

Didier Briel Consulting

Translation and consulting



OmegaT

Didier Briel
didier@didierbriel.fr
SERISS
March 2018

Contents

- Concepts
- Main features
- Exchange with other CAT tools
- Project management strategy
- Team projects
- Technical support



Concepts

Translation memory

Computer-Aided Translation

Free software

Concepts

Translation memory

- A translation memory is a “database” where translations are recorded
- The records can contain paragraphs, sentences or segments of sentences, stored as bilingual units
- The standard for translation memory exchange is TMX, an XML format

Concepts

Computer-Aided Translation

- Sometimes called “Translation Memory”
- It is an assistance for the translator, not a machine translation
- The computer allows to reuse sentences or segments of sentences already translated
- More generally, CAT is the whole assistance available to the translator through one or several programs (glossaries, dictionaries, etc.)

Concepts

Free software

- OmegaT is free software (GNU GPL v3)
- Free software is not freeware
- A program is free software if the program's users have the four essential freedoms
 - Run the program, for any purpose
 - Study how the program works, and change it so it does your computing as you wish
 - Redistribute copies so you can help your neighbour (including for money)
 - Distribute copies of your modified versions to others



Main features

OmegaT

Main features

- Fuzzy matches
- Automatic propagation of translations
- Glossaries
- Looking up terms ("concordance") in the project, in reference memories, in glossaries and in reference documents
- Projects with an unlimited number of folders and files, in all accepted formats
- Internet search system
- Team projects with an unlimited number of translators
- Right to left and bidirectional writing management
- Tag protection and validation (checking formatting)
- Morphological recognition
 - Stemming (fuzzy matches, glossaries)
 - Stop words (fuzzy matches)
 - Tokenization (compute "words", especially for Chinese and Japanese)
 - More than a hundred languages available (36 Lucene, 100+ Hunspell)

OmegaT

Main features

- Spell checking (Hunspell)
- Language checking (LanguageTool)
 - Basic rules applicable to most languages
 - Specific rules in 29 languages
- Dictionaries (mono and multilingual) in StarDict and Lingvo DSL format
- Auto-completion
 - Glossary entries
 - Abbreviations
 - History completion and prediction
- Aligner for all formats recognised by OmegaT
- Machine translation (Google Translate, Microsoft Translator, Apertium, Yandex, MyMemory)
- Scripts
 - Equivalent to Word and Excel macros
- Public plugin API
 - Additional file formats (Okapi)
 - Local machine translation (Apertium)
 - EPWING Japanese dictionary support (Hiroshi Miura)
 - More than 20 modules available

OmegaT

Supported formats

- More than 50 supported formats
- Microsoft Office (2007 and above)
 - Word, Excel, PowerPoint, Visio
- Open Document Format
 - LibreOffice, OpenOffice
- HTML and XHTML
- XLIFF
- TTX and SDLXLIFF (Trados)
- MQXLIFF (memoQ)
- TXML (Wordfast Pro)
- IDML (InDesign)
- PDF (plain text and through Icenix Infix export)



Exchange with other CAT tools

OmegaT

Exchange with other CAT tools

- Translation memories
 - TMX version 1.4b level 1 and 2
 - Read
 - Create
 - Textual information is preserved, as well as formatting placeholders
 - Unlimited number of translation memories
- Glossaries
 - Text files
 - Columns separated by tabulations or commas (CSV)
 - Source, Target and Comment columns
 - Read
 - Create
 - TBX
 - Read
 - Unlimited number of glossaries
 - Connection to the TaaS (Terminology as a Service) terminology website
 - 24 official European languages plus Russian and Turkish
 - Collaboration with memoQ and Trados Studio users



Project management strategy

Basic concepts

Recurrent vs. one-off projects

Project management strategies

Basic concepts

- OmegaT uses one translation memory per project (the project memory) and an unlimited number of “external” memories
- Each time the target documents are created, a corresponding translation memory is created
- Everything is processed in memory in OmegaT
 - The size of a translation memory cannot be “unlimited”
 - The translation memory is not a database
- A single project to do everything cannot really be considered

Project management strategies

Recurrent vs. one-off projects

- Recurrent projects, single memory
 - The project does not change
 - Only the content of the *source* folder changes when new jobs are received
 - The project translation memory contains all the segments translated in this project
 - Exported translation memories are not used
- One-off projects, centralised memories
 - A project is created for each new translation
 - This project uses a centralised repository of existing memories
 - Fuzzy matches give more information, since they contain the name of the memory
 - At the end of the project, the exported translation memory is copied in the central repository
 - The repository can be organised by subjects, clients, etc.



Team projects

Concepts

Setup

Usage

Team projects

Concepts

- OmegaT allows using “team projects”
- It is an alternative to “server” versions of other CAT tools
- Data are stored on a Subversion or Git server
- An unlimited number of translators can work simultaneously on the same documents
- It is possible to work both offline and online
- Apart from the core of the project, data can come from various sources (Subversion, Git and HTTP) through mapping

Team projects Setup

- The initial project must be created on a Subversion or Git server
 - Local network
 - Internet
 - Free hosting
 - Paid hosting
 - To create the project, the project manager uses Subversion or Git tools
 - TortoiseSVN under Windows
 - And OmegaT features
 - Manual or command-line driven creation of a base project
 - Commit Source Files

Team projects

Usage

- Project > Download Team Project
 - The user enters username and password if required
- All project data are downloaded
 - Project
 - Source documents
 - Glossaries
 - Translation memories
 - Dictionaries

Team projects

Usage

- The project can then be used like a usual project, with a few differences
 - Each time the project is open
 - All data are updated
 - On each save (3 minutes by default)
 - Modified data from the translation memory and the writeable glossary are sent to the server
 - Possible conflicts are solved (the most recent change wins)
 - Data modified by other translators are downloaded
- If necessary, target documents can be created online with Commit Target Documents



Technical support

Volunteer support

Professional support

Technical support

Volunteer support

- Developers on Sourceforge
 - Reporting incidents
 - Requesting enhancements
- Other users on the Yahoo group
 - Support
 - Tips and tricks
 - Discussions on current and future features
 - 2643 members
 - 24/7 (presence in most time zones)
 - Main language English, but support is available in numerous languages (French, German, Japanese, Russian, etc.)
 - Writing in mother tongue is welcome

Technical support

Professional support

- Affordable professional support
 - SDL Trados Studio 2017 Freelance costs 695 € (without discount)
 - For this price, you can buy numerous hours of support
- As the source code is available, you can choose your partner freely
- It is sometimes possible to use internal skills
 - Employee trained in OmegaT troubleshooting and development