6.10 Disambiguation

*[Ed. note: This data category is not completely stable yet.]*

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)6.10.1 Definition**

The [Disambiguation](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "Disambiguation) data category is used to indicate occurrences of specific concepts that may require special handling in the localization of the document.

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

* Informing translation systems that this fragment of text may not be literally translated, but subject to specific proper name translation rules or official translations, as well as a very specific meaning of the phrases.
* Informing content management and translation systems about the type of the underlying entity in order to enable processing based on a specific type of the target, for example, when handling personal names, product names or geographic names, chemical compounds, protein names and similar.

Disambiguation is achieved by associating a selected fragment of text with an external web resource that can be referenced by a translation or linguistic review agent in order to access the correct meaning or lexical use of the text and thereby informing its translation.

A fragment of text can be disambiguated at different granularities, i.e. as a lexical concept, as an ontology concept, or as a named entity.

As a lexical concept, the external reference can provide synonyms and example usage, e.g. using service such as Wordnet.

As an ontology concept, the external reference can provide a formal conceptual definition within a framework of related concepts.

As a named entity, the external reference can provide a description of the real world entity the text intends to convey. For instance, the word 'City' in 'I am going to the City' may be disambiguated in one of the WordNet synsets that can be represented by 'city', an ontology concept of a City that could represent a subclass of a “PopulatedPlace” in the conceptual granularity level, or the central area of a particular city, e.g. City of London, as interpreted in the entity granularity level. Linked data network, such as DBpedia, increasing interlink ontological and named entity definitions for the same things as authored in different languages, offering a mechanism to locate translations from the source language description.

Two types of disambiguation are possible:

* Disambiguation for target type class, which describes the type class of the underlying concept or entity of the fragment.
* Disambiguation for target identity, which describes the actual underlying external resource that conveys the intended meaning of the fragment.

Text analysis engines, such as named entity recognizers, named entity, concept and word sense disambiguation components can offer an easy way to create this information. Content management tools can present and visualize this information or use it to index their content. Machine translations systems may use it for training and translation when dealing with proper names and edge cases.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**6.10.2 Implementation**

The [Disambiguation](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "Disambiguation) data category can be expressed with global rules, or locally on an individual element. The information applies to the textual content of the element. There is no inheritance.

*[Ed. note: Below will need a test case in the test suite.]*

When using disambiguation specifying the target identity, the user [MUST](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rfc2119) use only one of the two addressing modes:

* Using disambigSource and one of disambigIdent or disambigIdentPointer (at a global rule) to specify the collection and the identifier itself.
* Using one of disambigIdentRef or disambigIdentRefPointer (at a global rule) using a IRI for the disambiguation target.

GLOBAL: The disambiguationRule element contains the following:

* A required selector attribute that contains an [absolute selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selectors) which selects the nodes to which this rule applies.
* At least one of the following:
  + To specify the target type class, exactly one of the following:
    - A disambigClassPointer attribute that contains a [relative selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selectors) pointing to a node specifying the entity type class behind the selector.
    - A disambigClassRefPointer attribute that contains a [relative selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selectors) pointing to a node that holds a IRI that specifies the entity type class behind the selector.
  + To specify the target identity, exactly one of the following:
    - A disambigIdentPointer attribute that contains a [relative selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selectors) pointing to a node that represents a unique identifier for the disambiguation target.
    - A disambigIdentRefPointer attribute that contains a [relative selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selectors) pointing to a node that holds a IRI that represents a unique identifier for the disambiguation target.

For an example, see [Example 54](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#EX-disambiguation-html5-rdfa-companion-document).

LOCAL: The following local markup is available for the [Disambiguation](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#Disambiguation) data category:

* An optional disambigConfidence attribute with the value of a rational number in the interval 0 to 1. This represents the confidence of the agents producing the annotation that the union of the values for the other disambiguation attributes in this instance are accurate. 1 represents the highest level of confidence.
* An optional disambigGranularity attribute that contains a string, specifying the granularity level of the disambiguation. The value can be one of the following identifiers: lexicalConcept, ontologyConcept, or entity
* At least one of the following:
  + To specify the target type class:
    - A disambigClassRef attribute that contains a IRI, specifying the type class of the concept or entity behind the selector.
  + To specify the target identity, exactly one of the following:
    - When addressing [mode 1](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#disambiguation-use-cases):
      * A disambigSource attribute that contains a string representing the disambiguation identifier collection source.
      * A disambigIdent attribute that contains a string, representing the disambiguation identifier for the disambiguation target that is valid within the specified disambiguation source.
    - When addressing [mode 2](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#disambiguation-use-cases):
      * A disambigIdentRef attribute that contains a IRI that represents a unique identifier for the disambiguation target.

Note: Any node selected by the disambiguation data category with the disambigConfidence attribute specified MUST be contained in an element with the its:toolsRef (or in HTML5 its-tools-ref) attribute specified for the disambiguation data category.

Example 52: Local mixed usage of Usage of disambigClassRef, disambigGranularity, and disambigIdentRef in HTML.

**<!DOCTYPE html>**

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

**<head>**

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

**<title>**Disambiguation: Local Test**</title>**

**</head>**

**<body>**

**<p><span**

its-disambig-class-ref="http://nerd.eurecom.fr/ontology#Place"

its-disambig-ident-ref="http://dbpedia.org/resource/Dublin"

its-disambig-granularity="entity"

its-disambig-confidence=”0.7”**>**Dublin**</span>**

is the **<span**

its-disambig-source="Wordnet3.0"

its-disambig-ident="301467919"

its-disambig-granularity="lexicalConcept"

its-disambig-confidence=”0.5”**>**capital**</span>** of Ireland.**</p>**

**</body>**

**</html>**

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

**Note:**

For referring to disambigClassRef values, implementors are encouraged to use an existing repository of entity types as long as they satisfy their requirements. For example, the Named Entity Recognition and Disambiguation ontology [[NERD]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#nerd).

Furthermore, valid target types depend on the disambiguation granularity: types of entities are distinct from types of lexical concepts or ontology concepts. While this distinction exists, the specification does not prescribe a way of automatically inferring a disambiguation level from a target type.

When serializing the ITS mark-up in HTML5, the preferred way is to serialize in RDFa Lite or Microdata due to the existing search and crawling infrastructure that is able to consume this kind of data.

Example 53: Local mixed usage of entityTypeSourceRef, enttiyTypeRef, disambigSourceRef, disambigIdentRef in HTML+RDFa Lite.

See [Example 54](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "EX-disambiguation-html5-rdfa-companion-document) for the companion document with the mapping data.

**<!DOCTYPE html>**

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

**<head>**

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

**<link** href=EX-disambiguation-html5-rdfa.xml rel=its-rules**>**

**<title>**Entity: Local Test**</title>**

**</head>**

**<body>**

**<p><span** property=http:**//**xmlns.com/foaf/0.1/name about=http://dbpedia.org/resource/Dublin typeof=http:/nerd.eurecom.fr/ontology#Place>Dublin**</span>** is

the capital of Ireland.**</p>**

**</body>**

**</html>**

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

Example 54: Companion document, having the mapping data for [Example 53](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "EX-disambiguation-html5-rdfa).

**<its:rules** xmlns:its="http://www.w3.org/2005/11/its" version="2.0"**>**

**<its:disambiguationRule** selector="//\*[@typeof]" disambiguationClassRefPointer="@typeof"**/>**

**<its:disambiguationRule** selector="//\*[@about]" disambigIdentRefPointer="@about"**/>**

**</its:rules>**

[Source file: [examples/html5/EX-disambiguation-html5-rdfa.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/html5/EX-disambiguation-html5-rdfa.xml)]