

XML Text Memory

Table of contents

- [About xml:tm](#)
- [Why use xml:tm?](#)
- [History](#)
- [xml:tm specification](#)

About xml:tm



xml:tm (XML-based Text Memory) is the vendor-neutral open XML standard for embedding text memory directly within an XML document using XML namespace syntax. xml:tm leverages the namespace syntax of XML to embed text memory information within the XML document itself.

At the core of xml:tm is the concept of "text memory". Text memory comprises two components:

1. **Author Memory.** The XML document is segmented and a full history of all segments and revisions is maintained in the XML document itself.
2. **Translation Memory.** When an xml:tm namespace document is ready for translation the namespace itself specifies the text that is to be translated. The tm namespace can be used to create an [XLIFF](#)-format document for translation. xml:tm allows for much more focused and better defined translation memory matching than is possible using standard TM technology. In particular, it includes the following:
 - *Exact Matching.* Author memory provides exact details of any changes to a document. Where text units have not been changed for a previously translated document xml:tm provides the basis for declaring an "Exact match" with the previously translated target language document.
 - *In-Document Leveraged Matching.* xml:tm can also be used to find in-document leveraged matches
 - *Database Leveraged Matching.* When an xml:tm document is translated the translation process provides perfectly aligned source and target language text units. These can be used to create traditional translation memories.
 - *In-Document Fuzzy Matching.* The text units contained in the leveraged memory database can also be used to provide fuzzy matches of similar previously translated text from within the same document.
 - *Fuzzy Matching.* The text units contained in the leveraged memory database can also be used to provide fuzzy matches of similar previously translated text.
 - *Non-Translatable Text.* Text units that are made up solely of numeric, alphanumeric, punctuation or measurement items can be identified during authoring and flagged as non translatable, thus reducing the translation count

metrics.

To learn more about xml:tm, please read "[Using XML technology to reduce the cost of authoring and translation](#)", and "[How to Leverage the Maximum Potential of XML for Localization](#)" by Andrzej Zydrón.

Why use xml:tm?

xml:tm can greatly simplify XML-based globalization workflows. Among its benefits are:

1. **Integration of linguistic assets with content means TM is always up to date.** Because translation and author memories are included directly in the document, you don't have to worry about whether TM matches really apply or not and any updates to the translation are automatically reflected in the TM so there is no need for a separate TM database to update: relevant xml:tm-format documents can be loaded on the fly to create memories that are always perfectly up to date.

2. **xml:tm works seamlessly with DITA.** DITA offers tremendous advantages to document creators by breaking content down into reusable chunks. xml:tm goes one step further and breaks these chunks down into segments in order to help globalization processes. Because xml:tm is an XML namespace, it can be integrated not only into DITA, but into almost any XML format, thus saving time and money and eliminating the problems that often arise from converting XML to other formats for translation.

3. **xml:tm was designed to integrate seamlessly with other standards**, so if you are already working in a standards-based environment, xml:tm will fit right in. xml:tm works with the following standards:

- **[SRX \(Segmentation Rules eXchange\)](#)**. xml:tm mandates the use of SRX for text segmentation of paragraphs into text units.
- **[Unicode Standard Annex #29-9](#)**. xml:tm mandates the use of Unicode Standard Annex #29 for tokenization of text into words.
- **[XLIFF 1.2](#)**. xml:tm mandates the use of XLIFF for the actual translation process. xml:tm is designed to facilitate the automated creation of XLIFF files from xml:tm enabled documents, and after translation to easily create the target versions of the documents.
- **[GMX-V \(Global Information Management Metrics eXchange - Volume\)](#)**. xml:tm mandates the use of GMX-V for all metrics concerning authoring and translation.
- **[TMX \(Translation Memory eXchange\)](#)**. xml:tm facilitates the easy creation of TMX documents, aligned at the sentence level.
- **[DITA \(Darwin Information Technology Architecture\)](#)**. xml:tm complements the DITA standard by allowing text reuse at the sentence level within DITA documents.
- **[W3C ITS](#)**. xml:tm mandates the use of W3C ITS Document Rules for identifying translatable text within an XML document as well as W3C ITS Best Practices with regard to XML document localization.

History

xml:tm 1.0 was adopted by OSCAR as an official standard for the globalization industry on

February 26, 2007.

xml:tm specification

xml:tm can be downloaded ([ZIP file](#)) or [viewed online in XHTML format](#).