

translate5

translate5 - Towards a General-Purpose MT Evaluation Tool

Marc Mittag, MittagQI – Quality Informatics, Mössingen

translate5: Open source coming from translation industry

- At first: VM-distribution
- Origin and development history; about my background
- Current use cases in translation industry
 - Proofreading with
 - Terminology recognition
 - With MQM-based error annotation and Quality categorization
 - Setting of Status
 - Simple workflow with mailing of changed segments
 - Terminology correction in large Translation Memories
 - Detailed analysis of correction needs

Created traction in MT-research

- Column-based approach on data
- SQL back-end
- Ability to serve as a hub
 - Ideal to dock all the various tools of MT-research to one hub
 - to make them easily usable for everyone
 - ensure, they won't get lost, if their developer moves on
 - optionally Integrate them in translate5 installer and updater
 - For working on column-based data
 - Helps to bring the MT-research tools to application in the industry
- Continuously developed Open Source Quality assurance tool backed by the industry
- Synergy through usage of common Open Source tool
 - Big potential in bringing MT-research-ideas to translation industry
 - And translation industry perspectives to MT-research

translate5 installation and update

- Possible on Linux and Windows with a few clicks

<http://confluence.translate5.net/display/TIU/Installation+and+update>

- On Mac similar to Linux for experienced users
- On local student systems: Installation via preconfigured Ubuntu Virtualbox instance

<http://www.translate5.net/downloads/translate5-ubuntu-virtualbox-preinstalled.zip>
(coming soon)

Update via

```
sudo -u www-data ./install-and-update.sh
```

Current usage in MT-research

- Live example at <http://translate5-metashare.dfki.de/>
- Live example: Annotating multiple target alternatives in translate5
 - Download test data via <http://confluence.translate5.net/display/BUS>
 - CSV example with custom MQM issue types
 - CSV example with multiple target alternatives

Example of serving as a hub for MT-research

Integration of MT-ComparEval

- Write small translate5-plugin
- Optionally integrate MT-ComparEval in installation and update routine
- Show live, where integrate would make sense

Current main features from user perspective

Showing in live example with people trying themselves

- big amounts of data (files and segments)
- Flexible filtering and sorting of columns
- Flexible columns (show specific columns to specific users)
- Multiple import formats
- Commenting on segments
- MQM-tagging
- TermTagging
- User management
- Workflow
- Task overview
- Locking of specific segments

Future application in MT-research

- Serving as a constant database-driven hub
 - to evaluate MT-data with different existing and future research tools, like
 - MT-CompareEval (show it online)
 - BLEU-calculation in a translate5-column
 - Appraise
 - QuEst
 - Word-Aligners
 - Part of Speech Tagger
 - to research and compare existing annotation projects by
 - Copying the data from existing in new task
 - Re-translate data in a new column of an existing task
 - Compare the new column with existing ones, etc.
 - to do semi-automatic MQM-markup

Future application in MT-research and translation industry

- Show mock-ups about
 - TM/MT-integration (currently at the end of this presentation)
 - Word-Aligner (to be added)
 - Semi-automated MQM-annotation with diff (to be added)
- Further development ideas
 - Integrate Keystroke-logging
 - Export / Visualize postediting time

Important architecture concepts

- Event-triggered plug-in system
(show plugin-examples in source-code and translate5 confluence)
- REST-API
(show REST-API in source-code and translate5 confluence)
- Column-based segment-database (flexible base numbers of columns by task)
(show database in phpMyAdmin and Entity-concept in source code)
- Queuing, Management and Load-Balancing of external and internal processes including dependency management, etc.
(show termTagger and segment counting in action)
- Pragmatic object-orientated structure
- Based on
 - Zend Framework, PHP (backend)
 - ExtJs, Javascript (frontend; show ExtJs documentation)
 - MySQL database, other SQL-DB systems possible

Discussion

- Which tools would be interesting to integrate?
- How would the development for them work?
- Which of these proposals make sense for the development project at MT-Marathon?

Visualize word alignment in translate5

source target(mqm)

Norway's rakfisk: Is this the world's smelliest fish? Norwegen u.a. rakfisch: Das ist der böse riechendste Fisch der Welt?



Semi-automatic MQM annotations (Mock-up 1; before MQM-selection)

- On segment save the changed string gets selected automatically
 - The MQM-issue popup opens up
 - The user can select an issue (via mouse, keyboard or keyboard shortcut)
- If more than one section changed, after the MQM-assignment for the first section, the next gets selected.
- After saving the segment, a diff algorithm automatically selects the changed section in the corresponding non-editable version of the column and tags that, too.

Nr.	Source text	Target text	Target text	Comments
109	Marketing Mercedes-Benz Trucks	Marketing Mercedes-Benz Trucks	Marketing Mercedes-Benz Trucks	
110	Single Source Publishing	Single Source Publishing	Single Source Publishing	
111	(SSP)	(SSP)	(SSP)	
112	Überblick Funktionsweise Single SSP	Single SSP functionality at a glance	Single SSP functionality at a glance	
113	Aufbau einer Struktur auf inhaltlicher Ebene mit dem Ziel Inhalte möglichst flexibel medienneutral wiederzuverwenden	Development of a structure on a content level with the possible	Development of a structure on a content level with the goal of reusing contents as media	
114	<1>1</1>	<1>1</1>	<1>1</1>	
115	Bildung von in sich geschlossenen Inhaltselementen (Topics), die für sich alleine stehen können und eine Sinn ergeben	Formation of self-contained content elements (Topics) that can stand alone and make sense	Formation of self-contained content elements (Topics) that can stand alone and make sense	
116	<1>2</1>	<1>2</1>	<1>2</1>	
117	<1>Verwaltung und Strukturierung der Topics pro Ausgabemedium </1>zu einem <2>sinnvollen Informationsgebilde (</2>Topic-Map)	<1>Management and structuring of the topics for each output medium </1>to create a <2>sensible body of information (</2>Topic Map)	<1>Management and structuring of the topics for each output medium </1>to create a <2>sensible body of information (</2>Topic Map)	
118	<1>3</1>	<1>3</1>	<1>3</1>	



Semi-automatic MQM annotations (Mock-up 3; after segment save)

- On segment save the changed string gets selected automatically
 - The MQM-issue popup opens up
 - The user can select an issue (via mouse, keyboard or keyboard shortcut)
- If more than one section changed, after the MQM-assignment for the first section, the next gets selected.
- After saving the segment, a diff algorithm automatically selects the changed section in the corresponding non-editable version of the column and tags that, too.

Nr.	Source text	Target text	Target text	Comments
109	Marketing Mercedes-Benz Trucks	Marketing Mercedes-Benz Trucks	Marketing Mercedes-Benz Trucks	
110	Single Source Publishing	Single Source Publishing	Single Source Publishing	
111	(SSP)	(SSP)	(SSP)	
112	Überblick Funktionsweise Single SSP	Single SSP functionality at a glance	Single SSP functionality at a glance	
113	Aufbau einer Struktur auf inhaltlicher Ebene mit dem Ziel Inhalte möglichst flexibel medienneutral wiederzuverwenden	Development of a structure on a content level with the [3aim 3] of reusing contents as flexibly as possible regardless of the media in question	Development of a structure on a content level with the [3goal 3] of reusing contents as flexibly as possible regardless of the media in question	
114	<1>[3]1[3]1.	<1>[3]1[3]1.	<1>[3]1[3]1.	
115	Bildung von in sich geschlossenen Inhaltselementen (Topics), die für sich alleine stehen können und eine Sinn ergeben	Formation of self-contained content elements (Topics) that can stand alone and make sense	Formation of self-contained content elements (Topics) that can stand alone and make sense	
116	<1>[3]2[3]1.	<1>[3]2[3]1.	<1>[3]2[3]1.	
117	<1>Verwaltung und Strukturierung der Topics pro Ausgabemedium </1> zu einem <2>sinnvollen Informationsgebilde (</2>Topic-Map)	<1>Management and structuring of the topics for each output medium </1> to create a <2>sensible body of information (</2>Topic Map)	<1>Management and structuring of the topics for each output medium </1> to create a <2>sensible body of information (</2>Topic Map)	
118	<1>[3]3[3]1.	<1>[3]3[3]1.	<1>[3]3[3]1.	

translate5 MT/TM-integration



translate5

Eingeloggter Benutzer: Thomas Lauria
Loginname: manager

[Abmelden](#) [Meine Einstellungen](#) Deutsch ▾
[Task Overview](#) [User Administration](#) [TM Overview](#)

TM Overview

Add TM Aktualisieren

TM Name			Color		Resource
MyMemory	Deutsch (de)	English (en)			MyMemory
TM of Client XY	Deutsch (de)	English (en)			OpenTM2
Collected Phrases	English (en)	Deutsch (de)			My Slow TM
TM of Client XY	Deutsch (de)	Italienisch (it)			OpenTM2
Moses MT	English (en)	Deutsch (de)			Moses MT

translate5 MT/TM-integration

translate5
open source initiative

Login name: manager

English

Task Overview

Add Task Refresh

Actions

Properties to task "ENG_Minipaket"

Associate users to task Task-properties Associate TMs to task

	Name	Deutsch (de)	Englisch (en)
OpenTM2 TMs			
<input type="checkbox"/>	TM of Client XY	Deutsch (de)	Englisch (en)
<input checked="" type="checkbox"/>	TM of Client YX	Deutsch (de)	Englisch (en)
My Slow TM			
<input type="checkbox"/>	Collected Phrases	Deutsch (de)	Englisch (en)
MyMemory			
<input checked="" type="checkbox"/>	MyMemory	Deutsch (de)	Englisch (en)
Moses MT			
<input checked="" type="checkbox"/>	Moses MT	Deutsch (de)	Englisch (en)

Save TMs

close window

translate5 MT/TM-integration



translate5

Files << Segment list and editor

Work files - Editor modes ▾ Hide tags Short tag view Full tag view Reset sorting / filtering QM Subsegment Statistics

Nr.	Source text	Pivot language	Target text	Target text	Comments
1	This file is a based on a part of the php-online-Dokumentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	
2	Apache 2.x on Unix systems - Manual	xxxxxxxxxxxxxxxxxxxxxApache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	
3	PHP Manual	xxxxxxxxxxxxxxxxxxxxxPHP Manual	PHP Manual	PHP Manual	
4	Installation and Configuration	xxxxxxxxxxxxxxxxxxxxxInstallation und Konfiguration	Installation und Konfiguration	Installation und Konfiguration	
5	Installation on Unix systems	xxxxxxxxxxxxxxxxxxxxxInstallation auf Unix-Systemen	Installation auf Unix-Systemen	Installation <u>auf Unix-Systemen</u>	
6	Apache 1.3.x on Unix systems	xxxxxxxxxxxxxxxxxxxxxApache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	
7	Apache 2.x on Unix systems				

MT / TM Matches Concordance Search

	Source Lang	Target lang	MatchRate / TM
1	Installation on Unix systems	Installation auf Unix-Systemen	100%
2	Installation on Unix systems	Installation auf Unix-Systemen	100%
3	Installation on Linux systems	Installation auf Linux-Systemen	95%
4	Installation on a Unix system	Installation auf einem Unix-System	85%
5	Installation on PC systems	Installation auf PC-Systemen	85%
6	Attributes: Last Edit: Marc Mittag 07/05/2015 Created: Thomas Lauria 07/05/2014	Another Result....	80%
7		Another Result....	80%
8		Another Result....	80%
9	(This is shown as ToolTip on MouseOver as long the mouse is over the icon)	Another Result....	80%
10		Another Result....	80%

Reference files +

translate5 MT/TM-integration



translate5

Dateien << **Segmentliste und Editor**

Arbeits-Dateien - DummyDirectory - install-unix.apach...

Editor Modi | Tags verbergen | **Tag-Kurzansicht** | Tag-Vollansicht | Sortierung / Filterung zurücksetzen | QM Subsegment Statistik

Nr.	Ausgangstext	Relaisprache	Zieltext	Zieltext	Kommentare
1	This file is based on a part of the php-online-Dokumentation. Its translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. Its only purpose is the generation of demo-data for translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	
2	Apache 2.x on Unix systems - Manual	xxxxxxxxxxxxxxxxxxxxxApache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	Apache 2.0 auf Unixsystemen - Manual	
3	PHP Manual	xxxxxxxxxxxxxxxxxxxxxPHP Manual	PHP Manual	PHP Manual	
4	Installation and Configuration	xxxxxxxxxxxxxxxxxxxxxInstallation und Konfiguration	Installation und Konfiguration	Installation und Konfiguration	
5	Installation on Unix systems	xxxxxxxxxxxxxxxxxxxxxInstallation auf Unix-Systemen	Installation auf Unix-Systemen	Installation auf Unix-Systemen	
6	Apache 1.3.x on Unix systems	xxxxxxxxxxxxxxxxxxxxxApache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	Apache 1.3.x auf Unix-Systemen	
7	Apache 2.x on Unix systems				
8	Lighttpd 1.4 on Unix systems	xxxxxxxxxxxxxxxxxxxxxLighttpd 1.4 on Unix systems	Lighttpd 1.4 on Unix systems	Lighttpd 1.4 on Unix systems	

MT / TM Matches | Concordance Search | **TestTM One Results**

Searchfield for Source: Searchfield for Target: 2 of 3 TM Resources answered

	Source Lang	Target Lang	TM - Resource
1	Installation on Unix systems	Installation auf Unix-Systemen	TM of Client XY - OpenTM2
2	Installation on Unix systems	Installation auf Unix-Systemen	TestTM One - MyTM
3	Installation on Linux systems	Installation auf Linux-Systemen	TM of Client XY - OpenTM2
4	Installation on a Unix system	Installation auf einem Unix-System	TM of Client XY - OpenTM2
5	Installation on PC systems	Installation auf PC-Systemen	TestTM One - MyTM
	loading...		Collected Phrases - My Slow TM

Referenz-Dateien

translate5 MT/TM-integration

Authenticat
Login name:
Task: testfiles
TaskGuid: {35

Logout



Segment list and editor

Editor modes | Hide tags | Short tag view | Full tag view | Reset sorting / filtering | QM Subsegment Statistics

Nr.	Source text	Target text	Autostatus
1	This file is a based on a part of the php-online-Dokumentation. It's translation is done by a pretranslation based on a very fast winalign-Project and is not at all state of the translation art. It's only purpose is the generation of demo-data for translate5.	Diese Datei ist Teil der php-online-Dokumentation. Ihre Übersetzung ist durch eine Vorübersetzung entstanden, die auf einem sehr schnell durchgeführten winalign-Project basiert und in keiner Art und Weise dem State of the Art eines Übersetzungsprojekts entspricht. Sein einziger Zweck ist die Erzeugung von Demo-Daten für translate5.	
2	Apache 2.x on Unix systems - Manual	Apache 2.0 auf Unixsystemen - Manual	
3	PHP Manual	PHP Manual	

MT / TM Matches | Concordance Search

	Source text	Target text	Score
1	Installation on Unix systems	Installation auf Unix-Systemen	100%
2	Installation on Unix systems	Installation auf Unix-Systemen	100%
3	Installation on Linux systems	Installation auf Linux-Systemen	95%
4	Installation on a Unix system	Installation auf einem Unix-System	85%
5	Installation on PC systems	Installation auf PC-Systemen	85%