TILBURG ◆ UNIVERSITY

Tilburg University

TraMOOC: Translation for Massive Open Online Courses

Kordoni, Valia; Cholakov, Kostadin; Egg, Markus; Way, Andy; Birch, Lexi; Kermanidis, Katia; Sosoni, Vilelmini; Tsoumakos, Dimitrios; van den Bosch, Antal; Hendrickx, I.H.E.; Papadopoulos, Michael; Georgakopoulou, Panayota; Gialama, Maria; van Zaanen, Menno; Buliga, Ioana; Jermol, Mitja; Orlic, Davor

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

Link to publication in Tilburg University Research Portal

TraMOOC: Translation for Massive Open Online Courses

List of partners
Humboldt Universität zu Berlin (UBER), Germany (coordinator)
Dublin City University (DCU), Ireland
The University of Edinburgh (UEDIN), UK
Ionian University (IURC), Greece
Stichting Katholieke Universiteit (Radboud University & Radboud UMC), The Netherlands
EASN Technology Innovation Services BVBA (EASN TIS), Belgium
Deluxe Media Europe Ltd (Deluxe Media Europe Ltd), UK
Stichting Katholieke Universiteit Brabant Universiteit van Tilburg, The Netherlands
IVERSITY GMBH (iversity.org), Germany
KNOWLEDGE 4 ALL FOUNDATION (K4A), UK

Project duration: February 2015 — January 2018

Summary

Massive open online courses have been growing rapidly in size and impact. TraMOOC aims at developing high-quality translation of all types of text genre included in MOOCs from English into eleven European and BRIC languages (DE, IT, PT, EL, DU, CS, BG, CR, PL, RU, ZH) that are hard to translate into and have weak MT support. Phrase-based and syntax-based SMT models will be developed for addressing language diversity and supporting the language-independent nature of the methodology. For a high quality, automatic translation approach and for adding value to existing infrastructure, extensive and advanced bootstrapping of new resources will be performed. An innovative multi-modal automatic and human evaluation schema will further ensure translation quality. For human evaluation, an innovative, strict-access control, time- and cost-efficient crowdsourcing setup will be used. Translation experts, domain experts and end users will also be involved. Separate task mining applications will be employed for implicit translation evaluation: (i) topic detection will be applied to source and translated texts and the resulting entity lists will be compared, leading to further qualitative and quantitative translation evaluation results; (ii) sentiment analysis performed on MOOC users' blog posts will reveal end user opinion/evaluation regarding translation quality. Results will be combined into a feedback vector and used to refine parallel data and retrain translation models towards a more accurate second phase translation output. The project results will be showcased and tested on the Iversity MOOC platform and on the VideoLectures.net digital video lecture library.