered by the following:

- Support of ITS 2.0 format by available CAT and other authoring tools is either missing or limited
- Even if one day a new edition of your preferred CAT tool were enriched with ITS metadata support, would it be implemented properly, fully and in the most convenient way for you? Who has never had problems with core functionality or usability of their CAT tool?
- Even if one day all authoring tools could fully support ITS 2.0 format, what would you do with legacy

Translation Memories (TM)? New pieces of content enriched with ITS metadata markup will not fully match the same pieces in TM's free of metadata. You will not be able to reuse in-context or 100% matches easily. You would get 80-90% matches instead of 100% matches because of differences in internal tags in translation units stored in the TM and pay much more to translators as a result.

- Not to mention that full support of ITS 2.0 metadata, local markup, global rules and external data by existing Translation Memory formats and engines still remains an open issue.

The solution could be to separate data processing (i.e. localization of production versions of source files in some CAT, Quality Assurance and Engineering Testing environments) from metadata preview (i.e. visualization of metadata-enriched versions of source files with some other utility) while providing some synchronization between the production and preview environments at the level of processed fragments of text (translation units or sentences within translation unit).

The proposed WICS solution from Logrus includes rendering any available auxiliary file in HTML5 format enriched with ITS 2.0 metadata. This can be done with the means of generating an additional reference file (full source) that is provided to the translator and editor in addition to the pieces of source text to be translated (translatable source) in some CAT tool. The reference file has a standard HTML5 formatting with embedded ITS 2.0 metadata tags, and visualization vehicles of HTML5 and/or Java script. The preview will not require additional proprietary special software – the standard browser is enough, on any platform (see the picture):
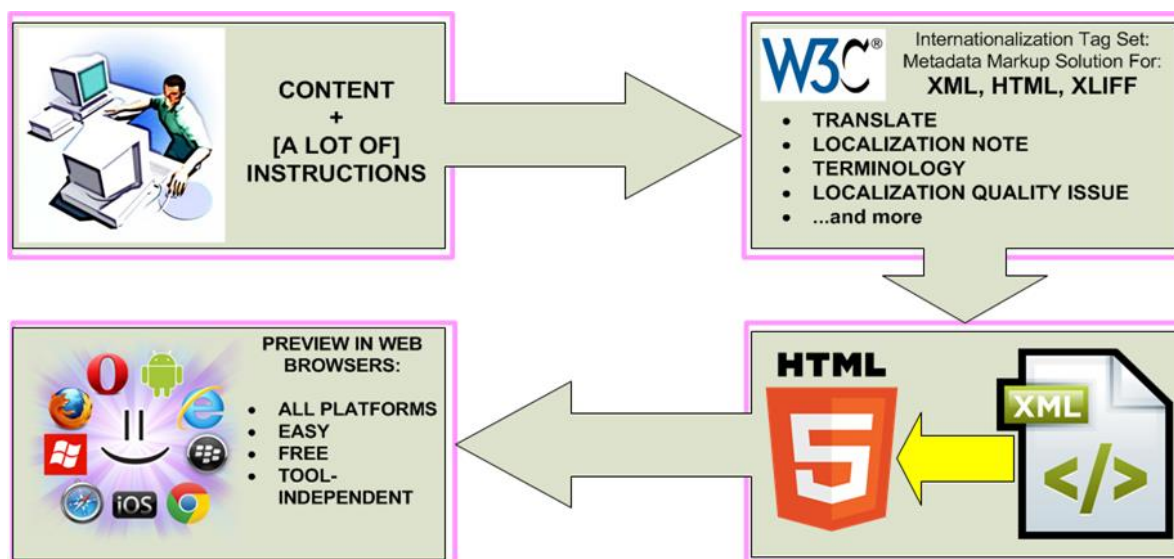


WICS software displays ITS metadata embedded into XML, XLIFF, or HTML files in Web browsers. To support XML and XLIFF formats, text conversion utilities were developed to transform these formats into HTML5 format.

ITS 2.0 information can be highlighted, color-coded and even augmented with popups. Content can be converted to HTML5 even if the content publication format is different, without "full WYSIWYG", but for even better preview for the text worker. Additional information can be shown to the text worker such as definitions, comments, instructions, parts of speech, semantic information, reference web sites (both extranet and intranet), reviewer's comments, etc.

Actual translation or editing will be carried out in another format in a certain CAT tool, but a parallel preview is certain to improve the view of the context for translators, editors, reviewers and other text workers. Such a technology will enable companies to quickly and semi- (and sometimes fully) automatically render context of

the content in a wide variety of scenarios and tools. The reference WICS file will serve as help and instructional information for human processing over a multitude of systems and processes, for authoring, translation, MT post-editing, knowledge transfer and accumulation, etc.

So the main questions motivating this project launch were 1) how do we preview the metadata and 2) how do we load the metadata with localization-related information. The answers are 3) preview metadata in Web browser outside any CAT tool and 4) develop the preview solution based on HTML5 format and Java Script to provide a universal, flexible, open solution.

# 3. INPUT DATA

WICS accepts the following formats of input files containing ITS metadata to be previewed:

- XML
- HTML
- XLIFF with some limitations set forth by W3C on recommended ITS markup for XLIFF; see:
    - http://www.w3.org/International/its/wiki/XLIFF_Mapping
    - http://www.w3.org/International/its/wiki/XLIFF_1.2_Mapping
    - http://www.w3.org/International/its/wiki/XLIFF_2.0_Mapping

To demonstrate the differences between implementations of ITS markup for XML and XLIFF, a limited demo XML-to-XLIFF converter was developed. When it comes to conversion of HTML or XML files into XLIFF format, you can compare WICS output XLIFF file with the following W3C's sample files provided just as reference:

https://github.com/w3c/its-2.0-testsuite/tree/master/its2.0/xliffsamples

Please don't think that these particular W3C's examples illustrate just extraction of content and metadata into output XLIFF format. In their extraction algorithm, ITS markup governs the fragmentation of the output into translation units, not the source XML\HTML markup nor customer-defined rules. If the output XLIFF files should observe your own fragmentation rules (INI files) and existing translation memories, you would need a customized solution similar to our XML-to-XLIFF demo converter.

# 4. OUTPUT DATA

WICS converters produce the following output format enriched with ITS metadata transferred from the source files:

- HTML
- XLIFF (for the example input XML file only)

HTML-to-HTML conversion has been introduced for technical reasons to ensure that all ITS rules are contained within the document without any external rule files. That simplifies parsing rules and implementation of rule priority and inheritance logic.

# 5. PREVIEWING METADATA – VISUAL DESIGNS

## 5.1. Computer Screen Space Allocation in Real-Life Scenarios

The following screen shot illustrates the typical screen space allocation when you translate content in some

CAT tool. Just about 1/3 of the screen is available for reference information, e.g. previewing any metadata:
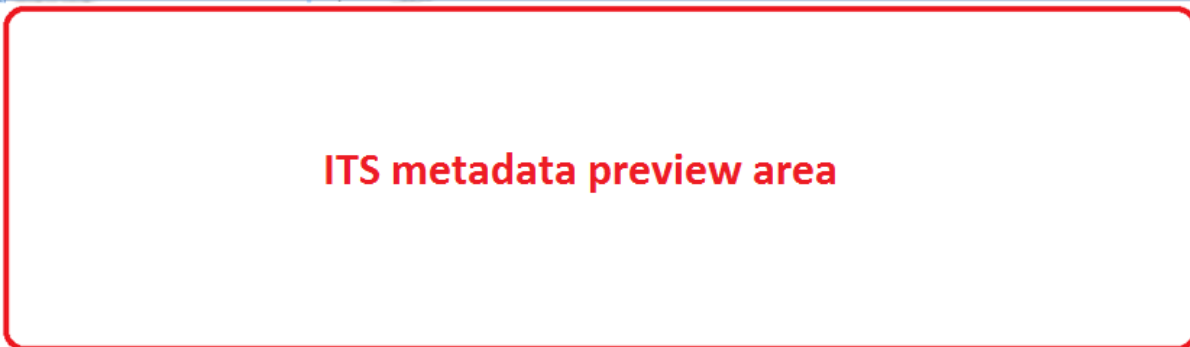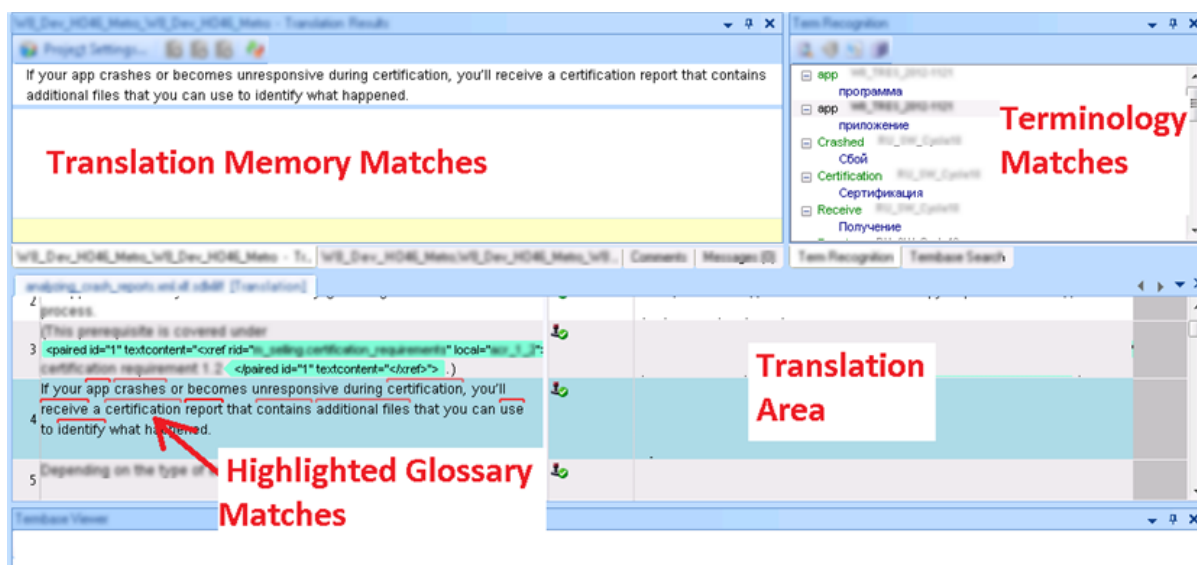


## 5.2.    The Two Levels of Preview

You might want to have at least two levels of metadata preview: compact and detailed. In compact preview mode, the content is highlighted and some content navigation features are available. You can see which pieces of content have metadata of which type linked to them, but you don't see the metadata. In detailed preview mode, the metadata items are expanded and navigation through metadata items is supported.

The sketch below illustrates the general look and feel of compact design:
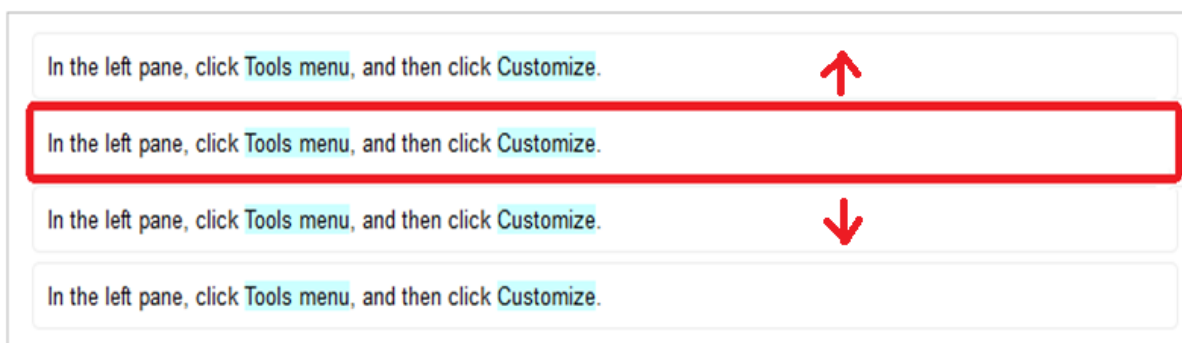


The content is presented by several strings of text. The strings are made identical just to simplify the picture. In this mode, you can navigate through the content. The selected fragment of content is enclosed in a red frame;

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

the pieces of content that have some metadata associated with them are highlighted in blue.

The following sketch illustrates the detailed metadata preview mode:
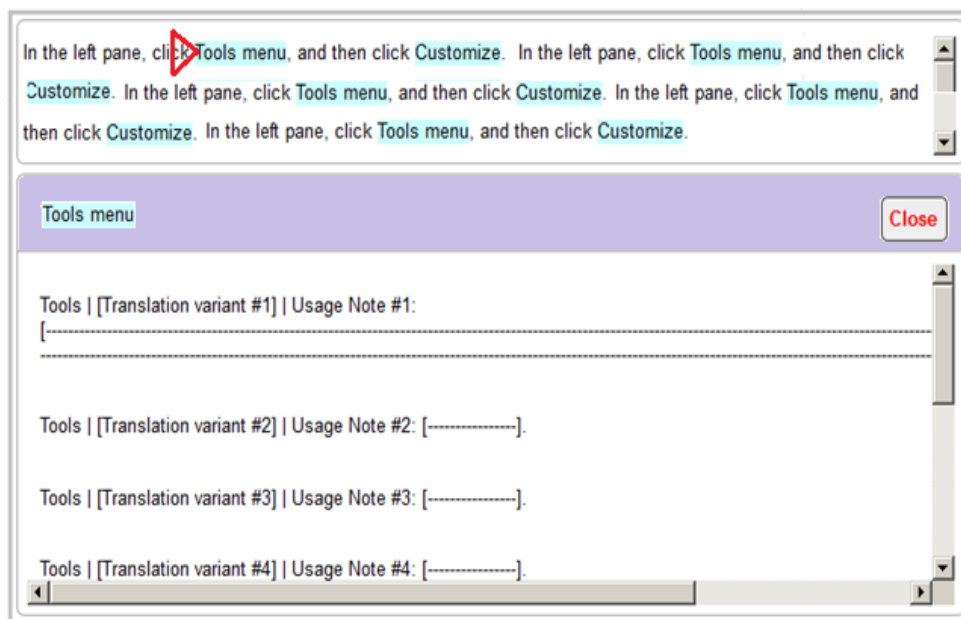


This design shows the selected fragment of content with several metadata items inside. The active metadata item is marked with red triangle. Below the content panel are the metadata item title and the metadata information panel. In this example some terminology notes are displayed.

The screen shot below demonstrates more complicated design with all metadata items expanded. The metadata items and their information panels are numbered to see which panel corresponds to which metadata item. Some metadata items are nested:

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)
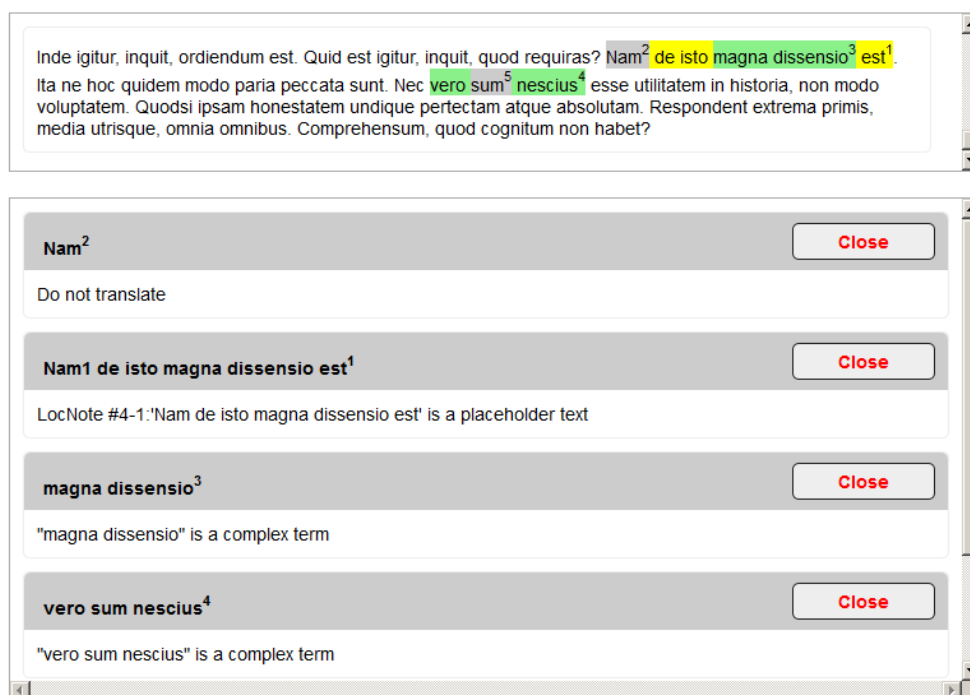
## 5.3. Visual Markers of Metadata Categories

As there are 19 different ITS metadata categories, you would need the same number of different visual markers to highlight the pieces of content associated with some metadata.
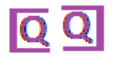
The metadata categories can be highlighted with different background colors of the text:

- Translate (gray)
- Language Information (gray)
- Localization Note (yellow)
- Terminology (green)
- Text Analysis (pink)

- Or with different foreground colors of the text:

- Localization Quality Issue (red)
- Elements Within Text – Nested (blue)

- Or with pairs of "visual" tags consisting of small pictures. The first picture represents an opening "visual" tag, the second one represents a closing "visual" tag:

- Directionality: 
- Elements Within Text: 
- Domain: 
- Locale Filter: 
- Provenance: 
- Target Pointer: {Src} {Tgt}
- Id Value: 
- Localization Quality Rating: 
- MT Confidence: 
- Allowed Characters: 
- Storage Size: 

The visual markers may look as follows:

**Markup:**

```
<p>Text with <span its-within-text='yes'>bold</span>.</p>
```

**Visualization:**

Text with <bold>.

The highlighting colors and visual marker assignments are fully customizable.

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

## 5.4. Visualization of Nested Metadata

In real-life scenarios, the metadata may have several levels of nesting; this applies both to metadata of the same type and to mixed scenarios, for example:

- [Localization Note [Localization Note]]
- [Localization Note [Terminology]]
- [Terminology [Terminology]]
- [Terminology [Text Analysis]]
- [Id Value [Id Value [Id Value]]]

## 5.5. Visualization of "Global" Metadata — ITS: Domain

Visualization of such metadata categories as Domain is non-trivial as Domain metadata are not linked to any specific pieces of content nor content markup elements. ITS Domain information needs to be displayed as complementary whenever ITS metadata of any other kind are displayed. If the content is not associated with any other metadata but Domain, some special processing needed to display Domain information.

## 5.6. Static vs. Dynamic Visualization

There are at least two ways of implementing ITS metadata preview in a Web browser: static and dynamic.

**The static preview** may be implemented with preliminary conversion of HTML files enriched with ITS markup into HTML files without ITS markup: all ITS markup are converted into HTML markup, ITS metadata are either incorporated into output HTML file as additional HTML content or stored in additional external HTML file(s) and referenced from the main HTML file.

**The dynamic preview** may be implemented with Java script and JavaScript libraries stored in external .js files. This script analyzes the content, local and global ITS rules, applies ITS rules, extracts ITS metadata and displays these metadata dynamically when you preview the HTML file in a Web browser. This task is very complex and hard to implement.

In this particular project, the dynamic preview has been implemented as Demo with limited support of ITS metadata categories due to the complexity of task.

**The hybrid preview** is a compromise between the requirement to add as less markup as possible to the source content to be previewed and the requirement to provide universal and rich preview. The hybrid preview comprises both static and dynamic approaches. As in static preview, the HTML (or XML) files containing ITS metadata need to be converted to their preview HTML versions with some additional HTML markup added to the source one. As in dynamic preview, the visualization is implemented with preview JavaScript connected to the output HTML files along with some metadata-supporting JavaScript embedded to the output files.

In this project, the hybrid preview has been implemented for all ITS metadata categories and the source files of the following formats:

- XML
- HTML

## 5.7. Demo Screen Shots
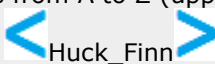
THE STATIC PREVIEW (DEMO)

The following screen shots have been taken in a Web browser from intermediate sample HTML files enriched with ITS markup. They illustrate the markup and <u>static</u> preview of ITS metadata of all types supported by HTML format. These files have been generated from the example files included in W3C ITS 2.0 Test Suite. For each sample input file the two static preview derivatives were generated:

1. With the source (XML or HTML) markup displayed (the upper screen shot)
2. With the source content displayed only (the lower screen shot)

*ITS: Allowed Characters*

<html lang="**en**" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="**utf-8**" /> <title>**Example**</title> </head> <body> <p>**Login names can only use letters from A to Z (upper or lowercase) and the character underscore (\_) and minus (-). For example:**

<code its-allowed-characters="**[a-zA-Z_\-]**"> **Huck_Finn** </code>**.**</p> </body> </html>

Login names can only use letters from A to Z (upper or lowercase) and the character underscore

(\_) and minus (-). For example: Huck_Finn .

*ITS: Directionality*

<html lang="**en**" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="**utf-8**" /> <title>**Directionality**</title> <link href="**dir3htmlrules.xml**" rel="**its-rules**" /> </head>

<body> <p>**In Hebrew, the title** <q lang="**he**" class="**rtLeft**"> הבינאום פעילות, **W3C** </q> **means "Internationalization Activity, W3C".**</p> </body> </html>

In Hebrew, the title הבינאום פעילות, W3C means "Internationalization Activity, W3C".

*ITS: Domain*

<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="**utf-8**" /> <link href="**domain1htmlrules.xml**" rel="**its-rules**" /> <meta name="**keywords**"

content="**automotive**" /> <title>**Domain HTML**</title> </head> <body> <p>**The new car from XYZ - buy it now!**</p> </body> </html>

The new car from XYZ - buy it now!

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)
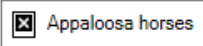
```
<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" /> <link
href="withintext1htmlrules.xml" rel="its-rules" /> <title>Appaloosa</title> </head>
<body> <section> <span itemprop="keyword">horse</span> <span

itemprop="keyword">appaloosa</span> </section> <p> The Appaloosa <cite>  The
name comes from "Palouse horse" in reference to the Palouse River in Northern Idaho.

 </cite> are rugged horses originally breed by the <span itemprop="keyword">Nez-
Perce</span> tribe in the US Northwest. </p> <p> They are often characterized by their
spotted coats, as shown here: <img src="appaloosa.png" alt="Appaloosa horses" /> </p>
<!--footer--> <p> Copyright: <em>Zebulon Inc.</em> </p> </body> </html>
```

horse appaloosa

The Appaloosa  *The name comes from "Palouse horse" in reference to the Palouse River in*

*Northern Idaho.*  are rugged horses originally breed by the Nez-Perce tribe in the US Northwest.

They are often characterized by their spotted coats, as shown here:   Appaloosa horses

Copyright: *Zebulon Inc.*

*ITS: External Resource*

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<link href="externalresource1htmlrules.xml" rel="its-rules" /> <title>Video element
example</title> </head> <body> <video width="640" height="360"
src="http://www.example.com/video/v2.mp" poster="video-image.png">If your
browser doesn't support the <code>video</code> element, you can <a
href="http://www.example.com/video/v2.mp">download the video</a>
instead.</video> </body> </html>
```

Not supported by Web browsers:

*ITS: ID Value*

```
<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" /> <title>Id
Value</title> <link href="idvalue1htmlrules.xml" rel="its-rules" /> </head> <body> <p

id="settingsMissing">   <b id="text"> Can't find settings file. </b> <var id="desc">

The module cannot find the default settings file. You need to re-initialize the system.

</var>  </p> </body> </html>
```

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

**Can't find settings file.** *The module cannot find the default settings file. You need to re-initialize the system.*

*ITS: Language Information*

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<link href="languageinfo1htmlrules.xml" rel="its-rules" /> <title>EXAMPLE</title>
</head> <body> <p>The motto of Québec is: <q lang="fr-CA">Je me souviens</q> .</p>
<p>La devise du Québec est : <q lang="fr-CA">Je me souviens</q> .</p> </body>
</html>
```

The motto of Québec is: Je me souviens .
La devise du Québec est : Je me souviens .

*ITS: Locale Filter*

```
<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>Locale filter</title> <link href="locale1htmlrules.xml" rel="its-rules" /> </head>

<body> <div id="Canada">         <p>This legal notice is only for Canadian locales.</p>
</div> </body> </html>
```

This legal notice is only for Canadian locales.

*ITS: Localization Note*

```
<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>EXAMPLE</title> <link href="locnote6htmlrules.xml" rel="its-rules" /> </head>
<body> <span id="1"> This is a hippoptamus: <img src="hippo.png" alt="Hippo in the
Chobe River" /> . </span> <span id="2"> This is an elephant: <img src="elephants.png"
alt="Elephants taking a mud bath" /> . </span> <span id="DisableInfo">The {0} has been
disabled.</span> </body> </html>
```

This is a hippoptamus: [x] Hippo in the Chobe River . This is an elephant: [x] Elephants taking a mud bath
The {0} has been disabled.

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <script
type="application/its+xml"> <its:rules xmlns:its="http://www.w3.org/2005/11/its"
xmlns:h="http://www.w3.org/1999/xhtml" version="2.0"> <its:param
name="locQualityIssueParam">q1</its:param> <its:locQualityIssueRule
selector="//h:span[@id=$locQualityIssueParam]" locQualityIssueType="typographical"
locQualityIssueComment="sentence without capitalization" locQualityIssueEnabled="yes"/>
<its:locQualityIssueRule selector="//h:span[@id='q2']" locQualityIssueType="misspelling"
locQualityIssueComment="should be 'Music can'" locQualityIssueEnabled="yes"/> </its:rules>
</script> <meta charset="utf-8" /> <title>Example</title> </head> <body> <p> <span
id="q1">music </span> is an art form whose medium is sound and silence. <span
id="q2">Musci acn </span> take many different forms and is experienced by individuals
in a range of social settings ranging from being alone to attending a large concert. </p>
</body> </html>
```

music is an art form whose medium is sound and silence. Musci acn take many different forms and is experienced by individuals in a range of social settings ranging from being alone to attending a large concert.

*ITS: Localization Quality Rating*

```
<html its-loc-quality-rating-profile-ref="http://example.org/qamodel/v13" its-loc-quality-
rating-score="100" its-loc-quality-rating-score-threshold="95" lang="en"

xmlns="http://www.w3.org/1999/xhtml">      <head> <meta charset="utf-8" /> <title>Rikki-
tikki-tavi</title> </head> <body> <p>This is the story of the great war that Rikki-tikki-
tavi fought single-handed, through the bath-rooms of the big bungalow in Segowlee
cantonment. Darzee, the Tailorbird, helped him, and Chuchundra, the musk-rat, who
never comes out into the middle of the floor, but always creeps round by the wall, gave

him advice, but Rikki-tikki did the real fighting.</p> </body>      </html>
```



This is the story of the great war that Rikki-tikki-tavi fought single-handed, through the bath-rooms of the big bungalow in Segowlee cantonment. Darzee, the Tailorbird, helped him, and Chuchundra, the musk-rat, who never comes out into the middle of the floor, but always creeps round by the wall, gave him advice, but Rikki-tikki did the real fighting.

*ITS: MT Confidence*

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>Sentences about European capitals.</title> </head> <body> <p> <span its-mt-
```

confidence="**0.8982**" its-annotators-ref="**mt-confidence|file:///tools.xml#T1**"> **Dublin is the capital of Ireland.** </span> <span its-mt-confidence="**0.8536**" its-annotators-ref="**mt-confidence|file:///tools.xml#T2**"> **The capital of the Czech Republic is Prague.** </span> <span its-mt-confidence="**0.7982**" its-annotators-ref="**mt-confidence|file:///tools.xml#T3**"> **Madrid is the capital of Spain.** </span> <span its-mt-confidence="**0.9982**" its-annotators-ref="**mt-confidence|file:///tools.xml#T4**"> **Berlin is the capital of Germany.** </span> </p> </body> </html>

Dublin is the capital of Ireland. The capital of the Czech Republic is Prague. Madrid is the capital of Spain. Berlin is the capital of Germany.

*ITS: Preserve Space*

The *Preserve Space* data category is not supported for HTML format – see ITS Specification:

> The Preserve Space data category is not applicable to HTML documents in HTML syntax because xml:space (and by extension Preserve Space) has no effect in documents parsed as text/html. However, the data category can be used in HTML in XHTML syntax.

*ITS: Provenance*

<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="**utf-8**" /> <title>**Test**</title> <script id="**pr1**" type="**application/its+xml**"> <its:provenanceRecords xml:id="pr1" xmlns:its="http://www.w3.org/2005/11/its"> <its:provenanceRecord tool="onlinemtex" revPerson="James Joyce" revTool="mycat-v1.0" revOrg="acme-CAT-v2.3" provRef="http://www.examplelsp.com/excontent987/production/prov/e6354"/> </its:provenanceRecords> </script> <script id="**pr2**" type="**application/its+xml**"> <its:provenanceRecords xml:id="pr2" xmlns:its="http://www.w3.org/2005/11/its"> <its:provenanceRecord person="John Doe" org="legaltrans-ex" revPerson="Tommy Atkins" revOrg="vistatec" provRef="http://www.vistatec.com/job-12-7-15-X31/reviewed/prov/re8573469"/> </its:provenanceRecords> </script> <link href="**provenance1htmlrules.xml**" rel="**its-rules**" /> </head> <body> <p class="**prov-records-ref**"> <a href="**#pr1**"></a>**This paragraph was translated from the machine.** </p> <p class="**prov-records-ref**"> <a href="**#pr2**"></a>**This text was translated directly by a person.** </p> </body> </html>

This paragraph was translated from the machine.
This text was translated directly by a person.

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

*ITS: Storage Size*

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>Storage Size</title> </head> <body> <p>String to translate:</p> <p
contenteditable="true" id="123" its-storage-size="25" its-storage-encoding="ISO-8859-1" its-
line-break-type="NEL">  Papua New-Guinea  </p> <p contenteditable="true" id="139" its-
storage-size="25" its-storage-encoding="ISO-8859-1" its-line-break-type="NeL">
Dominican Replubic  </p> </body> </html>
```

String to translate:

Papua New-Guinea

Dominican Replubic

*ITS: Target Pointer*

```
<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>Target Pointer</title> <link href="targetpointer1htmlrules.xml" rel="its-rules" />
</head> <body> <div id="one"> <p class="srcfolder">{Src}Remember last folder</p>
<p class="targetfolder">{Tgt}lastFolder</p> </div> <div id="two"> <p class="srcfile">
{Src}Custom file filter:</p> <p class="targetfile">{Tgt}testFile.txt</p> </div>
</body> </html>
```

{Src}Remember last folder

{Tgt}lastFolder

{Src}Custom file filter:

{Tgt}testFile.txt

*ITS: Terminology*

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>Terminology test: default</title> </head> <body> <p>We need a new <span its-
term="yEs" its-term-info-ref="#TDPV">motherboard</span> </p> </body> </html>
```

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

We need a new motherboard

*ITS: Text Analysis*

```
<html lang="en" xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<link href="textanalysis1htmlrules.xml" rel="its-rules" /> <title>Text Analysis: Local
Test</title> </head> <body> <p> <span id="dublin"
typeof="http://nerd.eurecom.fr/ontology#Location"
property="http://dbpedia.org/resource/Dublin">Dublin</span> is the <span id="capital"
typeof="http://nerd.eurecom.fr/ontology#City" about="301467919"
property="Wordnet3.0">capital</span> of Ireland.</p> </body> </html>
```

Dublin is the capital of Ireland.

*ITS: Translate*

```
<html xmlns="http://www.w3.org/1999/xhtml"> <head> <meta charset="utf-8" />
<title>HTML Link Example</title> <link href="translate1htmlrules.xml" rel="its-rules" />
</head> <body> <img src="smiley.gif" alt="Smiley face" width="42" height="42" /> <p
id="100" title="Text"> This is the first paragraph. It has some <code>un-translatable
code</code> . </p> </body> </html>
```



This is the first paragraph. It has some un-translatable code .

## THE DYNAMIC PREVIEW (DEMO)

The following screen shots have been taken in a Web browser from intermediate sample HTML file enriched with ITS markup. The preview is supported by WICS JavaScript and CSS files, and visual marker pictures linked to the sample HTML file. The screen shots illustrate the markup and dynamic navigation and preview of ITS metadata of the following key categories:

- *Elements Within Text*
- *Localization Note*
- *Terminology*
- *Translate*

The following screen shot contains the content navigation menu (the two rows of buttons on top of the window); the content preview panel with the scroll bar; the fragments (paragraphs) of the content enclosed in gray frames (the active fragment is highlighted with a red frame); the pieces of content highlighted with different background color or enclosed in visual markers. The highlighted pieces of text have some ITS metadata associated to them:

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

inponebat (Elements Within Text = "no". The element splits the text flow of its parent element and its content is an independent text flow)⅃,
eadem ille sanabat. At iste non dolendi status K[non imperium]⅄ non vocatur voluptas. Non autem hoc: igitur ne illud quidem. Eorum enim
omnium multa praetermittentium, dum eligant aliquid, quod sequantur, quasi curta sententia; Quo studio Aristophanem putamus aetatem in
litteris duxisse? Non enim, si omnia non sequebatur, idcirco non erat ortus illinc.

Inde igitur, inquit, ordiendum est. Quid est igitur, inquit, quod requiras? Nam de isto magna dissensio est. Ita ne hoc quidem modo paria
peccata sunt. Nec vero sum nescius esse utilitatem in historia, non modo voluptatem. Quodsi ‹ipsam› honestatem undique pertectam
atque absolutam. Respondent extrema primis, media utrisque, omnia omnibus. Comprehensum, quod cognitum non habet?

The following screen shot presents:

- Current ITS metadata item (the tip) highlighted with blue shadow around "vero sum nescius"
- The two nested ITS metadata items: "sum" is marked as not translatable term with *Translate* category and highlighted with gray background; "sum" is nested into "vero sum nescius" item which, in turn, is linked to *Terminology* metadata and highlighted with green background:

inponebat (Elements Within Text = "no". The element splits the text flow of its parent element and its content is an independent text flow)⅃,
eadem ille sanabat. At iste non dolendi status K[non imperium]⅄ non vocatur voluptas. Non autem hoc: igitur ne illud quidem. Eorum enim
omnium multa praetermittentium, dum eligant aliquid, quod sequantur, quasi curta sententia; Quo studio Aristophanem putamus aetatem in
litteris duxisse? Non enim, si omnia non sequebatur, idcirco non erat ortus illinc.

Inde igitur, inquit, ordiendum est. Quid est igitur, inquit, quod requiras? Nam de isto magna dissensio est. Ita ne hoc quidem modo paria
peccata sunt. Nec vero sum[2] nescius[1] esse utilitatem in historia, non modo voluptatem. Quodsi ‹ipsam› honestatem undique pertectam
atque absolutam. Respondent extrema primis, media utrisque, omnia omnibus. Comprehensum, quod cognitum non habet?

| vero sum nescius[1] | Close |
|---|---|
| "vero sum nescius" is a complex term | |

| sum[2] | Close |
|---|---|
| Do not translate | |

The following screen shot illustrates the preview mode with all tips (ITS metadata items) within active fragment of the content displayed. Each tip and related piece of content are marked with some number to help the users to determine which metadata in bottom panel relate to which piece of content in upper panel:

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

[Paragraph #3.] Cur tantas regiones barbarorum pedibus obiit, tot maria transmisit? Quae hic rei ‹publicae› vulnera ⊥ inponebat (Elements Within Text = "no": The element splits the text flow of its parent element and its content is an independent text flow)⊥, eadem ille sanabat. At iste non dolendi status ⟨[non imperium]⟩ non vocatur voluptas. Non autem hoc: igitur ne illud quidem. Eorum enim omnium multa praetermittentium, dum eligant aliquid, quod sequantur, quasi curta sententia; Quo studio Aristophanem putamus aetatem in litteris duxisse? Non enim, si omnia non sequebatur, idcirco non erat ortus illinc.

Inde igitur, inquit, ordiendum est. Quid est igitur, inquit, quod requiras? Nam[2] de isto magna dissensio[43] est[1]. Ita ne hoc quidem modo paria peccata sunt. Nec vero sum[6] nescius[5] esse utilitatem in historia, non modo voluptatem. Quodsi ‹ipsam[7]› honestatem undique pertectam atque absolutam. Respondent extrema primis, media utrisque, omnia omnibus. Comprehensum, quod cognitum non habet?

| **Nam de isto magna dissensio est[1]** | Close |
|---|---|
| LocNote #4-1:'Nam de isto magna dissensio est' is a placeholder text | |

| **Nam[2]** | Close |
|---|---|
| Do not translate | |

| **magna dissensio[3]** | Close |
|---|---|
| "magna dissensio" is a complex term | |

| **dissensio[4]** | Close |
|---|---|

## THE HYBRID PREVIEW

The following screen shots have been taken in a Web browser from intermediate sample HTML files enriched with ITS markup. The preview is supported by JavaScript and CSS files, and the same visual marker pictures embedded to the intermediate sample HTML files. The screen shots illustrate the markup and dynamic preview of ITS metadata:

- The upper pane contains the source text with some pieces highlighted or enclosed in visual markers to indicate that some ITS metadata are linked to these pieces of content
- The bottom pane contains the ITS metadata information displayed

*ITS: Allowed Characters*

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

Login names can only use letters from A to Z (upper or lowercase) and the character underscore (_) and minus (-). For example: ◄ `Huck Finn` ►[1].

---

[1] **Huck_Finn**                                                    **Close**

Allowed characters: [a-zA-Z_-]

---

*ITS: Directionality*

---

In Hebrew, the title ◄ "W3C, פעילות הבינאום" ►[1] means "Internationalization Activity, W3C".

---

[1] **פעילות הבינאום, W3C**                                          **Close**

Reading right-to-left

---

*ITS: Domain*

---

[1]

The new car from XYZ - buy it now!

---

[1]                                                                  **Close**

Domain: automotive

---

*ITS: Elements Within Text*

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

Text with ◀ bold ▶ [1].

| | |
|---|---|
| [1] **bold** | **Close** |
| Internal tag | |

⊥ horse ⊥ [1] appaloosa

The Appaloosa ❰ *The name comes from "Palouse horse" in reference to the Palouse River in Northern Idaho.* ❱ [2]
are rugged horses originally breed by the Nez-Perce tribe in the US Northwest.

They are often characterized by their spotted coats, as shown here: ☒ Appaloosa horses [3]

Copyright: *Zebulon Inc.*

| | |
|---|---|
| [1] **horse** | **Close** |
| External tag – end of text unit | |
| [2] **The name comes from "Palouse horse" in reference to the Palouse River in Northern Idaho** | **Close** |
| Independent insert | |
| [3] **Appaloosa horses** | **Close** |
| Do not translate | |

*ITS: External Resource*

<Not supported by Web browsers in HTML files>

*ITS: ID Value*

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

⌐⌐ **Can't find settings file.** ⌐₁⌐ _The module cannot find the default settings file. You need to re-initialize the_ _system._ ⌐²⌐³

---

**¹ Can't find settings file.** [Close]

ID: text

**² The module cannot find the default settings file. You need to re-initialize the system.** [Close]

ID: desc

**³ Can't find settings file. The module cannot find the default settings file. You need to ⌐** ~~system.~~ [Close]

ID: settingsMissing

_ITS: Language Information_

---

The motto of Québec is: «Je me souviens» .

La devise du Québec est : « Je me souviens¹ » .

**¹ Je me souviens** [Close]

Language: fr-CA

_ITS: Locale Filter_

*ITS: Localization Note*

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

This is a hippoptamus: ☒ Hippo in the Chobe River [1] . This is an elephant: ☒ Elephants taking a mud bath [2] . The {0} has been disabled. [3]

---

[1] **Hippo in the Chobe River**                                    [ Close ]

Do not translate
Do not translate the word "Chobe"

---

[2] **Elephants taking a mud bath**                                 [ Close ]

Do not translate
This text will be displayed over the image.

---

[3] **The {0} has been disabled.**                                  [ Close ]

The variable {0} has three possible values: 'printer', 'stacker' and 'stapler options'.

---

*ITS: Localization Quality Issue*

---

music [1] is an art form whose medium is sound and silence. Musci acn [2] take many different forms and is experienced by individuals in a range of social settings ranging from being alone to attending a large concert.

---

[1] **music**                                                       [ Close ]

Type: typographical
Comment: sentence without capitalization
Enabled: yes

---

[2] **Musci acn**                                                   [ Close ]

Type: misspelling
Comment: should be 'Music can'
Enabled: yes

*ITS: Preserve Space*

The *Preserve Space* data category is not supported for HTML format – see ITS Specification:

> The Preserve Space data category is not applicable to HTML documents in HTML syntax because xml:space (and by extension Preserve Space) has no effect in documents parsed as text/html. However, the data category can be used in HTML in XHTML syntax.

> This paragraph was translated from the machine and subsequently revised.[1]
>
> This text was translated directly by a person.

---

[1] This paragraph was translated from the machine and subsequently revise    **Close**

Tool: http://www.provdata.com/tool/#onlinemtex
Revision person: James Joyce
Revision org: acme-CAT-v2.3
Revision tool: http://www.provdata.com/RevTool/#mycat-v1
External info: http://www.examplelsp.com/excontent987/production/prov/e6354

Revision person: John Smith
Revision org: Linguaserve I.S. S.A.

---

*ITS: Storage Size*

**String to translate:**

Papua New-Guinea[1]

Dominican Replubic

---

[1] **Papua New-Guinea**                                   **Close**

Storage size (bytes): 25
Encoding: ISO-8859-1
Line break type: lf

---

*ITS: Target Pointer*

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

{Src} Remember last folder[1]

{Tgt} lastFolder[2]

{Src} Custom file filter:

{Tgt} testFile.txt

| [1] **Remember last folder** | **Close** |
| --- | --- |
| Source | |

| [2] **lastFolder** | **Close** |
| --- | --- |
| Target | |

*ITS: Terminology*

We may define *discoursal point of view*[1] as *the relationship, expressed through discourse structure, between the implied author or some other addresser, and the fiction* .

We may define *refulgence*[2] as *a radiant or resplendent quality or state*

But I'm not sure how *quincunx* can be defined.

| [1] **discoursal point of view** | **Close** |
| --- | --- |
| the relationship, expressed through discourse structure, between the implied author or some other addresser, and the fiction | |

| [2] **refulgence** | **Close** |
| --- | --- |
| a radiant or resplendent quality or state | |

*ITS: Text Analysis*

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

Dublin[1] is the capital[2] of Ireland.

---

**[1] Dublin**                                                    **Close**

Ident: http://dbpedia.org/resource/Dublin
Class: http://nerd.eurecom.fr/ontology#Location

**[2] capital**                                                    **Close**

Source: Wordnet3.0
Ident: 301467919
Class: http://nerd.eurecom.fr/ontology#City

---

*ITS: Translate*

---

This is the first paragraph. It has some `un-translatable code`[2] .

---

**[1] Smiley face**                                              **Close**

Do not translate

**[2] un-translatable code**                                     **Close**

Do not translate

---

# 6. FILE FORMAT CONVERTERS

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

## 6.1. Introduction

The file format conversion has been introduced to make the visualization of ITS metadata easier by (i) reducing the variety of input formats (HTML, XML, XLIFF) to single universal format supported by all Web browsers (HTML), and by (ii) simplifying ITS markup in output files.

The following non-trivial data conversion tasks have been resolved:

- How to transform ITS XPath expressions in case of XML-to-HTML format conversion
- How to transform and merge the external ITS rules which could also overlap each other
- Which HTML elements to use as replacements of the source XML or XLIFF elements to carry ITS markup in output HTML files
- How to support nested ITS elements in output HTML files.

## 6.2. Conversion Tasks

- **XML-to-HTML:** Conversion of ITS-decorated XML files into ITS-decorated HTML5 files for displaying the contents and ITS metadata
- **HTML-to-HTML:** Reduction of ITS markup in ITS-decorated HTML5 files to combine all external ITS rules in single list of rules within the output HTML5 file
- **XLIFF-to-HTML:** Conversion of ITS-decorated XLIFF files into ITS-decorated HTML5 files for displaying the contents of <source> and <target> XLIFF elements and related ITS metadata
- **XML-to-XLIFF:** Extraction of text segments and their associated ITS metadata and creation of output XLIFF file with equivalent ITS decoration of translation units. This application is of limited functionality due to the customer-oriented nature and purpose of XML format. The conversion is demonstrated on only one example XML file
- **XML or HTML – to – Preview HTML (Hybrid Preview):** Conversion of ITS-decorated XML or HTML files into ITS-decorated HTML5 files with preview JavaScript connected and some auxiliary JavaScript and HTML markup embedded

# 7. PROJECT DELIVERABLES

The project deliverables include the following:

- Visual designs of metadata representations
- Sample XML, HTML and XLIFF input and the output files
- Data visualization utilities, their code and the instructions for programmers on building the executable binary files from the source code; data visualization scripts
- File format conversion utilities, their code and the instructions for programmers on building the executable binary files from the source code
- The project documentation

## 7.1. Visual Designs

For the description of original visual designs, see **Chapter 5: PREVIEWING METADATA – VISUAL DESIGNS** in this document.

The set of visual markers used to highlight different ITS metadata items within HTML content is published at

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

this address:

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/img

Implemented visual designs can be evaluated by previewing the sample files in approved Web browsers.

The sample HTML files with highlighted ITS metadata referenced above in Section 5.7 "Demo Screen Shots", **THE STATIC PREVIEW (DEMO)**, are published at this address (see Content subfolder for the preview versions of the source HTML files; Markup subfolder provides the special preview versions of both XML and HTML source files). To preview any sample HTML file, open it in supported Web browser:

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/samples/static_DEMO

The sample HTML files with highlighted ITS metadata referenced above in Section 5.7 "Demo Screen Shots", **THE HYBRID PREVIEW**, are published at this address (see Content subfolder for the preview versions of the source HTML files; Markup subfolder provides the special preview versions of both XML and HTML source files). To preview any sample HTML file, open it in supported Web browser:

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/samples/hybrid

To see the metadata linked to the content in Hybrid preview mode, just click the piece of text highlighted in some color or enclosed in visual markers.

## 7.2.  Ready to Use Executables, JavaScript Files and Samples

You can check the file conversion capabilities and ITS metadata preview functionality with ready to use utilities, preview scripts and sample files listed below.

### METADATA PREVIEW

**Dynamic preview script and sample file (DEMO):**

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/exe/dynamic_DEMO

To evaluate the dynamic preview script, open the following sample file in supported Web browser:

TGT_METADATA_ITEMS_NAVIGATION_TEST2 - JS.html

To prepare any other **XHTML/HTML5** file enriched with ITS markup for dynamic previewing, apply the file converter provided for dynamic preview (see below PREVIEW PREPARATION UTILITIES). You can test the conversion on the following sample file:

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/exe/dynamic_converter/sample_input

**Dynamic preview script and CSS, file converter source code and Programmer's Guide:**

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/code/dynamic_converter

**Hybrid preview script and sample files:**

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/samples/hybrid

To evaluate the hybrid preview script and additional markup, open any of the provided sample files in

supported Web browser.

To make any other **XML/HTML** file enriched with ITS markup ready for hybrid previewing, apply the file converter provided for hybrid preview (see below PREVIEW PREPARATION UTILITIES). You can test the conversion on the following sample files:

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/input/ITS2.0_Test_Suite/HTML

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/input/ITS2.0_Test_Suite/XML

**WICS file format conversion sample files:**

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples

The output HTML files illustrate how to transfer ITS metadata from XML or XLIFF files to HTML files and how to prepare the output files for previewing in Web browsers.

**XML-to-HTML conversion results:**

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/output/ITS2.0_Test_Suite/XML

**XLIFF-to-HTML conversion results:**

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/output/ITS2.0_Test_Suite/XLIFF

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/output/WICS_samples/XLIFF-to-HTML

**HTML-to-HTML conversion (markup optimization) results:**

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/output/ITS2.0_Test_Suite/HTML

**XML-to-XLIFF conversion results (single example demo):**

> https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/output/WICS_samples/XML-to-XLIFF

> The output illustrates what ITS markup can look like in real-life scenario when you apply some user-defined fragmentation rules and proprietary XML-to-XLIFF converter.

## PREVIEW PREPARATION UTILITIES

In general, to preview HTML files enriched with ITS markup, you need at least to embed some links to external JavaScript files into these HTML files. In practice, you would also need to add some attributes and images (visual markers).

**File converter for dynamic previewing:**

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/exe/dynamic_converter

**File converter for hybrid previewing:**

https://github.com/renatb/ITS2.0-WICS-viewer/tree/master/exe/hybrid_converter

### FILE FORMAT CONVERTERS

GUI and CLI versions of WICS data conversion utilities (file format converters):

https://github.com/renatb/ITS2.0-WICS-converter/tree/master/exe

For the user guide on WICS data conversion utilities, see User_Guide.html:

https://github.com/renatb/ITS2.0-WICS-converter/tree/master/docs/html

## 7.3.   GitHub Repositories for Code, Sample Files and Documentation

The GitHub repository of WICS data conversion utilities and documentation:

https://github.com/renatb/ITS2.0-WICS-converter

The GitHub repository of WICS data visualization utilities and documentation:

https://github.com/renatb/ITS2.0-WICS-viewer

The sample input files for WICS data conversion utilities are stored in the following location:

https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/input

The expected output of WICS data conversion utilities is provided for reference and stored in the following location:

https://github.com/renatb/ITS2.0-WICS-converter/tree/master/samples/output

## 7.4.   Supported Input File Formats

The following input file formats are supported for previewing ITS metadata:

- ITS-decorated XML
- ITS-decorated HTML
- ITS-decorated XLIFF (limited support)

## 7.5.   Supported Web Browsers

**Dynamic preview (Demo):**

- Mozilla Firefox 24.0

**Hybrid preview:**

- Microsoft Internet Explorer 10.0
- Mozilla Firefox 24.0

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

## 7.6.    Visually Supported ITS Metadata Categories

The following ITS data categories can be visualized in HTML files:

**Dynamic preview (Demo):**

*Elements Within Text*
*Localization Note*
*Terminology*
*Translate*

**Hybrid preview:**

*Allowed Characters*
*Directionality*
*Domain*
*Elements Within Text*
*External Resource*
*ID Value*
*Language Information*
*Locale Filter*
*Localization Note*
*Localization Quality Issue*
*Localization Quality Rating*
*MT Confidence*
*Provenance*
*Storage Size*
*Target Pointer*
*Terminology*
*Text Analysis*
*Translate*

The following ITS data categories are NOT supported for HTML format by design:
*Preserve Space*

## 7.7.    Known Issues and Limitations

**Implementation of ITS markup for XLIFF format:** See the note on XLIFF in **Chapter 3: INPUT DATA** above in this document.

**Limited functionality of XML-to-XLIFF file format converter:** See the note on XML-to-XLIFF conversion in **Section 6.2: Conversion Tasks** above in this document.

See also Known Issues and Limitations in WICS Data Converter Programming Guide (Programming_Guide.html):

https://github.com/renatb/ITS2.0-WICS-converter/tree/master/docs/html

**Limited functionality of dynamic preview demo:** The preview script supports a limited set of ITS metadata

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

categories (see above).

## 7.8. Applicable Software Licenses

**Third-party software used in WICS Data Converters:**

Here are a list of the third-party libraries used directly in the WICS converter code, along with reference URIs and license information. Most Perl modules are released under the Perl 5 artistic license. None of the licenses have restrictions against commercial use.

| Library | URI | License |
| --- | --- | --- |
| Perl | perl.org | Perl 5 Artistic |
| XML::LibXML | https://metacpan.org/module/XML::LibXML | Perl 5 Artistic |
| LibXML | http://www.xmlsoft.org/ | MIT |
| Wx | https://metacpan.org/module/Wx | Perl 5 Artistic |
| wxWidgets | http://www.wxwidgets.org | wxWindows |
| Wx::Perl::Packager | http://metacpan.org/module/Wx::Perl::Packager | Perl 5 Artistic |
| Dist::Zilla | dzil.org | Perl 5 Artistic |
| Par::Packer | https://metacpan.org/module/PAR::Packer | Perl 5 Artistic |
| HTML::HTML5::Parser | https://metacpan.org/module/HTML::HTML5::Parser | Perl 5 Artistic |
| HTML::HTML5::Writer | https://metacpan.org/module/HTML::HTML5::Writer | Perl 5 Artistic |
| Test::XML | https://metacpan.org/module/Test::XML | Perl 5 Artistic |
| Log::Any | https://metacpan.org/module/Log::Any | Perl 5 Artistic |
| HTTP::Message | https://metacpan.org/module/HTTP::Message | Perl 5 Artistic |
| IO::Compress::Gzip | https://metacpan.org/module/IO::Compress::Gzip | Perl 5 Artistic |
| Path::Tiny | https://metacpan.org/module/Path::Tiny | Apache 2.0 |
| Test::HTML::Differences | https://metacpan.org/module/Test::HTML::Differences | Perl 5 Artistic |
| Getopt::Lucid | https://metacpan.org/module/Getopt::Lucid | Apache 2.0 |
| Try::Tiny | https://metacpan.org/module/Try::Tiny | MIT |
| Data::Section::Simple | https://metacpan.org/module/Data::Section::Simple | Perl 5 Artistic |
| LWP::UserAgent | https://metacpan.org/module/LWP::UserAgent | Perl 5 Artistic |
| autodie | https://metacpan.org/module/autodie | Perl 5 Artistic |
| Test::Base | https://metacpan.org/module/Test::Base | Perl 5 Artistic |
| Test::Exception | https://metacpan.org/module/Test::Exception | Perl 5 Artistic |
| Test::NoWarnings | https://metacpan.org/module/Test::NoWarnings | GNU Library |
| File::Slurp | https://metacpan.org/module/File::Slurp | Perl 5 Artistic |
| Test::Warn | https://metacpan.org/module/Test::Warn | Perl 5 Artistic |
| URI | https://metacpan.org/module/URI | Perl 5 Artistic |

**Third-party software used in WICS data visualization utilities:**

JavaScript files for dynamic preview (demo) and for hybrid preview:

| Library | URI | License |
| --- | --- | --- |
| jquery-1.9.1.min.js | http://jquery.com | MIT License |

File format converter for hybrid preview:

| Library | URI | License |
| --- | --- | --- |
| SgmlReaderDll.dll | https://github.com/MindTouch/SGMLReader | Apache License, Version 2.0 |

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

# 8. ITS -RELATED PROMOTIONAL ACTIVITIES

**MultilingualWeb W3C Workshop: Making the Multilingual Web Work - ITS 2.0 Showcase**

Rome, 12.03.2013

As an integral part of the program, the Workshop featured a showcase of implementations of the Internationalization Tag Set 2.0 (ITS 2.0) specification:

http://www.multilingualweb.eu/documents/rome-workshop/rome-showcase

Renat Bikmatov (Logrus) presented poster-presentation "Universal Preview of ITS 2.0 Metadata in XML, XLIFF, and HTML Files" including live demonstration of ITS data visualization:

http://www.w3.org/International/multilingualweb/rome/posters/mlw-lt_rome2013poster-12.pdf

**FEISGILT 2013 - ITS 2.0 / LIS Track**

London, 11-12.06.2013

Renat Bikmatov (Logrus) presented "Practical visualization of ITSv2.0 categories for real world localization process":

http://www.w3.org/International/its/wiki/File:Practical_visualization_of_ITS_2.0_categories_for_real_world_localization_process.ppt

See also:

http://www.localizationworld.com/lwlon2013/feisgiltt/accepted.html

http://www.w3.org/International/its/wiki/Outreach_material#FEISGILTT_2013

**META-FORUM 2013 - ITS 2.0 Implementation Basket**

Berlin, 19-20.09.2013

Serge Gladkoff (Logrus) presented "Preview of ITS 2.0 Metadata in Web Browsers" at the "Technologies for the Multilingual Web" session and reported on "The WICS project" at META Exhibition.

# 9. FURTHER PROJECT DEVELOPMENT

**Extended testing of WICS utilities:**

File Format Converters and Metadata Preview utilities and scripts have been tested on W3C Test Suite files and a few sample files provided by Logrus to illustrate some real-life metadata usage scenarios. However, these samples do not cover all frequently used scenarios; more extended testing and debugging needed, as well as the feedback from any potential users.

**Implementation of global navigation through the content in hybrid preview mode:**

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)

The functionality of hybrid preview script needs to be extended to support the same kind of global navigation through the content, as in dynamic preview demo: the navigation through the fragments of content.

**Implementation of synchronization between the preview in production environment and the preview in Web browser:**

The navigation through the fragments of the content in production environment (e.g. CAT tool) and the navigation through the pieces of the content in preview window in Web browser need to be synchronized and automated: as soon as you open a translation unit in CAT tool, you should be able to automatically preview the same fragment of the content in "mirror" ITS-enriched HTML file opened in Web browser.

**Full-scale dynamic visualization of ITS metadata:**

The direct dynamic visualization of ITS metadata embedded in HTML or XLIFF files by means of connected JavaScript only (without any additional markup) would be a more general and complicated solution than the static or hybrid visualization of ITS metadata in derivative HTML files.

**Enhanced support of XLIFF format:**

Since W3C recommendations on integration of ITS metadata markup with XLIFF format and the scope of implementation of some ITS syntax features in XLIFF are works in progress, and the task of visualization of ITS metadata in XLIFF 1.2 and XLIFF 2.0 content is still an open issue.

**ITS metadata editor – new dimension of ITS metadata support in Web browsers:**

After implementation of metadata preview functionality, the next step would be implementation of metadata editing functionality in the same Web browser-based environment so that the language workers could add or update metadata related to their work and share this information through the whole production chain. Implementation of a localization metadata exchange environment would enable implementers to build the complete life cycle of metadata, which would be synchronized with the life cycle of the content.

**Automated content analyzer– natural language processing technology to apply metadata markup to the content:**

Once you have developed a manual editor capable of embedding ITS metadata markup into your ("your" means "customer-specific") content and linking your payload data to that markup, it's time to get started with development of automated Natural Language Processing tools which would help you to analyze your content against your metadata applicability rules and parse the content.

In general, the implementation of ITS metadata will drive the change in localization paradigm. In the classic paradigm, the localization instructions, glossaries and other reference materials are kept separated from the content. Now you can set metadata-level relations between different pieces of content and all those pieces of context information.

However, to get the most out of such symbiosis of content and metadata, the glossaries, instructions and reference data should be prepared and converted into the formats which would be compatible with ITS markup. The second prerequisite is the development of NLP tools, especially fuzzy searching tools, to match metadata with respective content.

HTML 5 - ITS 2.0 IMPLEMENTATION PROJECT: WORK IN CONTEXT SYSTEM (WICS)