Edinburgh Research Explorer

Modern MT: A New Open-Source Machine Translation Platform for the Translation Industry

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Baltic Journal of Modern Computing

OPEN ACCESS

Modern MT: A New Open-Source Machine Translation Platform for the Translation Industry

U. GERMANN[1], E. BARBU[2], L. BENTIVOGLI[3], N. BERTOLDI[3],
N. BOGOYCHEV[1], C. BUCK[1], D. CAROSELLI[2], L. CARVALHO[4],
A. CATTELAN[2], R. CATTONI[3], M. CETTOLO[3], M. FEDERICO[3],
B. HADDOW[1], D. MADL[1], L. MASTROSTEFANO[2], P. MATHUR[3],
A. RUOPP[4], A. SAMIOTOU[4], V. SUDHARSHAN[4],
M. TROMBETTI[2], J. van der MEER[4]

[1] University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom
[2] Translated srl, Via Nepal, 29, 00144 Rome, Italy
[3] Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Povo, Italy
[4] TAUS B.V., Oosteinde 9, 1483 AB De Rijp, Netherlands

ugermann@inf.ed.ac.uk

Abstract. *Modern MT* (www.modernmt.eu) is a three-year Horizon 2020 *innovation action* (2015–2017) to develop new open-source machine translation technology for use in translation production environments, both fully automatic and as a back-end in interactive post-editing scenarios. Led by Translated srl, the project consortium also includes the Fondazione Bruno Kessler (FBK), the University of Edinburgh, and TAUS B.V. *Modern MT* has received funding from the *European Union's Horizon 2020 research and innovation programme* under Grant Agreement No645487 (call ICT-17-2014).

Project Description

Modern MT aims to improve the state of the art in open source machine translation software by developing cloud-ready software that offers

- A **simple installation** procedure for a ready-to-go, REST-based translation service.
- **Very fast set-up times** for systems built from scratch using existing parallel corpora (e.g., translation memories). The goal is to process incoming data at approximately the 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.
- **High scalability** with respect to throughput, concurrent users, and the amount of data the system can handle.

A first version of the software is available at https://github.com/ModernMT/MMT.

Modern MT is also actively **collecting and curating parallel data** for internal use and public release from web crawls and contributions from translation stakeholders, to improve MT quality for everyone.