

Progress in ModernMT, a New Open-Source Machine Translation Platform for the Translation Industry

<http://www.modernmt.eu>

The ModernMT Consortium:

Translated srl, Rome, Italy

Fondazione Bruno Kessler (FBK), Povo, Italy

University of Edinburgh, Edinburgh, Scotland, Europe

Translation Automation User Society (TAUS), Amsterdam, Netherlands

Corresponding author: Ulrich Germann (ugermann@inf.ed.ac.uk)

Abstract

We report progress made in Year 2 of ModernMT, a three-year EU *Horizon 2020 Innovation Action* (2015–2017) that develops new open-source machine translation technology for use in translation production environments. ModernMT is designed to facilitate both fully automatic translation and interactive post-editing scenarios.

1 Project Goals

ModernMT aims to improve the state of the art in open-source machine translation technology by developing scalable, cloud-ready software that offers the following benefits.

- A **simple installation** procedure for turn-key RESTful¹ machine translation services.
- **Very fast set-up times** for systems built from scratch using existing parallel corpora (e.g., translation memories). Incoming data can be ingested at approximately the same speed at which it is uploaded.
- **Immediate integration of new data** (e.g., from newly post-edited MT output). Rebuilding or retuning the system will not be necessary.
- **Instant domain adaptation** by considering translation context beyond the individual sentence, without the need for domain-specific custom engines. The ModernMT system uses the translation input (from a single translation unit to an entire document), as well as additional context keywords (if provided by the user) to retrieve similar texts from its bitext database and to bias translations towards the style and lexical choice of these similar texts.

- **High scalability** with respect to throughput, concurrent users, and the amount of data the system can handle.

In addition, ModernMT is actively collecting, curating, cataloguing, and — where possible — releasing parallel data from web crawls and parallel data contributions from translation stakeholders, so that ModernMT users have access to data to build their own custom systems. Furthermore, additional data is being collected to set up a new MT service provider that offers high-quality MT services at an affordable price to MT users who prefer not to have to maintain their own systems.

2 Project Phases

The current roadmap of ModernMT can be described as follows.

Year 1 was dedicated to integrating existing statistical machine translation technology, mostly based on the *Moses* toolkit,² and prototyping of instant system adaptation and dynamic model updates.


Year 2 saw the development of a cloud-ready infrastructure and successful integration of adaptation and instant updates into the system. This included development of new database-backed back-ends for the language and translation models.

Year 3 will put focus on development of a ready-to-launch product and investigations into Deep Learning for use within the framework of ModernMT.

3 ModernMT is Open-source

The software is available at <https://github.com/ModernMT/MMT>.

Acknowledgements

 ModernMT has received funding from the European Union's *Horizon 2020 research and innovation programme* under Grant Agreement No 645487 (call ICT-17-2014).

© 2017 The authors. This article is licensed under a Creative Commons 3.0 licence, no derivative works, attribution, CC-BY-ND.

¹https://en.wikipedia.org/wiki/Representational_state_transfer

²<http://www.statmt.org/moses>