**ITS Tools Annotation**

In some cases, it may be important for instances of data categories to be associated with the processor that generated them. For example, the score of the MT Confidence data category is meaningful only when the consumer of the information also knows what MT engine produced it, because the score provides the relative confidence of translations from the same MT engine but does not provide a score that can be reliably compared between MT engines.

ITS provides uses the processInfo element to store information on processors that generated data category annotations independently from the data categories..

The attribute toolsRefs is used to associate a given tool that generated a given data category with the content of the element where information is annotated with that data category.

The value of toolRefs is a space-separated list of references where each reference is composed of two parts: a data category identifier and a URI pointing to the relevant toolInformation element. These two part are separated by the “|”, [??] character.

The data category identifier can be either: one of the pre-defined identifiers, or a user-defined value with a prefix.

The URI pointing to a tool for a given data category is overridden only when a new toolsRef attribute is defined with a new URI for the same data category.

Document:

<doc its:toolsRef="MTConfidence|file:///tools.xml#T1" xlmns:its="<http://www.w3.org/2005/11/its>">

 <p its:mtConfidenceScore="0.78">Text translated with tool T1</p>

 <p its:mtConfidenceScore="0.55">Text also translated with tool T1</p>

 <p its:mtConfidenceScore="0.34" its:toolRefs="mtConfidence|file:///tools.xml#T2">Text translated with tool T2</p>

</doc>

Separate document with the list of tools (tools.xml):

<its:processInfo>

 <its:toolInfo xml:id="T1" dataCategory="MTConfidence">

 <its:toolName>Bing Translator</its:toolName>

 <its:toolVersion>123</its:toolVersion>

 <its:toolValue></its:toolValue>

 <its:toolInfo>

 <its:toolInfo xml:id="T2" dataCategory="MTConfidence">

 <its:toolName>myMT</its:toolName>

 <its:toolVersion>456</its:toolVersion>

 <its:toolValue>FR-to-EN-General</its:toolValue>

 <its:toolInfo>

<its:processInfo>

### 6.20 MT Confidence

#### [Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22contents)6.20.1 Definition

The [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category is used to communicate the self-reported confidence of a machine translation engine in the accuracy of a translation it has provided. It is not intended to provide a comparable between machine translation engines and platforms. It is solely for providing self-reported confidence by the specific system that produced the actually used raw machine translation. This data category does NOT aim to establish any sort of correlation between the self-reported confidence and either human evaluation of MT usefulness, or post-editing cognitive effort. For harmonization’s sake, MT Confidence is provided as a (rational) number from the interval <0;1>.

**Note:**

Implementers are expected to interpret the floating point number and present it to human and other consumers in other convenient forms, such as percentage (0-100%) with up to 2 decimal digits, font or background color coding etc.

This data category can be used for several purposes, including, but not limited to:

* Automated prioritising of raw machine translated text for further processing based on empirically set thresholds.
* Providing readers, translators, post-editors, reviewers and proofreaders of machine translated text with self-reported relative accuracy prediction.

MT confidence scores can be displayed e.g.on websites machine translated on the fly, by simple web-based translation editors or Computer Aided Translation (CAT) tools.

#### [Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22contents)6.20.2 Implementation

The [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category can be expressed locally on individual elements. For elements, the data category information [inherits](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22def-inheritance) to the textual content of the element, including child elements, but excluding attributes.

LOCAL: the following local markup is available for the [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category:

* A mtConfidence attribute with a value that is a rational number in the interval 0 to 1.

Example 95: The [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category expressed locally

**<text** xmlns:its="http://www.w3.org/2005/11/its"**>**

 **<body>**

 **<p>**

 **<span** its:mtConfidence="0.8982"**>**Dublin is the capital city of

 Ireland.**</span>**

 **</p>**

 **</body>**

**</text>**

[Source file: [examples/xml/EX-mtConfidence-local-1.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-mtConfidence-local-1.xml)]

Example 96: The [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category expressed locally in HTML5

**<!DOCTYPE html>**

**<html** lang=en**>**

 **<head>**

 **<meta** charset=utf-8**>**

 **<title>**Sentences about Dublin and Prague MTed from Czech with mtConfidence locally.**</title>**

 **</head>**

 **<body>**

 **<p>**

 **<span** its-mt-confidence =0.8982 **>** Dublin is the capital of Ireland.**</span>**

 **<span** its-mt-confidence =0.8536 **>** The capital of the Czech Republic is Prague.**</span>**

 **</p>**

 **</body>**

**</html>**

[Source file: [examples/html5/EX-mtConfidence-html5-local-1.html](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/html5/EX-mtConfidence-html5-local-1.html)]

### [o to the table of contents.](%22%20%5Cl%20%22contents)6.13 Text Analysis Confidence

#### [o to the table of contents.](#contents)6.13.1 Definition

The **Text Analysis Confidence** data category is used to communicate the self reported confidence of a text analytics engine accuracy of an annotation it has provided. It is not intended to serve as a comparison metric between TA engines, but solely for providing self-reported confidence by the specific system that produced the annotations. This data category does NOT aim to establish any sort of correlation between the self-reported confidence and either human evaluation of text annotation usefulness.

This data category can be used for several purposes, including, but not limited to:

* Automated prioritisation of annotations for further processing based on a quality threshold.
* Providing readers, translators, post-editors, reviewers and proofreaders of machine translated text with self-reported relative accuracy prediction.

This data category provides aTA Confidence is provided as a rational number from the interval <0;1> or as percentage (0-100%) with up to 2 decimal digits.

The TA source and confidence can be displayed on websites on the fly, by simple content management systems. TA Confidence MAY be displayed for human consumers as segment annotation or as color-coded font or background.

The scope of TA Annotation covers ITS 2.0 data categories that could be expected to come from an automatic text analysis source, such as Disambiguation Identity, Disambiguation EntityType, Terminology or Domain.

#### [o to the table of contents.](#contents)6.15.2 Implementation

The TA Annotation data category applies solely to documents, elements, and attributes within the ITS 2.0 domain. The information applies to the ITS2.0 metadata content of the element,and not any content or non-ITS 2.0 attributes or elements.

For elements, the data category information [inherits](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22def-inheritance) to the annotation of textual content of the element, including child elements, but excluding attributes..

LOCAL: The following local markup is available for the Text Analysis Annotation data category:

* A taConfidence attribute with a value that is a rational number in the interval 0 to 1.

Example ###: The TA Annotation data category expressed locally

<body>
 <p><span its:textAnalysisAgentRef=”http://enrycher.ijs.si” its:taConfidence=”0.95”

Its:disambigType=”entity”

its:entityTypeRef=”[http:/nerd.eurecom.fr/ontology](http://nerd.eurecom.fr/ontology)#Place”>Dublin</span> is the capital city of Ireland.</p>
 </body>
</text>

[Source file: TBS]

Example 32: The [TA Annotatio](#locNote-datacat)n data category expressed locally in HTML5

<html lang="en">
 <head>
 <meta charset="utf-8" />
 <title>Entity: Local Test</title>
 </head>
 <body >
 <p><span
its-entity-type-source-ref="<http://nerd.eurecom.fr/ontology>"
its-entity-type-ident-ref="[http:/nerd.eurecom.fr/ontology](http://nerd.eurecom.fr/ontology)#Place"
its-disambig-source-ref="<http://dbpedia.org/>”
its-disambig-ident-ref="<http://dbpedia.org/resource/Dublin>”
its-disambig-type=”entity”

its-ta-confidence =”0.95”

>Dublin</span> is the capital of Ireland.</p>
 </body>
</html>

 [Source file: TBS]

**Impact on Data Categories**

Here we take the examples from certain data categories to show how they would work with ITS, using the example above, which is extended with its:toolInfo element that matches the content from those examples. We keep the location of the its:toolInfo the same. We use the single element below in all examples

<its:processInfo>

<its:toolInfo xml:id="T3" dataCategory="MTConfidence">

 <its:toolName>Bing Translator</its:toolName>

 <its:toolValue>en-t-cs</its:toolValue>

 <its:toolInfo>

<its:toolInfo xml:id="T4" dataCategory="MTConfidence">

 <its:toolName>vanilla Moses</its:toolName>

 <its:toolValue>medical:EN-ES\_LA</its:toolValue>

 <its:toolInfo>

<its:toolInfo>

<its:toolInfo xml:id="T5" dataCategory="Text Analysis Confidence">

 <its:toolName> http://enrycher.ijs.si </its:toolName>

 <its:toolInfo>

**MT Confidence score**

</its:processInfo>

Copy of Example 95: The [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category expressed locally

**<text** xmlns:its=<http://www.w3.org/2005/11/its>

its:toolsRef="mtConfidence|file:///tools.xml#T3"**>**

 **<body>**

 **<p>**

 **<span** its:mtConfidence="0.8982"**>**Dublin is the capital city of

 Ireland.**</span>**

 **</p>**

 **</body>**

**</text>**

[Source file: [examples/xml/EX-mtConfidence-local-1.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-mtConfidence-local-1.xml)]

Copy of Example 96: The [MT Confidence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html%22%20%5Cl%20%22mtconfidence) data category expressed locally in HTML5

**<!DOCTYPE html>**

**<html** lang=en**>**

 **<head>**

 **<meta** charset=utf-8**>**

 **<title>**Sentences about Dublin and Prague MTed from Czech with mtConfidence locally.**</title>**

 **</head>**

 **<body** its-tools-ref="its-mt-Confidence|file:///tools.xml#T4"**>**

 **<p>**

 **<span** its-mt-confidence=0.8982 **>** Dublin is the capital of Ireland.**</span>**

 **<span** its-mt-confidence-score=0.8536 **>** The capital of the Czech Republic is Prague.**</span>**

 **</p>**

 **</body>**

**</html>**

Example ###: The TA Annotation data category expressed locally

<body its:toolsRef="taConfidence|file:///tools.xml#T5">
 <p><span its:taConfidence=”0.95”

Its:disambigType=”entity”

its:entityTypeRef=”[http:/nerd.eurecom.fr/ontology](http://nerd.eurecom.fr/ontology)#Place”>Dublin</span> is the capital city of Ireland.</p>
 </body>
</text>

[Source file: TBS]

Example 32: The [TA Annotatio](#locNote-datacat) data category expressed locally in HTML5

<html lang="en">
 <head>
 <meta charset="utf-8" />
 <title>Entity: Local Test</title>
 </head>
 <body its:toolsRef="its-ta-confidence|file:///tools.xml#T5">
 <p><span
its-entity-type-source-ref="<http://nerd.eurecom.fr/ontology>"
its-entity-type-ident-ref="[http:/nerd.eurecom.fr/ontology](http://nerd.eurecom.fr/ontology)#Place"
its-disambig-source-ref="<http://dbpedia.org/>”
its-disambig-ident-ref="<http://dbpedia.org/resource/Dublin>”
its-disambig-type=”entity”

its-ta-confidence =”0.95”

>Dublin</span> is the capital of Ireland.</p>
 </body>
</html>

[Source file: [examples/html5/EX-mtConfidence-html5-local-1.html](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/html5/EX-mtConfidence-html5-local-1.html)]