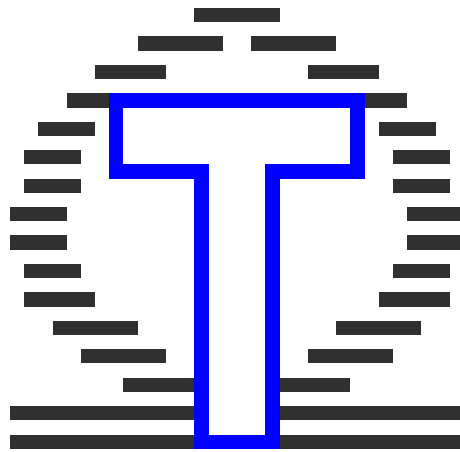


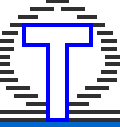
OmegaT



Dublin Computational Linguistic Research Seminars

Didier Briel

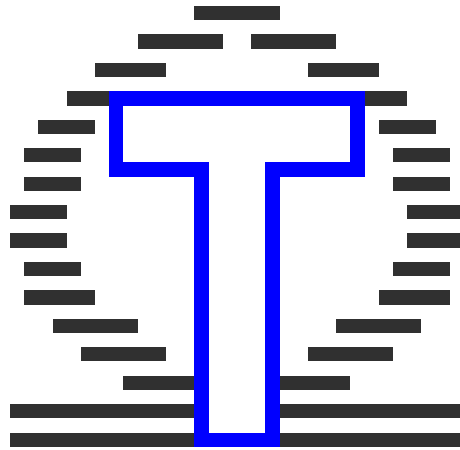
June 2012



Contents

- OmegaT workflow
- Main features
- Plugins
- Exchange with other CAT tools
- Supported formats
- The OmegaT project
- Availability
- Support

OmegaT

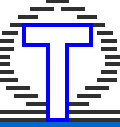


OmegaT workflow

Main characteristics

Translation of a file

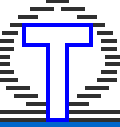
Demonstration



OmegaT workflow

Main characteristics

- Completely stand-alone
 - None of its features depends on the installation of other software (e.g., Microsoft Office)
- Available on all platforms compatible with Java 1.5 and later
- No intermediate format
 - No preparation
 - Import or conversion
 - No “clean-up”
 - Instantaneous dynamic modification of projects (adding/changing/removing documents)
- No database
 - All data are processed in memory
 - Very fast
 - Data size is limited
- Automatic propagation of translations

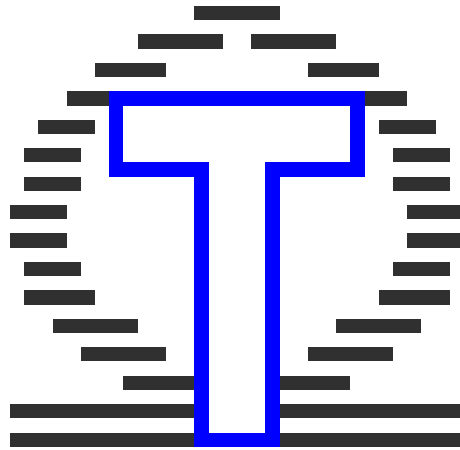


OmegaT workflow

Translation of a file

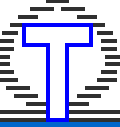
- Creating a project
- If needed, conversion of the source file
- Installing glossaries and translation memories
- Translation
- Generating the target documents
- If needed, conversion of the target file

OmegaT



Main features

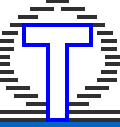
RTL and bidi issues



Concepts

Main features

- Fuzzy matching
- Automatic propagation of translations
- Glossaries
- Search terms in the project, in reference memories and in reference documents
- Projects can contain an unlimited number of folders and files, in all supported formats
- Right to left and mixed (bidi) writings
- Tag check (formatting check)
- Spell checking (Hunspell or JMySpell)
- Mono and multilingual dictionaries in StarDict and Lingvo DSL format
- Machine Translation (Google Translate, Apertium, Belazar)

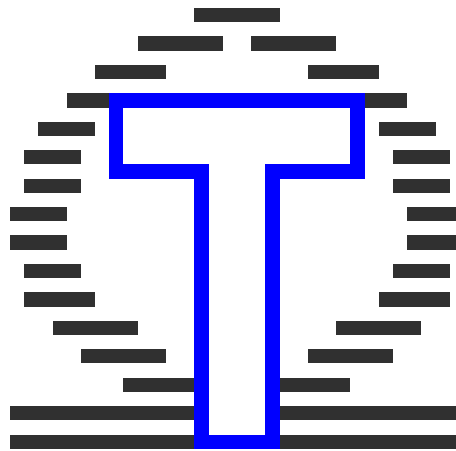


Concepts

Main features: RTL and bidi issues

- A CAT tool must be able to display simultaneously LTR and RTL languages
 - English to Arabic
 - Hebrew to English
- OmegaT switches the segment orientation automatically
 - Unicode markers at the beginning of each segment are used to set directionality
 - The orientation can be forced manually (Shift+Ctrl+O)
- RTL languages need to switch to LTR within the segment
 - Foreign names
 - Phone numbers
- José García-Verdugo (Universidad de Castilla-La Mancha, Spain, Arabic-Spanish teacher)
 - *“I chose OmegaT because of how well it handles right-to-left and left-to-right issues which, being far from perfect, is actually much better than a number of expensive commercial apps.”*
 - <http://tech.groups.yahoo.com/group/OmegaT/message/19344>

OmegaT



Plugins

Licence compatibility

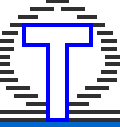
Architecture

Stubs

Available plugins

Tokenizers

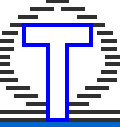
Scripting



Plugins

Licence compatibility

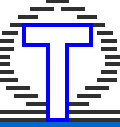
- Plugins are distributed separately, mainly for legal reasons
 - OmegaT is GPL v2 strict, because of JAXB
 - Most plugins are using Apache components
 - The Apache licence is not compatible with GPL v2
 - The Apache license is compatible with GPL v3
 - A small part of OmegaT (*stubs*) is re-licenced under GPL v3
 - This is possible, because OmegaT sources say *either version 2 of the License, or (at your option) any later version*



Plugins

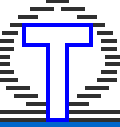
Licence compatibility

- Any licence compatible with GPL v3 can be used in a plugin
 - Apache
 - Modified BSD
 - *3-clause BSD*
 - ...
 - <http://www.gnu.org/licenses/license-list.en.html#GPLCompatibleLicenses>
- Contrary to GPL, some of these licences (e.g., BSD) do not require distribution of the source code when packed with other software
 - *Permissive non-copyleft*



Plugins Architecture

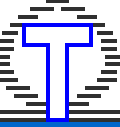
- External plugins are loaded at runtime from the *plugins* folder
- A plugin interacts with OmegaT mainly through *interfaces*, which is why *stubs* are needed, to reference these interfaces
- *Stubs* are “empty” source code, containing just the declaration of the needed objects and methods
- At runtime, these empty methods will be replaced by the real ones
- How to create a plugin is documented in OmegaT sources, in the *docs_devel* folder
 - It includes a complete sample plugin project



Plugins

Available plugins

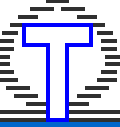
- Provided by OmegaT
 - LanguageTool
 - Language checker
 - Tokenizers
 - Improves fuzzy matches and glossaries
 - Scripting
 - Allows to run scripts within OmegaT
- Provided by third parties
 - Okapi
 - Additional file filters
 - http://www.opentag.com/okapi/wiki/index.php?title=Okapi_Filters_Plugin_for_OmegaT
 - Glossum
 - Online glossaries
 - <http://www.glossum.com/omegat>



Plugins

Tokenizers

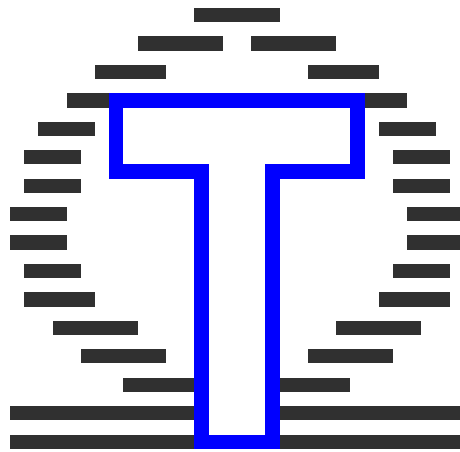
- Tokenizers are developed by Lucene (Apache)
- Available in 24 languages
 - Based on Lucene 2.9.0
- OmegaT uses them for three things
 - Computing tokens
 - Enhancement of spell-checking
 - Enhancement of “word” recognition for CJK languages
 - Removing *stop words* for fuzzy matches
 - Computing lemmas
 - Used both for fuzzy matches and glossaries
 - Not available in some languages
 - Czech
 - Stop words only
 - Chinese, Japanese and Korean (*CJK*)
 - Tokenization only



Plugins Scripting

- Allows to launch processes acting on data held in memory in OmegaT
- Allows to create graphical interfaces and interactive functions
- Numerous languages supported by default
 - NetRexx
 - XSLT
 - Groovy
 - JavaScript
 - Easy addition of other languages
 - Jacl (Tcl/Tk)
 - Jython (Python)
- Samples scripts are included
 - Search and replace
 - GUI scripting
 - ...

OmegaT



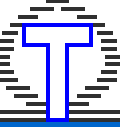
Supported formats

Native formats

Okapi plugin

Custom coding

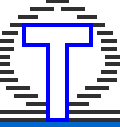
Okapi Rainbow



Supported formats

Native formats

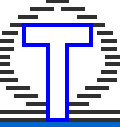
- Documentation formats
 - Plain text (any encoding supported by Java), including Unicode
 - (X)HTML (including complete website tree structure)
 - StarOffice, OpenOffice.org, LibreOffice and OpenDocument
 - Open XML (Microsoft 2007/2010)
 - Help & Manual
 - HTML Help Compiler
 - LaTeX
 - DokuWiki
 - CopyFlow Gold for QuarkXPress
 - DocBook
 - Typo3 LocManager
 - Icenix Infix (PDF)
 - XLIFF source = target
 - TXML Wordfast source = target



Supported formats

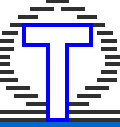
Native formats

- Localisation formats
 - Android resources
 - Java *.properties*
 - *Key-value* files
 - Mozilla DTD
 - Windows resources (RC)
 - WiX localisation
 - ResX
 - Flash XML export
 - Camtasia for Windows
 - Magento CE localisation
 - PO (Portable Object File) (reading existing translations)
 - SubRip subtitles (SRT)
 - SVG images



Supported formats Okapi plugin

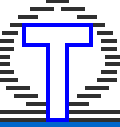
- InDesign IDML
- JSON
- Qt TS
- Trados TagEditor TTX (reading existing translations)
- Transifex projects
- Wordfast Pro TXML (beta) (reading existing translations)



Supported formats

Custom coding

- Trivial modifications of the source code allow creating a file filter
- The process is documented
 - http://www.omegat.org/en/howtos/new_filter.html
- For XML filters, it is very easy, which probably explains why there isn't a generic XML filter

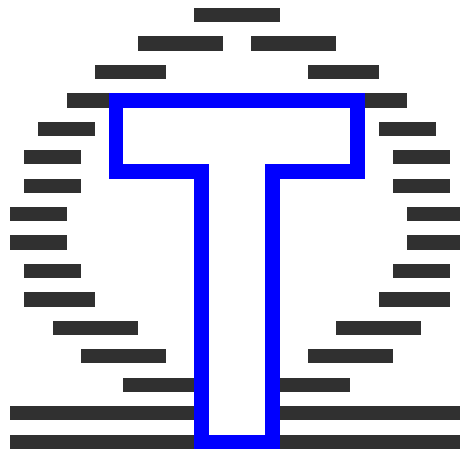


Supported formats

Okapi Rainbow

- Okapi Rainbow can create OmegaT projects, with added value
 - Additional file filters
 - Additional formats
 - MIF (Adobe FrameMaker)
 - Reading existing translations for some formats
 - Generic XML filter
 - Translatable text tagging
 - ITS (Internationalization Tag Set)
 - Pre-translation with machine translation

OmegaT

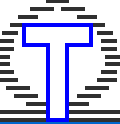


Exchange with other CAT tools

Translation memories

Bilingual documents

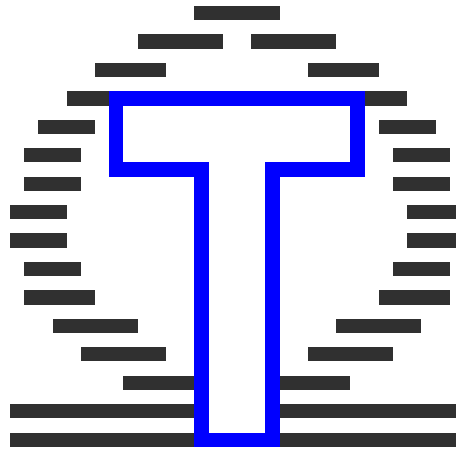
Glossaries



Exchange with other CAT tools

- Translation memories
 - Unlimited number of translation memories
 - TMX version 1.4b level 1 and 2
 - Import
 - Export
 - Textual information is preserved, as well as formatting placeholders
- Bilingual documents (often called unclean documents)
 - PO
 - XLIFF (with Rainbow to read existing translations)
 - TTX (with the Okapi plugin)
 - TXML (with the Okapi plugin to read existing translations)
- Glossaries
 - Unlimited number of glossaries
 - Glossaries can be read in several formats
 - TSV (tab separated values)
 - CSV (comma separated values)
 - TBX (TermBase eXchange)
 - Only one of them is the writeable glossary
 - TSV format only

OmegaT



The OmegaT project

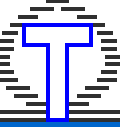
History

Key people

Project philosophy

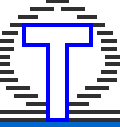
User uptake

Funding



The OmegaT project History

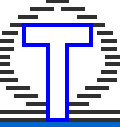
- First development
 - In C++ in 2000
- First public release
 - In Java in 2001
- Sentence segmenting
 - OmegaT 1.6 in 2005
- Spell-checking
 - OmegaT 1.8 in 2008
- “Instant” project loading with on-demand matching
 - OmegaT 2.0 in 2009
- First plugin
 - OmegaT 2.0 in 2009
- Multiple translations support
 - OmegaT 2.5 in 2011
- Team projects
 - OmegaT 2.6 in 2012



The OmegaT project

Key people

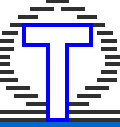
- Initial developer
 - Keith Godfrey (US/UK)
- Project co-ordinator and website manager
 - Marc Prior (Germany)
- Past development managers
 - Maxym Mykhalchuk (Italy/Ukraine)
 - Henry Pijffers (Netherlands)
- Current Development team
 - Development manager
 - Didier Briel (France)
 - Lead developer
 - Alex Buloichik (Belarus)
- Localisation manager and user group owner
 - Jean-Christophe Helary (Japan)
- Documentation manager
 - Vito Smolej (Germany)



The OmegaT project

Project philosophy

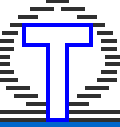
- OmegaT is an informal organisation
 - There is no legal structure
- Users are encouraged to contribute
 - This does not necessarily requires programming skills
 - Translation
 - Video tutorials
 - Presentation to local translator groups
 - ...



The OmegaT project

User uptake

- No accurate figures
 - OmegaT does not track usage
- From a 2010 survey
 - *OmegaT is used 1/3 as much as Wordfast, DejaVu and MemoQ, and 1/8 as much as the market leader Trados (source: Wikipedia)*
 - <http://www.translationtribulations.com/2010/07/results-of-june-translation-tools.html>



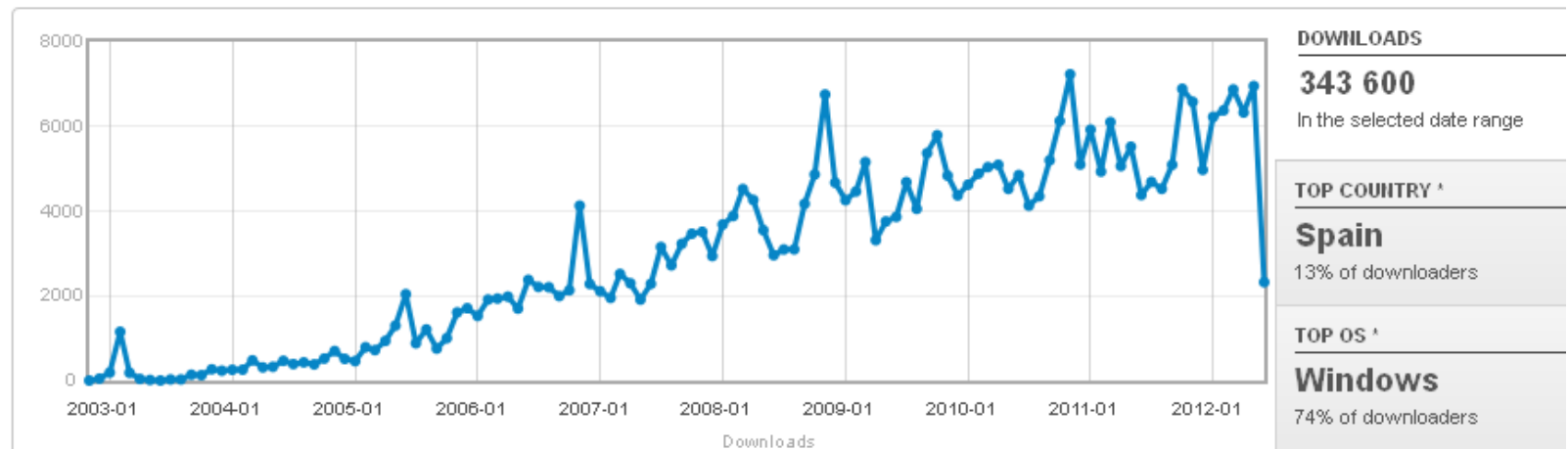
The OmegaT project

User uptake: downloads

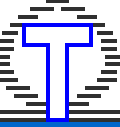
- Downloads since the beginning on Sourceforge

Home (Change File)

Date Range: 2002-11-28 to 2012-06-13



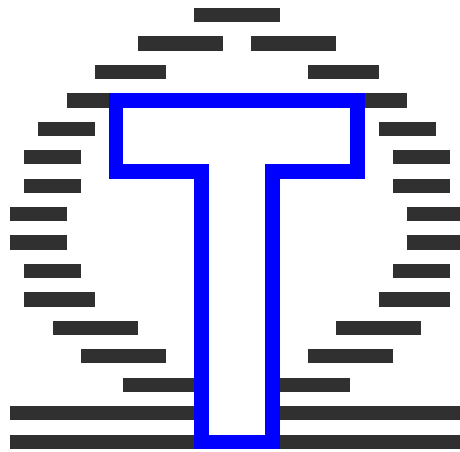
- Java Webstart usage is not counted



The OmegaT project Funding

- The great majority of work on the OmegaT project has been completed on a voluntary basis
- Donations are welcome
 - The OmegaT project will use them for specific purposes
 - Funding a Mac for the lead developer
 - Funding developments
- If a development is really wanted, it is possible to sponsor it
 - The development team negotiates the features and the price with the sponsor

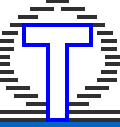
OmegaT



Availability

Download/Java Webstart

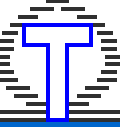
Source code



Availability

Download/Java Webstart

- OmegaT is available in two versions
 - Standard
 - Fully documented (at least in English)
 - Latest
 - New features are not documented
 - New features may be less stable
- For four platforms
 - Windows
 - Linux
 - Generic
 - Mac
- With or without the Java Runtime Environment
- In addition, OmegaT can be used with Java Webstart
 - No installation (from a user point of view)
 - Updates are done automatically
 - Plugins cannot be used
- All versions can be downloaded from omegat.org
- Plugin are on a separate website
 - <https://sourceforge.net/projects/omegat-plugins/>

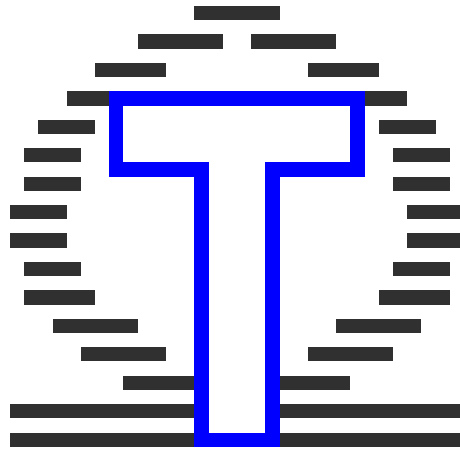


Availability

Source code

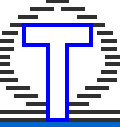
- Source code is available with all releases
- */trunk* sources are available
 - Subversion
 - svn co
<https://omemat.svn.sourceforge.net/svnroot/omemat/trunk/omemat>
 - Tarball
 - <http://omemat.svn.sourceforge.net/viewvc/omemat/?view=tar>
- Prototypes are sometimes available from Git
 - <http://omemat.git.sourceforge.net/git/gitweb.cgi>
- Plugin */trunk* sources are available from Git
 - <http://omemat-plugins.git.sourceforge.net/git/gitweb-index.cgi>

OmegaT



Support

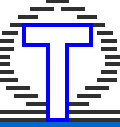
User and developer support
Bug and enhancement trackers
Contribution



Support

User and developer support

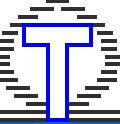
- User mailing list
 - <http://groups.yahoo.com/group/omegat>
 - 1750 members
 - 24/7 (presence in most time zones)
 - Main language English, but support is available in numerous languages (Japanese, Russian, German, French, etc.)
 - Writing in mother language is encouraged
- Developer and power user mailing list
 - https://sourceforge.net/mailarchive/forum.php?forum_name=omegat-development
 - 125 members
 - English language
 - Technical discussions



Support

Bug and enhancement trackers

- Bug reports
 - https://sourceforge.net/tracker/?group_id=68187&atid=520347
 - 500 bugs reported since 11/2002
 - 47 open as of 13/06/2012
 - Oldest open bug is from 07/2006
- Enhancement requests (*RFE*)
 - https://sourceforge.net/tracker/?group_id=68187&atid=520350
 - 795 requests since 11/2002
 - 340 open as of 13/06/2012
 - Oldest open request is from 06/2004



Support Contribution

- Before sending a contribution, it is often useful to discuss it
 - Submission of an enhancement request
 - Discussion in the user mailing list
 - Discussion in the development mailing list
 - Private discussion with the development manager
- Commit access to Subversion is limited to a few core developers
- Contributions are sent to the development manager
- What to submit
 - Zip archive
 - Patch
 - All changed source files
 - Test data if needed
 - Explanation of what the patch does