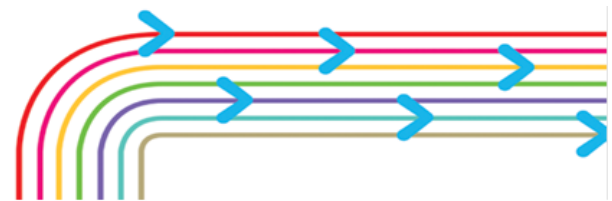




Better Translation Technology

XTM for Language Service Providers Explained



XTM For Language Service Providers Explained

Documentation for XTM for LSPs Explained

Published by XTM International Ltd.

© Copyright XTM International Ltd. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying, without the written permission of XTM International Ltd.

Updated January 2012



XTM-International Ltd, PO Box 2167, Gerrards Cross, SL9 8XF, UK
Tel.: +44 (0)1753 480479 email: sales@xtm-intl.com <http://www.xtm-intl.com>

Table of Contents

Table of Contents.....	3
1. Introduction.....	4
2. Open Architecture.....	6
3. Open Standards.....	7
5. Advanced Translation Technology.....	10
6. Web 2.0 based.....	10
7. Sub-contracting.....	12
8. Content Management Systems Vendors.....	12
9. Organizations with translation requirements.....	12
10. Conclusion.....	12

XTM For Language Service Providers Explained

1. Introduction

There is a new generation of Computer Assisted Translation (CAT) tools available based on the latest Web 2.0 technology. These systems are more open, cheaper, more efficient and easier to use, and much quicker to implement than previous generation desktop-based CAT systems.

XTM is a Highly Scalable, Open Architecture, Open Standards, Open Source, advanced Translation Management (TMS) and CAT tool from XTM-INTL. XTM is designed on SOA (Service Oriented Architecture) principles as a Web 2.0 suite of modular components that address the direct needs of the following types of customer:

- Localization Service Providers
- Content Management Systems Vendors
- Organizations with localization requirements

The user interface of XTM is completely browser-based, including the fully featured revolutionary translation editor. This type of architecture is ideal for the highly collaborative nature of the translation process, which requires that Project Managers, Translators, Reviewers and Correctors and all work together effectively sharing all linguistic assets in real time.

This modular, highly flexible approach affords XTM significant advantages. In addition, XTM uses XML as the native format for all text and data. If the source files are not in XML, then they are converted to XML during processing and then back again into the original format on completion. This provides substantial benefits in terms of elegant and effective processing based on an Open Standards format: faster, more efficient and accurate. XTM is completely in tune with XML. Rather than a multiple disparate filters, XTM has one highly tuned extraction and matching engine driven by XML standards-based parameter files.

One of the biggest problems facing Localization Service Providers is how to stay competitive. Traditional desktop-based CAT tools are inherently expensive to run and maintain. They offer very little in the way of process automation, apart from providing some benefit in the form of translation memory for updated documents that have previously been translated. Moreover, running and maintaining a translation project using these tools is very labour-intensive, and leads to the creation of little islands of translation that are not shared between translators working on the same language:

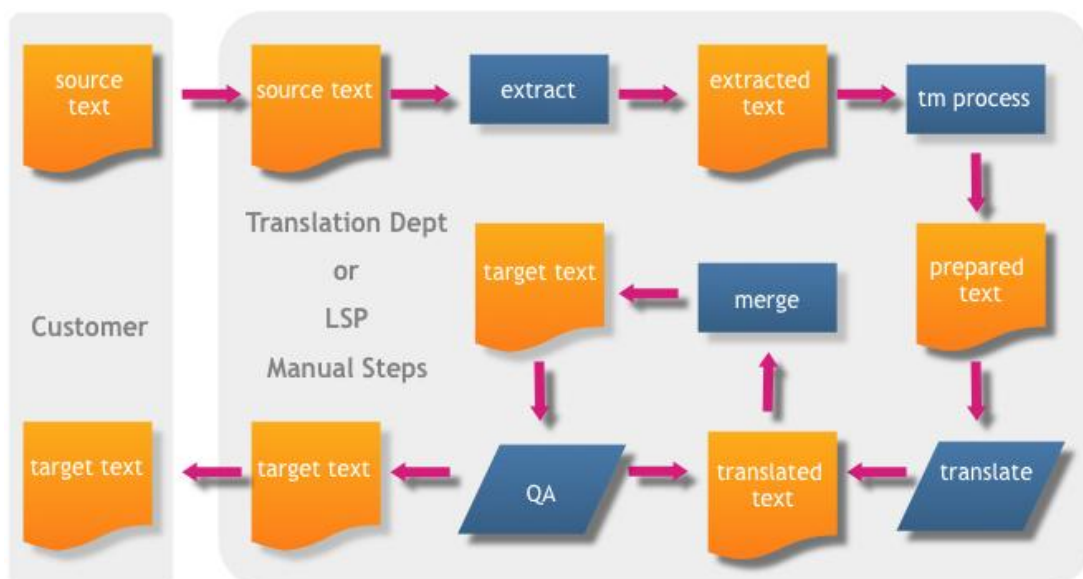
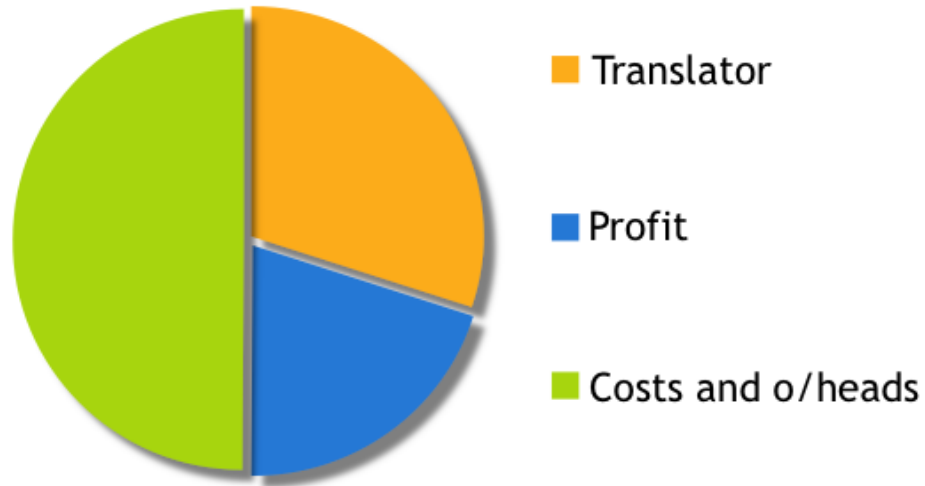


Figure 1: Traditional Translation Workflow

XTM For Language Service Providers Explained

The net effect of this way of working is that half of the cost of translation is taken up by project management and manual data handling:



Source Professor Reinhard Schäler LRC - ASLIB 2002

Figure 2: True Cost of Translation

XTM completely breaks this mould. It uses the power of modern software techniques and architecture to free up the desktop, and to provide a high degree of automation and sharing. This eliminates most of the costs and overheads associated with running a translation project:

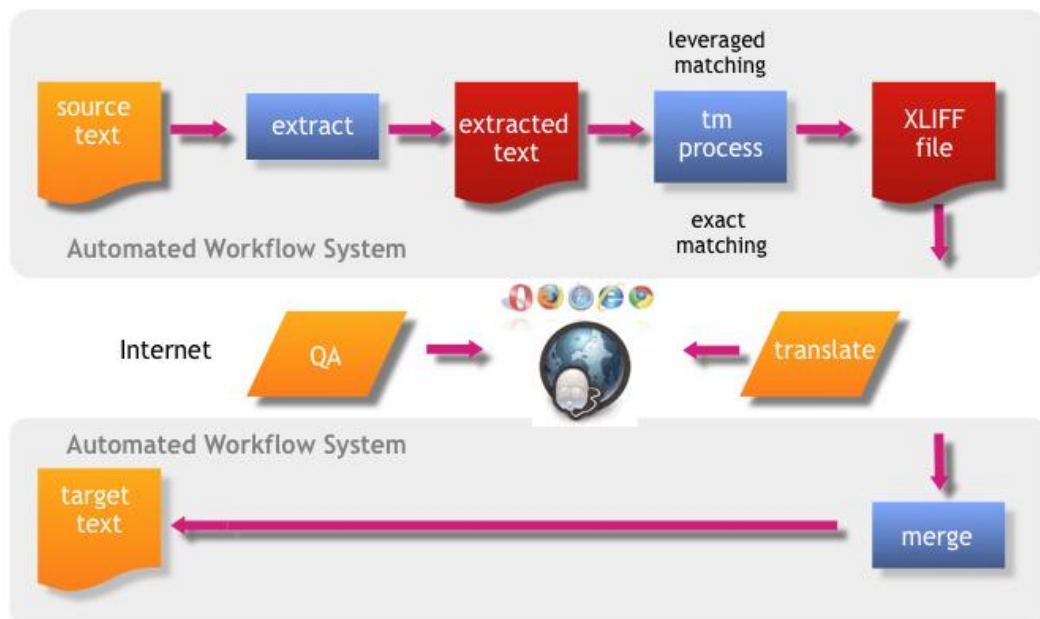


Figure 3: OAXAL Automated Workflow

2. Open Architecture

XTM comprises the following modular components, based on SOA (Service Oriented Architecture) principles:

1. XTM Engine: The Translation Memory (TM) engine, including Statistical Machine Translation
2. XTM Workbench: The Translation editor
3. XTM QA: The QA Manager responsible for quality assurance and spell checking, including dictionary management
4. XTM Open APIs: The open API access to all XTM modules
5. XTM TM Manager: The Translation Memory Management module
6. XTM Terminology: The Terminology Module
7. XTM Workflow: The Translation Workflow Module
8. XTM Portal: A web-based portal for translation customers for online ordering



Figure 4: XTM Modules

The whole concept behind the architecture of XTM is to ensure that the individual modules can be easily integrated with existing systems. The XTM modules are all written in Java, which ensures cross-platform compatibility. The database layer of XTM is handled via an abstraction layer, which means that XTM can work seamlessly with MySQL, MS SQL Server, PostgreSQL and Oracle without any problems. XTM supports Linux, Windows, MacOS and Solaris platforms.

The XTM Engine is accessible via Web Services. All the other XTM modules are accessible via Web Services. XTM has been successfully integrated with .Net, Java, PHP and C/C++ systems. The use internally of the OASIS XLIFF (XML Localization Interchange File Format) as the translation format also means that XTM can exchange data freely with tools that support XLIFF natively, of which there is a steadily increasing number.

3. Open Standards

XTM is based on the new OASIS Open Architecture for XML Authoring and Localization (OAXAL) reference model, comprising the following Open Standards:

1. Unicode
2. XLIFF - OASIS XML Localization Interchange File Format
3. SRX – ETSI LIS (formerly LISA OSCAR) Segmentation Rules
4. TMX – ETSI LIS (formerly LISA OSCAR) Translation Memory Exchange
5. GMX/V – ETSI LIS (formerly LISA OSCAR) Word and Character Count Standard
6. TBX – ETSI LIS (formerly LISA OSCAR) Term Base Exchange
7. W3C ITS – Internationalization Tag Set
8. xml:tm – ETSI LIS (formerly LISA OSCAR) based text memory

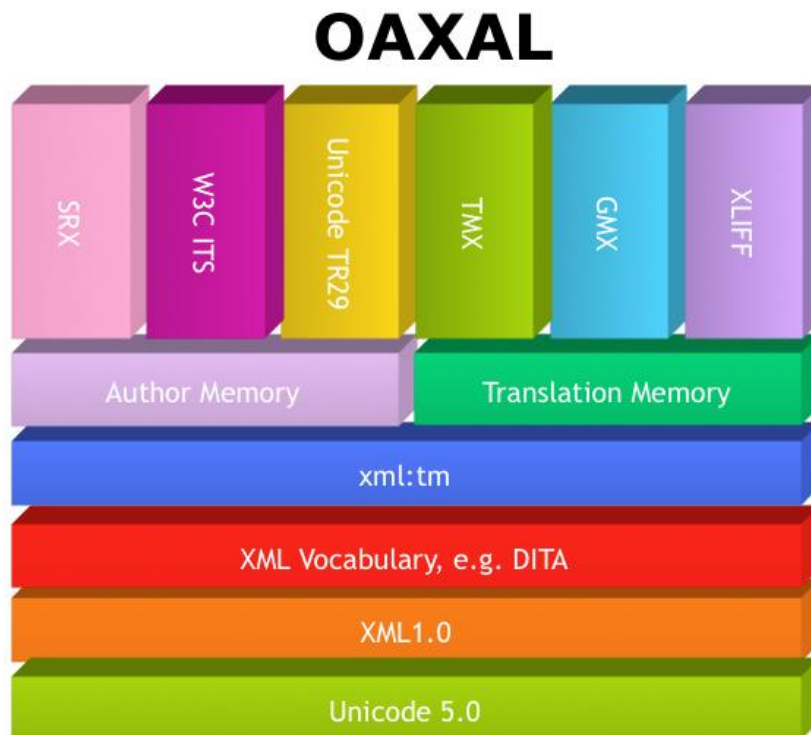


Figure 5: XTM provides a complete OAXAL implementation

The benefits of Open Standards are that they not only provide good engineering practice, but also allow for the easy interchange with other compliant systems.

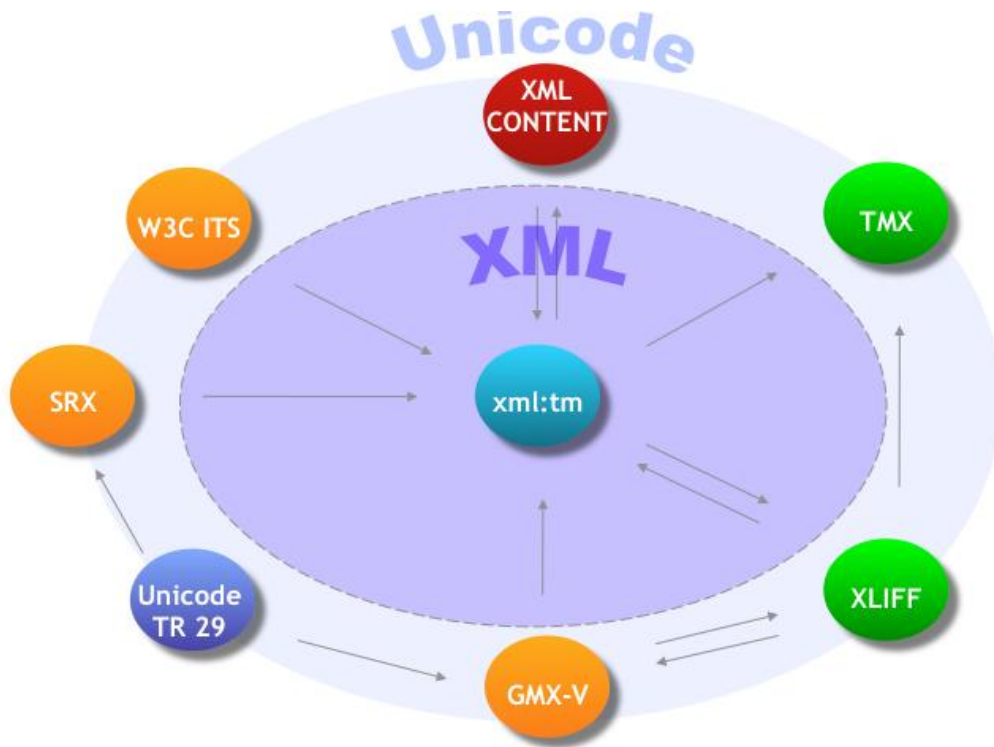


Figure 6: OAXAL Standards Interaction

The real power of XTM is that it is primarily XML based. XTM converts all non-XML file formats to XML prior to processing, and then back into the native format post translation without ANY loss.

Obviously, if the file format is already XML, such as In Design, XHTML, OOXML, ODF or DITA, then XTM processes the files natively. Having one internal format means that processing is greatly simplified. Internally, XTM also uses XLIFF as the translation format, so that interchange with XLIFF compliant systems is guaranteed.

XTM For Language Service Providers Explained

XTM can handle the following file formats:

- XML (Generic)
- XHTML
- HTML
- ODF (Open Office)
- DITA
- DocBook
- SVG
- RTF
- PDF
- PO/POT
- Adobe In Design (.INX & .IDML)
- Adobe Frame Maker
- OOXML (Microsoft Office 2007)
- rtf
- Microsoft Office 95 - 2003
- Visio (.VDX)
- SDF
- Java Property files
- Net resource files
- Plain text files
- YAML
- ASP, ASPX
- Trados TTX
- PHP
- JSON
- INI Files

In addition, XTM can import translation memories in TMX format from Trados and other CAT tools, as well as being able to accept the following formats from other CAT tools:

- Trados TTX files
- SDLXLIFF
- XLIFF files

Using Open Standards guarantees that you will not be locked into a proprietary trap and forced upgrade treadmill. It also makes for much easier interoperability with clients and partners who also use these standards.

5. Advanced Translation Technology

XTM introduces the concept of document-centric translation memory based on the LISA OSCAR xml:tm standard. This advanced technology allows for the maintenance of Author Memory within a CMS environment, providing every sentence in a document with a unique identifier. Furthermore, using this approach enables in-document matching when translating the document during its life cycle, along with in-document exact (ICE) matching, in-document leveraged and fuzzy matching. XTM also makes use of traditional database matching technology, but enhances this with linguistic stemming and advanced web search technology. XTM can also deliver high quality Statistical Machine Translation (SMT), based on the acclaimed Google SMT engine during the translation cycle, thereby further reducing the costs of localization.

The key concepts of XTM are automation, ubiquity in terms of browser interface, maximum re-utilization of linguistic assets through its server based architecture, and extensive use of available Internet resources.

Localization is a highly collaborative task that requires the real-time interaction between Project Managers, Translators, Correctors and Reviewers. XTM provides an ideal solution that allows all parties to cooperate interactively, and with maximum effect.

6. Web 2.0 based

A core design principle of XTM is that only a web interface is required. Thanks to the significant advances in browser technology over the past 8 years, the power and capabilities of the current generation of browsers allows for advanced applications like XTM to be run straight from the browser window. An example of this is the extensive use of Ajax technologies, seen in Gmail and Google Docs. In addition, Salesforce.com has also shown what can be achieved with current technology. There are many more examples. XTM supports both Firefox/Mozilla as well as Internet Explorer.

XTM For Language Service Providers Explained

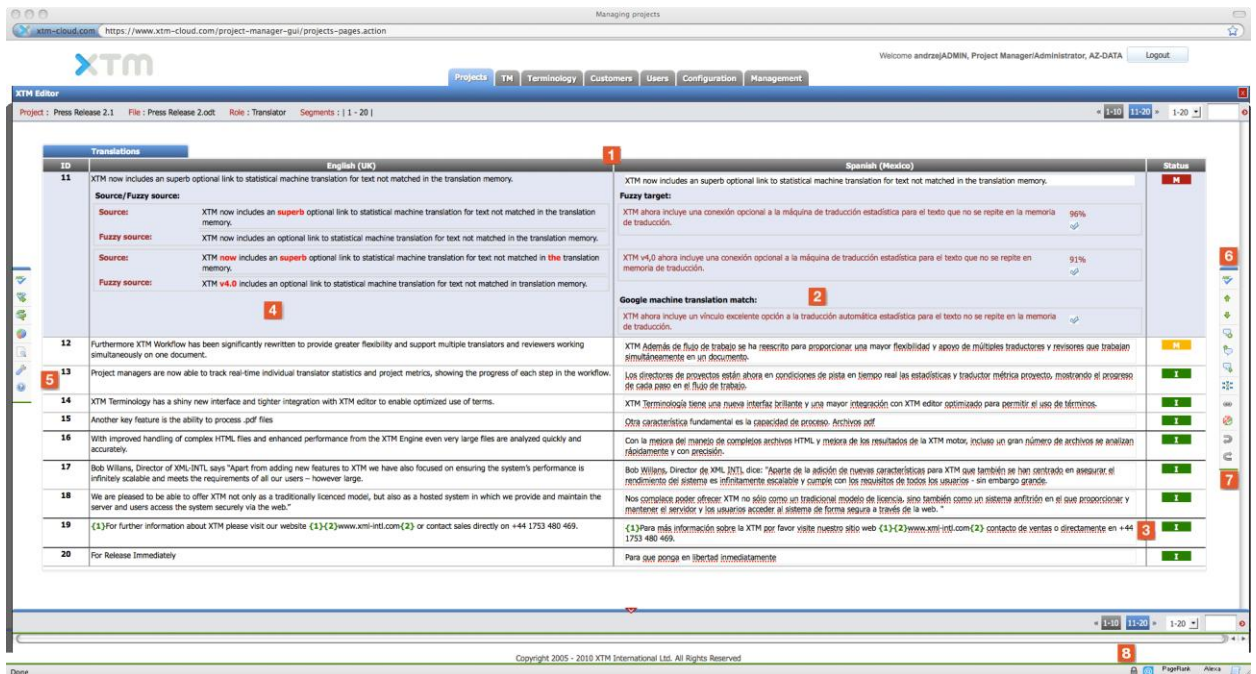


Figure 7: XTM Workbench

1. The pairs of source and target segments are presented to the translator.
2. Leveraged and fuzzy matches from the TM are provided for the translator.
3. In-context exact matches require no attention from the translator or reviewer.
4. Terms are highlighted in blue in the source text and in the approved translation provided.
5. PDF previews of the target document are generated on the fly.
6. The quality assurance module spell-checks and QAs the entire document.
7. XTM stores the full revision history of each segment.
8. Online access is fully encrypted.

The XTM translator workbench also provides the following facilities:

- Integrated concordance searching based on advanced linguistic techniques.
- Automated QA checking
- Integrated terminology identification and target term resolution
- Interactive Search and Replace
- Integrated monolingual and bilingual word look up, utilizing online web resources such as Wiktionary, Google Dictionary and Google Search
- Unlimited translation file size
- Multiple translators can work on the same file, at the same time
- Proofing and translation can be done concurrently on the same document

XTM For Language Service Providers Explained

The benefits of a browser-based user interface are multi-fold:

- All translators can share the same dynamic memories in real time across the Internet.
- No special software has to be installed and maintained on the translator's PCs.
- Multiple translators can work on the same file concurrently.
- The most up-to-date version of the translation is held centrally. There are no files to send out.

XTM comprises the complete set of tools required by the translator, including terminology, dictionary and translation memory management.

The speed and reliability of such an approach is significant compared with the alternative of desktop-based translation. There are no individual licenses to buy, and no desktop software to maintain. XTM does not preclude the fact that in certain circumstances, the translators may have to work on their local desktop with their own translation tools. To this end, XTM provides the ability to download the XLIFF translation file and edit it using the increasing number of low cost XLIFF translation editors.

The key feature of the XTM architecture and browser-based translation workbench is that most of the tasks, apart from the actual translation, are automated. This means that the translator merely has to concentrate on the task at hand: translation. The user interface demonstrates this. It is the essence of simplicity, in the same way as many Google web applications are. The user is not assailed with countless confusing icons and menu options. The interface is finely-tuned to the translator's needs and requirements.

7. Sub-contracting

XTM also contains a very powerful sub-contracting feature. This allows Translation companies, which have been sub-contracted to perform translation for one or more languages, to use XTM in a totally transparent manner, without having to set up the details of the translators involved in the main system. This enables sub-contractors to participate in the translation workflow without compromising any of their participants, whose details are kept confidential.

8. Content Management Systems Vendors

XTM has been successfully integrated with leading commercial and Open Source CMS systems. XTM is designed to provide the maximum benefit from the automation of the translation process for such systems. XTM also provides the important benefit of Author Memory, which is provided by means of the xml:tm standard used to maintain transparent unique identifiers for all pieces of text in a document.

9. Organizations with translation requirements

XTM can be used to great effect by commercial and government organizations with localization requirements. The benefits are that full control of the translation process is maintained, along with a high degree of automation allowing for substantial reductions in translation costs.

10. Conclusion

XTM is above all an affordable, flexible and technically advanced solution for LSPs. XTM can be easily integrated with existing LSP workflow and project management systems. Written in Java and supporting multiple databases, it is designed for integration. XTM is currently being integrated in multiple top-tier LSP systems. XTM has also been integrated with XTRF; the SRC, CRM and Project Management tool. As both systems are browser-based and written in Java, this is a formidable tool for medium-to-small LSPs. In addition, XTRF TM will be shortly made available as Software as a Service (SaaS) solution for medium and small sized LSPs.