



**Machine Translation Enhanced
Computer Assisted Translation**

D4.1 - First version of MateCat Tool

Author: Alessandro Cattelan

Dissemination level: Public

Date: September 30th, 2012

Grant agreement no.	287688
Project acronym	MateCat
Project full title	Machine Translation Enhanced Computer Assisted Translation
Funding scheme	Collaborative project
Coordinator	Marcello Federico (FBK)
Start date, duration	November 1 st 2011, 36 months
Dissemination level	Public
Contractual date of delivery	September 30 th , 2012
Actual date of delivery	September 30 th , 2012
Deliverable number	D4.1
Deliverable title	First version of MateCat Tool
Type	Prototype
Status and version	Final
Number of pages	6
Contributing partners	Translated
WP leader	Translated
Task leader	Translated
Authors	Alessandro Cattelan
Reviewers	Marco Trombetti
EC project officer	Kimmo Rossi
MateCat partners are:	Fondazione Bruno Kessler (FBK), Italy
	Université Le Mans (LE MANS), France
	The University of Edinburgh (UEDIN), UK
	Translated S.r.l. (TRANSLATED), Italy

For copies of reports, updates on project activities and other MateCat-related information, contact:

FBK

MateCat

Marcello Federico

Povo - Via Sommarive 18

I-38123 Trento, Italy

federico@fbk.eu

Phone: +39 0461 314 552

Fax: +39 0461 314 591

Copies of reports and other material can also be accessed via <http://www.matecat.com>

© 2012, Alessandro Cattelan

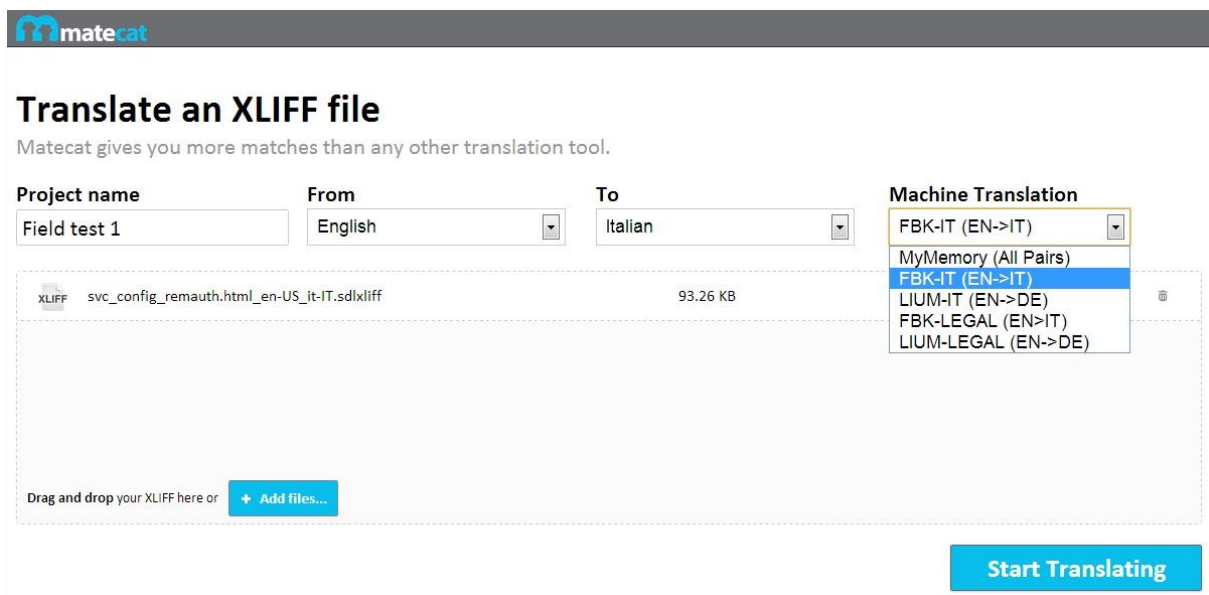
No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

Executive summary

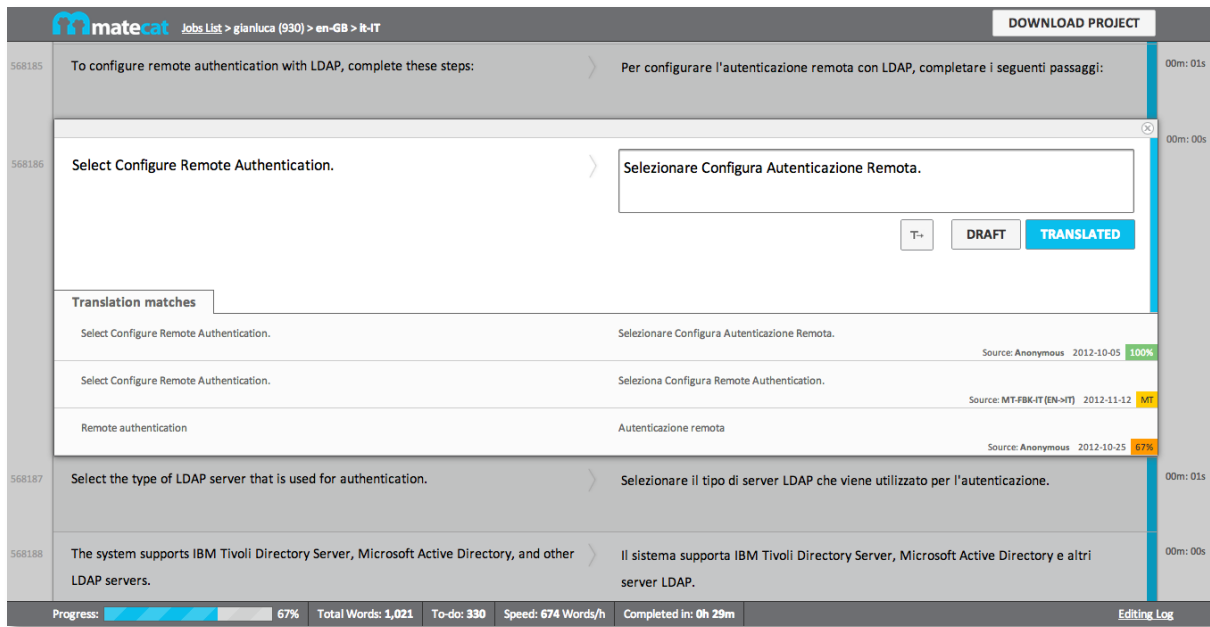
The first beta release of the MateCat tool has been finalized and is now available online. The current release is tailored for machine translation developers and researchers. It provides a basic platform to analyze translators' behaviour so as to adapt the software and to measure the improvements of the MT engine. The current web-based translation environment allows to measure the time to edit and post-editing effort of a translation performed using a specific MT server connected via a set of standard API.

CAT Tool Description

The CAT tool consists of three main components. The first one is the file importer which allows to import the files to translate (XLIFF format), set the language pair and specify the machine translation engine to test.




The second component is the translation editor where the user performs the translation. The text is presented in a tabular view and the document is broken down into minimal units (segments). For the active segment (i.e. the segment the translator is editing), three translation matches are presented with their relative quality match and source (TM, MT). The editor also displays the time to edit for each segment and some information on the job such as progress, total words, words remaining to translated, expected time before completion.



The screenshot shows the MateCat web interface for a translation job. At the top, there's a navigation bar with the MateCat logo, user information 'Jobs List > gianluca (930) > en-GB > It-IT', and a 'DOWNLOAD PROJECT' button. The main content area is split into source and target text. The source text is 'To configure remote authentication with LDAP, complete these steps:' and the target text is 'Per configurare l'autenticazione remota con LDAP, completare i seguenti passaggi:'. Below this, there's a section for 'Select Configure Remote Authentication.' with a corresponding Italian translation 'Selezionare Configura Autenticazione Remota.' and buttons for 'T->', 'DRAFT', and 'TRANSLATED'. A 'Translation matches' table is visible below, listing three matches with their respective source information and completion percentages (100%, MT, and 67%). The bottom status bar provides job statistics: Progress: 67%, Total Words: 1,021, To-do: 330, Speed: 674 Words/h, Completed in: 0h 29m, and an 'Editing Log' link.

The third component is the editing log which displays statistics on the time to edit and post-editing effort for each translated segment and for the entire translation job. From the same component, users can download a CSV report for each job to perform a more accurate statistical analysis.



[< Back to Translation](#)
[Export All Data in CSV](#)

Job 984 - Editing Log

Slowest 5.000 segments by time-to-edit

Summary

Words	Avg Secs per Word	% of MT	% of TM	Total Time-to-edit	Avg PE Effort %	% of words in too SLOW edits	% of words in too FAST edits
2641	5s	96%	4%	03h:38m:59s	33%	1%	0%

Editing Details

Secs/Word	Job ID	Segment ID	Words	Suggestion source	Match percentage	Time-to-edit	Post-editing effort
163.8	984	580414	26	Machine Translation	86%	10m:58s	20%

Segment	You can move a volume to a new I/O group to balance the workload across the in the system without stopping host activity to the volumes.
Suggestion	È possibile spostare un volume di I / E in un nuovo gruppo di bilanciare il carico di lavoro tra i nel sistema host senza dover arrestare l'attività per i volumi.
Translation	È possibile spostare un volume in un nuovo gruppo I/O per bilanciare il carico di lavoro nel sistema senza arrestare l'attività dell'host sui volumi.
Diff View	È possibile spostare un volume di I / E in un nuovo gruppo di I/O per bilanciare il carico di lavoro tra i nel sistema host senza dover arrestare l'attività per i dell'host sui volumi.

Secs/Word	Job ID	Segment ID	Words	Suggestion source	Match percentage	Time-to-edit	Post-editing effort
28.8	984	580432	27	Machine Translation	86%	12m:57s	31%

Mirrored, compressed, and thin-provisioned volumes, as well as volumes that are used in FlashCopy mappings, can be moved to

CAT Tool URL

The first beta release of the CAT tool which only offers limited functionalities tailored to machine translation developers and researchers can be found at the following URL:

<http://demo.matecat.com>

The repository where the source code is stored can be found at the following URL:

https://www.assembla.com/code/matecat_source/git/nodes