

Platform for Online Sharing of Training Data and Building User Tailored MT

LetsMT! Consortium

E-mail: project@tilde.lv

www.letsmt.eu

Let's MT!

LetsMT! Mission

- ❑ To fully exploit the huge potential of existing open SMT technologies to build an innovative online collaborative platform for **data sharing and MT building**.

Challenge

- ❑ SMT systems for larger languages are of much better quality compared to systems for smaller languages
- ❑ Quality of SMT systems largely **depends on the size** of training data
- ❑ Parallel data accessible on the web is **just a fraction** of all translated texts. Most of them still reside in the local systems of different corporations, public and private institutions, desktops of individual users
- ❑ Current mass-market and online MT systems are of **general nature** and perform poorly for domain and user specific texts
- ❑ System adaptation is prohibitively **expensive service** not affordable to smaller companies or the majority of public institutions

Proposed solution

- ❑ Build an innovative online collaborative platform for data sharing and MT building
- ❑ Platform will support upload of public as well as proprietary MT training data and building of multiple MT systems, public or proprietary, by combining and prioritizing uploaded data
- ❑ LetsMT! will extend the use of existing state-of-the-art SMT methods that will be applied to data supplied by users to increase quality, scope and language coverage of machine translation
- ❑ The LetsMT! platform will be of particular significance for users of smaller languages where available MT solutions are of unsatisfactory quality or where they do not exist at all

LetsMT! platform features

- ❑ Uploading of parallel texts for users that will contribute their content
- ❑ Directory of web and offline resources gathered by LetsMT! as well as links provided by users to other sources not included in the LetsMT! repository
- ❑ Automated training of SMT systems from specified collections of training data
- ❑ Custom building of MT engines from selected pool of training data, for larger donors or paying customers
- ❑ Custom building of MT engines from proprietary non-public data, for paying customers
- ❑ MT evaluation facilities

LetsMT! platform functionality

- **website for upload** of parallel corpora and building of specific MT solutions
- **website for translation** where source text can be typed and translated
- **translation widget** provided for inclusion into websites to translate their content
- **browser plug-ins** that will provide the quickest access to translation
- **integration in CAT tools** and other applications

Lets MT! main target groups

- ❑ Translation industry
- ❑ Freelance translators
- ❑ Software developers and providers
- ❑ Web developers
- ❑ Public institutions
- ❑ Research community
- ❑ University education
- ❑ General users

Key Impact Areas

- ❑ Significant increase in available **language resources** for training of SMT systems
- ❑ **Improved quality** of SMT, especially for smaller languages
- ❑ Increase in **language coverage** for machine translation
- ❑ **Diversification** of free MT by tailoring for specific domains or user requirements
- ❑ Significant **increase in usage** of MT in web and applications through LetsMT! translation widgets, plug-ins and MT web-service
- ❑ Much wider use and greater impact of available **open-source SMT technologies**
- ❑ **Collaborative involvement** of different stakeholders from public sector, SMEs, universities, research and education community

Partnership

- ❑ Tilde (Project Coordinator, TILDE) - Latvia
- ❑ University of Edinburgh (UEDIN) - UK
- ❑ University of Zagreb (FFZG) - Croatia
- ❑ Kopenhagen University (UCPH) - Denmark
- ❑ Uppsala University (UUP) - Sweden
- ❑ Moravia (MOR) – Czech Republic
- ❑ ZooRobatic (SEM) – The Netherlands

Complementing Competencies



Synergy of Expertises

Software development & commercialization

- **Tilde**, UEDIN, UUP

Expertise in SMT

- **UEDIN**, UCPH, UUP

Expertise in building of parallel corpora

- **UUP**, UCPH

Expertise in under-resourced languages

- **FFZG**, UCPH, Tilde

Expertise in evaluation

- **UCPH**

SMT Users

- **SEM**, **Moravia**, **Tilde**, UCPH, FFZG

Application Scenarios

- ❑ Online MT service for the **localization and** translation industry
- ❑ Online MT service for global **business and financial news**
- + Showcase for patent translations for gisting purposes

Support Group

- ❑ SDI Media: subtitling and language dubbing services
Will evaluate translation quality, integration into SDI Media translation workflow
- ❑ Patent Office of the Republic of Latvia
Will evaluate translation quality, applicability for patent translation
- ❑ TAUS Data Association
- ❑ Directorate General de Traduccion

Support Group's objectives

- ❑ create strong cooperation between related projects and initiatives
- ❑ facilitate sharing of language resources for statistical machine translation
- ❑ involve external experts in requirements specification and evaluation of LetsMT! results
- ❑ share experience and competences in application of machine translation
- ❑ grow LetsMT! potential user community
- ❑ facilitate dissemination of project results

Project ID

- ❑ Funded under: The Information and Communication Technologies Policy Support Programme
- ❑ Area: CIP-ICT-PSP.2009.5.1 Multilingual Web: Machine translation for the multilingual web
- ❑ Total cost: €3.34m
- ❑ EU contribution: €1.67m
- ❑ Project reference: 250456
- ❑ Execution: From 01/03/2010 to 31/08/2012
- ❑ Contract type: The Information and Communication Technologies Policy Support Programme PB Pilot Type B

A series of seven concentric, downward-curving blue lines that create a semi-circular shape, framing the central text.

Thank you and Let's MT!

Let's MT!