



# Technology Showcase Guide

Jennifer DeCamp, Showcase Organizer



The 11th Conference of the Association for Machine Translation in the Americas

Vancouver, BC  
October 22-26  
[amta2014.amtaweb.org](http://amta2014.amtaweb.org)



## AMTA 2014 Technology Showcase

Organization	Exhibit	Booth	Page
Autodesk, Inc.	Term Translation Central	1	1
Basis	Rosette, Highlight	2	2
CAMT	CyberTrans	3	3
CNGL-DCU	Brazilator	4	4
CNGL-DCU	IHearU	5	6
eBay		6	8
FBK	MateCat Tool	7	9
Google	Google Translate	7	11
Iconic Translation Machines	IPTranslator	8	12
Microsoft	Microsoft Translate	10	13
Ntrepid	Virtus Translator	11	14
Pan American Health Organization	PAHOMTS 4.10	12	15
Raytheon BBN Technologies	BBN Web Monitoring System	13	16
Safaba Translation Solutions	Safaba MT v. 2.1	14	17
Sakhr	Sakhr MT	15	18
SDL Government	Government Language Platform (GLP), Translator Online Testing System (TOTS), Social Media Translation Engine (SMTE), Multilingual Chat (MLC)	16	19
Spoken Translation	Converser for Healthcare	17	21
Systran Software Inc.	Systran	18	22
Tilde	Let's MT and TermUnity	19	23
Yandex	Yandex Translate	20	24

**Organization: Autodesk Inc.**

**Exhibit: Term Translation Central**

**Booth: 1**

***For information, please contact: Dr. Ventsislav Zhechev at [ventsislav.zhechev@autodesk.com](mailto:ventsislav.zhechev@autodesk.com) or +41 32 723 91 22.***

One of the challenges of keeping the in-house MT infrastructure up-to-date is making sure it can handle new content. By its nature, new content will contain text that has never been translated before (e.g. as it would discuss new features of the product). Although current Statistical MT systems are quite robust, they cannot make wonders and are prone to produce bad translations for such new content that would then need to be repeatedly corrected by the translators—at least until the MT engines have been retrained. This is particularly problematic for cases where raw MT is published directly and the users of the system are expected to be the post-editors, as this could make the translations of new product features unhelpful or even misleading. To alleviate this issue, at Autodesk we have developed a system that would extract potential terms from new content up front so that they can be pre-translated before the actual content is localised. The pre-translated terms are then picked on the fly by the MT engines without it being necessary to retrain them. This allows us to always have up-to-date product-specific terminology in the MT output.

**Organization:           Basis Technology**

**Exhibit:                 HIGHLIGHT**

**Booth:                   2**

***For information, please contact: Jennifer Flather, Basis Technology Product Manager for HIGHLIGHT, at [jflather@basistech.com](mailto:jflather@basistech.com) or (202) 685-6783.***

HIGHLIGHT standardizes person and place names appearing in documents or spreadsheets according to the Intelligence Community (IC) Transliteration Standards for seven languages. Changes since HIGHLIGHT was shown at AMTA 2012:

- New languages: Mandarin, Korean, and Russian (in addition to Arabic, Dari, Farsi, and Pashto)
- Automatically finds names of person and places in text documents and links to existing or custom entity databases, with easy interactive correction.
- Three choices of operation (Automatic, Interactive, and Hybrid)
- Supports standardization of names originating in one document into another document
- HIGHLIGHT Pad: tailored for use with Computer Assisted Translation (CAT) tools. Users copy text from a CAT tool, paste it into HIGHLIGHT Pad, apply name standardization, then paste the results back into the CAT tool.
- Return on Investment Analytic dashboard.

**Organization:** United States Department of Defense  
Center for Applied Machine Translation  
(CAMT)

**Product:** CyberTrans

**Booth:** 3

**Organization:** Centre for Global Intelligent Content,  
School of Computing, Dublin City  
University (CNGL-DCU)

**Product:** Brazilator

**Booth:** 4

*For information, please contact: Lamia Tounsi at [lamia.tounsi@computing.dcu.ie](mailto:lamia.tounsi@computing.dcu.ie).*

The 2014 FIFA World Cup has become the biggest event yet for Twitter with 672 million tweets. Brazilator provides a live translation stream of tweets relating to this event and a live stream sentiment analysis of these tweets.

We covered 13 languages and 24 language pairs (both directions between English and Irish, Spanish, Italian, German, French, Croatian, Chinese, Japanese, Korean, Farsi, Portuguese, Greek). During the final match between Germany-Argentina, 6.5M tweet translations were requested.

We faced two main challenges dealing with social media text

- Non-standard text usage: Tweets typically contain noise, diverse styles, misspellings, abbreviations, web links, emoticons and hashtags , etc
- Volume vs. real time: Microsoft provided us with the infrastructure to build robust and efficient machine translation systems using the Microsoft Translator Hub. Twitter then directly invoked these systems.

We carried out data-based domain adaptation of SMT in a 4 step process:

1. Crawling in-domain parallel data with a domain-focused parallel crawler (ILSP-FC) from FIFA and UEFA websites
2. Collecting in-domain in-style monolingual data: #WC tweets
3. Building in-domain parallel glossaries (~700 soccer-related terms) by native speakers
4. Translating tuning and test sets (1K sentence pairs) by native speakers or via crowdsourcing

**Best MT results (BLEU scores) e.g. EN->GA 27.96, GA->EN 29.84**

Base-line Bing	GA	ES	IT	DE	FR	HR	ZH	JP	KO	FA	PT	EL
EN	n/a	-2.42	+0.17	-1.31	+7.85	-1.69	-0.3	-0.13	-3.35	-0.65	+3.92	+0.81
	n/a	-6.48	-12.6	-8.02	-5.21	-8.12	n/a	-3.19	-2.64	-9.12	-13.38	-11.84

Brazilator	GA	ES	IT	DE	FR	HR	ZH	JP	KO	FA	PT	EL
EN	27.96	47.02	37.05	34.81	66.34	27.14	19.77	16.47	8.89	22.53	52.71	33.18
	29.84	53.26	58.85	51.42	56.35	43.54	28.48	27.13	10.2	33.00	65.28	53.39

CNGL-DCU Team: Santiago Cortes, Piyush Arora, Chris Hokamp, Federico Fancellu, Alex Killen, Lamia Tounsi, Antonio Toral, Ankit Srivastava, Maria Alecu, Iacer Calixto, Sheila Castilho, Keith Curtis, Federico Gaspari, Akira Hayakawa, Teresa Lynn, Peyman Passban, Eziz Tursun, Ali Hosseinzadeh Vahid, Xiaofeng Wu, Xiaojun Zhang, Debasis Ganguly, Louise Irwin, Anna Kostekidou, Liangyou Li, Tsuyoshi Okita, Ximo Planells, David Racca, Joris Vreeke, Jian Zhang, Andy Way — Microsoft Research: Will Lewis, Declan Groves, Federico Garcea, Chris Wendt



**Organization:** Centre for Global Intelligent Content,  
School of Computing, Dublin City  
University (CNGL-DCU)

**Product:** IHearU

**Booth:** 5

For information, please contact: Xiaofeng Wu at xiaofeng.wu11@gmail.com.

IHearU is a handsfree Android app that allows people to interact and share content on Twitter using only their voice. The IHearU technology allows users to compose/reply twitter/direct messages by uploading real human voice messages; to translate non English tweets into English and to listen to twitter/direct messages, either in the original human voice which uploaded by IHearU or TTSed voice generated from text twitter.

The targeted audiences for this project are blind/visually impaired users, mobility impaired users with hand or arm limitations, older people struggling with technology, and users in any hands full activity such as playing golf, jogging, driving, etc.

The four most important features of IHearU are as follows:

1. Userfriendly (HandsEyesFree): Only very simple set of vocal commands are used;
2. Human Real Voice Upload: IHearU will upload human real voice, not the unpredictable ASR result.
3. Cloudbased translation system (CNGL Brazilor Engine): If a tweet is related to sport, IHearU will use CNGL's Brazilor translation engine to translate this tweet from 12 different languages into English, including Irish, German, French, Spanish, Italian, Portuguese, Croatian,Greek, Japanese, Korean, Chinese, and Farsi. For other languages and domains, IHearU will use Microsoft Bing translators.
4. Sentimental TTS: Each twitter will be classified with the CNGL sentimental classifier and the TTS will be configured with "happy/unhappy/neutral".





CNGL-DCU Team: Xiaofeng Wu, Santiago Cortes, Soham Alat, Lamia Tounsi, Joris Vreeke, Xiaojun Zhang, Qun Liu

**Organization: eBay**

**Product:**

**Booth: 6**

*For information, please contact: Hassan Sawaf at [hsawaf@ebay.com](mailto:hsawaf@ebay.com).*

**Organization:** FBK, Trento, Italy

**Product:** MateCat Tool

**Booth:** 9

***For information, please contact: Marcello Federico at federico@fbk.eu.***

We present the final release of the MateCat Tool, a new web-based CAT tool providing translators with a professional work environment, integrating translation memories, terminology bases, concordancers, and machine translation. The tool is completely developed as open source software and has been already successfully deployed for business, research and education. The MateCat Tool represents today probably the best available open source platform for investigating, integrating, and evaluating under realistic conditions the impact of new machine translation technology on human post-editing. A prototype version of the tool was already presented at AMTA 2012, which however was missing most of the interesting MT-features (such as online MT adaptation).

In the third year of the project, the MateCat tool was further developed to reach a high level of maturity and solidity, which makes it suitable for professional use. Translated S.r.l., the commercial partner of the consortium, has switched its entire production to MateCat, translating over 35 million words with it. In the last few months, over 3,000 translators have started using the MateCat tool for professional work.

Links to other presentations at the 2014 AMTA Conference:

Prashant Mathur, Cettolo Mauro, Marcello Federico and José G. C. de Souza. Online Multi-User Adaptive Statistical Machine Translation. 25 October 2014, 10 AM, Harbourside Ballroom

Mihael Arcan, Marco Turchi, Sara Tonelli, Paul Buitelaar. Enhancing Statistical Machine Translation with Bilingual Terminology in a CAT Environment. 23 October 2014, 2.30 PM, Harbourside Ballroom

Mauro Cettolo, Nicola Bertoldi and Marcello Federico. The Repetition Rate of Text as a Predictor of the Effectiveness of Machine Translation Adaptation. Research poster (25 October 2014, 2 PM – 3.30 PM, Harbourside Ballroom)

MateCat: an Open Source CAT Tool for MT post-editing. Marcello Federico, Nicola Bertoldi, Marco Trombetti, and Alessandro Cattelan. Post-Conference Tutorial (26 October 2014, 2.00 PM – 5.30 PM, Harbourside Ballroom)

Marcello Federico. User-Adaptative MT in the MateCat Tool. Invited Talk at the IAMT Workshop (22 October 2014, 12 PM, Harbourside Ballroom)

José G. C. de Souza, Marco Turchi and Matteo Negri. Towards a Combination of Online and Multitask Learning for MT Quality Estimation: a Preliminary Study. Poster at the IAMT Workshop (22 October 2014, 4 PM, Harbourside Ballroom)

Prashant Mathur and Mauro Cettolo. Optimized MT Online Learning in Computer Assisted Translation. Poster at the IAMT Workshop (22 October 2014, 4 PM, Harbourside Ballroom)

**Organization:** Google Translate

**Product:** Translate Community

**Booth:** 7

*For more information, please contact Keith Stevens at [kstevens@google.com](mailto:kstevens@google.com).*

Google Translate generates billions of translations per day for users worldwide in over 80 languages. For some of our languages, our translations are robust and grammatical. However, for most language pairs our translations leave a lot of room for improvement. A major source of this problem: a lack of high quality data.

Languages such as Hindi, Bengali, Khmer, Lao, Sinhala, Tibetan, Mongolian, and most of the world's languages lack sufficient data to train high quality Statistical Machine Translation models. There are no shared corpora for research on these languages, further compounding investigation into novel models for each language.

To work towards a better data future, we've created Translate Community, a platform for language enthusiasts. Together, our users can create crowdsourced corpora for all of the languages supported by Google Translate and any of the languages we're researching. Together, we're going to build a shared corpus of parallel documents that can be used for building initial SMT models, both inside Google and at other research institutions.



**Organization:           Iconic Translation Machines**

**Product:                 IPTranslator**

**Booth:                    8**

***For information, please contact: John Tinsley at [john@iptranslator.com](mailto:john@iptranslator.com).***

Iconic Translation Machines is a provider of high-quality cloud-based Machine Translation solutions for the Language Services industry that have been domain-adapted to specific technical areas, providing best-in-class performance and delivering significant gains in translator productivity.

Our mission is to open up Machine Translation to new industries and technical content types that are not traditionally accessible to language technology. We aim to do this through our Ensemble Architecture™ approach to MT and, in doing this, we intend to lower the barrier for entry to adopting Machine Translation by making high-quality solutions instantly available and easy to use.

IPTranslator is our flagship product for the patent and legal translation industry. It is ready to use right now! We have already built and fine-tuned Machine Translation systems across a number of technical domains and languages to deliver high-quality translations immediately from the cloud.

Our Instant Access MT lowers the bar for adopting Machine Translation. It allows you to evaluate our offerings straight away with no up-front cost for development and no waiting around.

During the technology showcase, we will demonstrate how the service works and answer questions on the technology.

On the same day as the showcase, Oct 24th, at 12:00 will we also be discussing this technology during the commercial track in a presentation entitled “From the Lab to the Market: Commercializing MT Research”.

**Organization: Microsoft**

**Product: Microsoft Translator**

**Booth: 10**

***For information, please contact: Chris Wendt or Christian Federmann at translator@microsoft.com.***

Microsoft Translator is a large-scale web service that powers translation in a variety of scenarios, in Microsoft Products and many others. Newest addition to the Translator service is speech translation, demonstrated via Skype. We describe the technology behind the speech translation service, the problems addressed by the implementation, and an outlook of what comes next. We give a demo of Skype Translator, and let the showcase visitor try it themselves.

Besides speech translation, we will walk visitors through the concept of P3 “Post-Publish Post-Editing”, using highly customized machine translation in combination with light and targeted human post-editing. The Microsoft Translator Hub provides the mechanism for a high degree of customization, works even with small amounts of custom training data. It is complemented by the collaborative translation framework for instant customization, receiving and managing community feedback on the automatic translation as well as corrections and votes.

**Organization: Ntrepid**

**Exhibit: Virtus Translator**

**Booth: 11**

***For information, please contact: Mark Arehart at [mark.arehart@ntrepidcorp.com](mailto:mark.arehart@ntrepidcorp.com).***

Virtus Translator is a customizable MT/CAT tool for translating structured data as found in databases and spreadsheets. Such data tends to be voluminous, repetitive, and tedious for a human translator. Virtus is optimized for certain data categories including persons, organizations, locations, biographical information (gender, marital status, occupation), industries, vehicles, and colors. Virtus imports data from Excel or delimited text files and auto-detects character encoding, language, and data category. The first line of translation is TM, and users can create their own TM to enhance or override the system TM. Users can also choose how the TM should be applied for each column of data being processed. Data not covered by the TM is translated using semantically annotated bilingual lexicons, category-specific translation rules, and transliteration rules for multiple transliteration standards. Users desiring an extra level of flexibility can override a variety of default translation settings per language and category. Virtus translates into English from Arabic, Chinese, French, Persian, Russian, and Spanish. Virtus is available as a desktop application and as a server with a desktop client. The server version includes an API that allows for integration into an existing data processing workflow.



**Organization: Pan American Health Organization (PAHO)**

**Exhibit: PAHOMTS®**

**Booth: 12**

*For information, please contact: Hermes Camelo at [camelohe@paho.org](mailto:camelohe@paho.org) or (202) 974-3783.*

PAHOMTS® is a fully automatic Machine Translation (MT) program developed in-house by a small team of computational linguists. It has been operational at PAHO since 1980. Translators postedit the raw output to produce high-quality translations with a 30-50% gain in productivity and a 33% cost savings.

At PAHO, MT has proven to be especially effective for conference documents, scientific papers, training materials, and technical abstracts. The software currently translates in any language combination of English, Spanish, and Portuguese. Each of its linguistically rich dictionaries are fully customizable and contains over 150,000 words, phrases, and context-sensitive rules. Many of these entries are from medical and public health domains.

PAHOMTS® comes with a translation API that can be integrated with Translation Memory systems. It creates aligned translations that can be fed into a translation memory or a bilingual corpus. The output preserves the format of the original document, including tables, and is easy to polish using PAHO's postediting macros.

PAHOMTS® is available for licensing by international and intergovernmental organizations, NGOs, educational institutions, and government agencies only.

New in this release:

- Microsoft Office 2010/2013 Add-Ins for Word, PowerPoint, and Excel
- Thousands of new dictionary entries
- Various enhancements and bug fixes.

**Organization: Raytheon BBN Technologies Corp.**

**Exhibit: BBN Broadcast Monitoring System for  
Open Source Media Analytics**

**Booth: 13**

*For information, please contact: Martha Lillie at [mlillie@bbn.com](mailto:mlillie@bbn.com).*

BBN Web Monitoring System for open source social media analytics. Under a recent effort with the CTTSO, BBN has integrated a variety of Human Language Technologies to support analysis of foreign language social media content, including automatic topic discovery, sentiment analysis, geolocation, entity extraction, and network graph analysis.



**Organization: Safaba Translation Solutions**

**Exhibit: Safaba Machine Translation: EMTGlobal 4.0**

**Booth: 14**

***For information, please contact: Dr. Alon Lavie at +1 (412) 983-2408 or at [alon.lavie@safaba.com](mailto:alon.lavie@safaba.com).***

Description

**SAFABA IS MACHINE TRANSLATION FOR GLOBAL BUSINESS.**

Safaba, the technology leader in machine translation (MT), provides global enterprises with automated translation solutions that deliver superior translation quality and simplify the path to global enterprise presence unlike any other translation solution.

Safaba solves the translation problems unique to each enterprise – continual content change in structured, nuanced language.

Global enterprises want to:

- > Launch products in new markets twice as fast...
- > Localize all corporate communication channels...
- > Dramatically reduce global operations costs...
- > Cut translation costs in half!

Safaba delivers:

- > Highly accurate translations in corporate language and style
- > Easy integration into corporate applications
- > Dramatically reduced translation and operational costs
- > Results!

Safaba (Hebrew: *SAFA BA*, meaning “Language Inside”) was founded by Drs. Alon Lavie and Robert Olszewski in Pittsburgh PA, an internationally recognized global hub of language technologies innovation and home to the world-renowned Language Technologies Institute (LTI) of Carnegie Mellon University (CMU).

**Organization: Sakhr**

**Exhibit: Sakhr MT**

**Booth: 15**

*For information, please contact: Kemp Gouldin at [gouldin@sakhr.us](mailto:gouldin@sakhr.us).*

**Organization: SDL Government, Inc.**

**Exhibit: Government Language Platform (GLP),  
Translator Online Testing System (TOTS),  
Social Media Translation Engine (SMTE),  
Multilingual Chat (MLC)**

**Booth: 16**

*For information, please contact: Melchior Baltazar, CEO, at  
mbaltazar@sdlgov.com or 703-851-6530.*

SDL Government (SDLGov) is a technology and services company that provides language translation and strategic communications solutions deployed by government organizations worldwide.

SDLGov language solutions are easily adaptable for specific domains and custom applications, providing a universal COTS architecture that can rapidly deliver linguistic technology tailored to a specific deployment scenario, including Training, HA/DR, AT/FP, CT, CI, OSINT, SOMEX and DOMEX multinational or coalition operations. Our solutions are securely deployed on-premise, behind an organization's firewall.

Exhibits include:

**1. Government Language Platform (GLP)**

The SDL Government Language Platform ("GLP") provides a comprehensive language translation solution that combines translation memory, terminology, workflow, and productivity technologies to provide superior linguistic output quality and consistency. The platform consists of the following components:

- **Language Weaver Enterprise Translation Server-Government (ETS-G)**, the machine translation platform that provides language engines in over 44 languages.
- **World Server-Government (WS-G)**, the linguistic management platform that consists of translation memory, terminology database, content integration and workflow management modules.
- **Trados Studio-Government (TS-G)**, the translation productivity suite that provides tools used by translators and editors in offline translation scenarios and integrates the translation supply chain into the overall translation process.

**2. Translation Online Testing System (TOTS)** TOTS is an automated evaluation system that instantly scores translation tests. With funding from DARPA and TSWG/CTSO, SDL Government has leveraged its translation technology to build

a solution that helps clients assess candidates most likely to perform at the desired ILR level. TOTS is capable of producing ILR ratings that closely match test results generated by a human rater for the same test.

3. **Social Media Translation Engine (SMTE).** SMTE provides fast and accurate translation of Tweets through a state of the art Statistical Machine Translation (SMT) approach that incorporates a supervisory rules engine and advanced terminology management tools. With funding from DARPA and TSWG/CTSO, SDL Government has leveraged its translation technology to developed the first commercially available informal machine translation language pair for Arabic Twitter.
4. **Multilingual Chat (MLC).** MLC provides instantaneous multilingual chat capabilities that allow multinational participants to communicate in their native language. Leveraging its translation technology, SDL Government has developed MLC to enable instant chat translation for one-to-one or one-to-many communications.

Our flagship offering, the Government Language Platform (GLP), provides a comprehensive language translation solution that combines translation memory, terminology, workflow, productivity and statistical machine translation technologies. SDLGov software solutions are designed to stand-alone, or to easily integrate with existing applications through SOAP and REST APIs, allowing organizations to add multilingual capabilities to the processes and operations they already have in place. Our language pairs cover 95% of the world's population.

**Organization: Spoken Translation**

**Exhibit: Converser for Healthcare**

**Booth: 17**

***For information, please contact: Mark Seligman at [mark.seligman@spokentranslation.com](mailto:mark.seligman@spokentranslation.com) or 510 843-9800 (direct) / 510 604-0025 (mobile).***

We'll demo Converser for Healthcare 4.0, a speech translation system, featuring new training-free speech recognition and improved interface for verification and correction of translation. Implementation issues will be discussed in Seligman and Dillinger, "Behind the Scenes in an Interactive Speech Translation System." We'll also demonstrate several multilingual and multiplatform speech translation offerings.

## **SYSTRAN Software, Inc.**

### **SYSTRAN Enterprise Server 8, Training Server in the Cloud, and SYSTRAN LDK**

#### **Booth 18**

*For information, please contact: David Hetland at [David.Hetland@systran.us](mailto:David.Hetland@systran.us) or 858-457-1900 ext. 419.*

SYSTRAN's market-leading language translation software solutions empower businesses and customers to understand multilingual information in real-time and communicate more effectively in 128+ language combinations. SYSTRAN's automatic customizable translation solutions are used to enhance customer support by delivering instant translations of knowledge base articles, email and chat, search, and any other application or business initiative to improve global customer satisfaction and retention. Use of SYSTRAN's solutions increases user productivity and time-savings while reducing costs.

#### **Featured Products**

##### **SYSTRAN Enterprise Server 8**

SYSTRAN Enterprise server 8 is the only comprehensive solution able to meet the full range of translation needs on an enterprise scale. SYSTRAN Enterprise Server 8 is a hybrid machine translation engine, which combines the best attributes of rule-based and statistical machine translation, resulting in a product that meets corporate expectations for quality, security, cost-effectiveness and productivity.

##### **Training Server in the Cloud**

At the 2012 AMTA, SYSTRAN exhibited SYSTRAN Enterprise Server 7 and the accompanied Training Server. With this package, corporate customers and Language Service Providers use the training server to independently train SYSTRAN Enterprise Server 7 to any selected domain or business objective and to produce publishable-quality translations. Through the use of the SYSTRAN Training Server, SYSTRAN Enterprise Server 7 automatically learns from existing and validated translations and updates itself as these translations are reused.

At this year's AMTA, SYSTRAN will introduce Training Server in the Cloud. Compared to the Training Server offered with SYSTRAN Enterprise Server 7, Training Server in the Cloud will provide faster, more advanced translation quality without the need of an enterprise server.

##### **SYSTRAN LDK**

On the basis of our abundant natural language processing technology and language resources, SYSTRAN provides not only the products and services but also Linguistic Development Toolkit that can be used by various external developers or cooperation partners to make use of the



engine. SYSTRAN LDK includes detailed modules of our natural language processing technology and all of the language pairs that our company supports.

*John Paul Barraza, Vice President of Operations, will present **Optimizing Large Post-Editing Project Workflows** at 4 p.m. Oct. 22, Harbourside Ballroom.*

**Organization: Tilde**

**Exhibit: Let's MT and TermUnity**

**Booth: 19**

*For information, please contact: Andrejs Vasiljevs (CEO) at [andrejs@tilde.com](mailto:andrejs@tilde.com)*

*Vienības gatve 75a*

*Rīga, Latvia LV1004*

*+371 2948 4754 (mobile)*

*+371 6760 5001 (corporate)*

At the core of Tilde's custom MT solutions is the synergy of the machine translation platform LetsMT® and the terminology service platform TermUnity, created in collaboration with the Universities of Edinburgh, Sheffield, Uppsala, Copenhagen, and others.

LetsMT® is a Moses-based cloud platform for gathering and processing public and user-provided multilingual data. The platform allows users to generate multiple MT systems by combining and prioritizing this data and applying language specific processing components.

Particular aspects of the demonstration will be methods for improving MT quality of morphologically rich languages, processing and preserving the formatting tags, and automating extraction and alignment of text from multilingual documents in different formats.

TermUnity provides the infrastructure to automate terminology identification, acquisition, and processing tasks. The automation of individual tasks is provided as a set of interoperable, cloud-based services integrated into workflows for identification of term candidates in user-provided documents, as well as the lookup of translation equivalents in online terminology resources – and on the Web – by automatically extracting multilingual terminology from comparable and parallel online resources.

The synergy of these services will be demonstrated by several methods for using terminology data for customization and quality improvement of domain-specific statistical machine translation. Training-level integration includes enrichment of monolingual and parallel data with terminology, adaptation of translation model by adapting and filtering phrase tables, and adaptation of language models. Translation-level integration is provided by pre-processing the source text to identify terms and mark-up them with translation hypotheses. Evaluation results show that the combination of these methods significantly improves translation quality.

Several use cases and solutions in government and several industries will be demonstrated.

The LetsMT® and TermUnity services have not been exhibited at previous AMTA conferences.

**Organization: Yandex**

**Exhibit: Yandex Translate**

**Booth: 20**

***For information, please contact: Farkhat Aminov at [aminov@yandex-team.ru](mailto:aminov@yandex-team.ru) or +7 (495) 739-70-00 (ext. 6049).***

Yandex.Translate ([translate.yandex.com](http://translate.yandex.com)) is a phrase-based SMT system available as a web service, mobile applications and a full-functional API. Yandex.Translate currently supports more than 40 languages and processes billions of translations daily.

Since the previous AMTA conference our team has added support for more than 30 new languages. Translations can now be done from any to any language. In addition, API is now available as a paid solution with SLA.