



LetsMT!: Do-It-Yourself Machine Translation Factory on the Cloud

**European Commission Competitiveness and Innovation Framework Programme
Call CIP-ICT-PSP.2009.5.1: Machine translation for the multilingual web
Small or medium-scale focused research project (STREP)
Project reference: 250456
<http://www.letsmt.eu>**

List of partners
Tilde, Latvia (coordinator)
University of Edinburgh, Human Communication Research Centre, UK
University of Zagreb, Faculty of Humanities and Social Sciences, Croatia
University of Copenhagen, Centre for Language Technology, Denmark
Uppsala University, Department of Linguistics and Philology, Sweden
Zoorobotics, Netherlands

Project duration: March 2010 — August 2012

Summary

LetsMT! has created a cloud-based platform for generation and running of SMT systems based on public and user-provided training data. Users can upload their parallel corpora to online repository and generate user-tailored SMT systems based on user selected data. LetsMT! hides complexity of MT generation by providing a cloud-based infrastructure and easy user interface to manage data, create and run multiple customized MT engines and use them for various translation needs. LetsMT! includes such features as storing of public and private training data, automated training of SMT systems from specified data, facilities for automated MT evaluation, facilities for running MT systems and web-based translation, API for integration of MT services, user and platform management facilities.

Publicly available parallel resources, such as OPUS, DGT and JRC-Acquis, supplemented by user-provided data, are used in LetsMT! as training data for development of SMT systems. Users can upload their data in a variety of formats (e.g. TMX, XLIFF and Moses formats, parallel documents in PDF, text and DOC formats, compressed gzip, zip and tar archives) that are automatically processed by validation and conversion tools. The system also includes a sentence alignment module for creation of new parallel resources for SMT training from scratch.

LetsMT! uses Moses as a language independent SMT solution and integrates it as a cloud-based service into the LetsMT! online platform. Moses toolkit has been adapted to fit into the rapid training, updating, and interactive access environment. The Moses SMT training pipeline involves a number of steps that each require a separate program to run. In the framework of LetsMT! this process is streamlined and made automatically configurable given a set of user-specified variables (training corpora, data for language model, dictionaries, tuning sets).

LetsMT! translation services can be used in several ways: through the web portal letsmt.com, through a widget for web-page translation, through browser plug-ins, and through integration in computer-assisted translation (CAT) tools SDL Trados and Kilgray memoQ. The platform provides API level access through a web service that (i) provides information about available SMT systems, their metadata and status, (ii) performs translation of text, (iii) allows SMT systems to be managed (load or unload), (iv) authenticates users and controls user access rights.