MomenT-2019 – The Second Workshop on Multilingualism at the intersection of Knowledge Bases and Machine Translation

https://moment2019.insight-centre.org/

19–23 August, 2019
Dublin, Ireland
Moment-2019 – The Second Workshop on Multilingualism at the intersection of Knowledge Bases and Machine Translation

https://moment2019.insight-centre.org/

19–23 August, 2019
Dublin, Ireland

© 2019 The authors. These articles are licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.
Preface from the co-chairs of the workshop

The increasing availability of knowledge bases (KBs), generated by academia and industry, attracts the attentions of researchers in several natural language processing (NLP) tasks with the aim of advancing state-of-the-art performance by making use of the vast amount of background knowledge available in on the web. However, most of the information that is found in KBs, like knowledge graphs, ontologies or terminological resources, is represented, in most of the cases, in one language only (e.g. English, German or Italian). Consequently, NLP applications that use these KBs are therefore limited to the language in which the information is stored. To make the information accessible beyond language borders, these KBs have to be translated into different languages. Since a manual enhancement of KBs is a very time-consuming and expensive process, machine translation (MT) can be applied for this purpose. Nevertheless, this translation task with MT is rather challenging, due to the sophisticated information of a certain domain knowledge, documented in knowledge graphs, the specific vocabulary in terminological dictionaries and the particular sentence structure of ontology labels.

In addition to the multilingual enhancement of monolingual KBs, a growing attention has also been paid to the integration of existing multilingual terminological knowledge into MT systems or computer-assisted translation (CAT) tools. An important open issue for this task is how to support translators with relevant information when dealing with specialised texts from different domains (IT, medical, law, etc.). Commercial or open source MT systems trained on generic data are the most common solutions, but they often struggle with the translation of the specific vocabulary found in this task. To reduce the post-editing effort involved in the translation process, a valuable alternative is to enhance the systems with existing multilingual knowledge, e.g. IATE or in-house terminological resources, which are agreed and curated resources that professionals use in their expert-to-expert communication. In this sense, the provision of multilingual KBs, which could be better integrated into the MT systems is a crucial step towards increasing the translation quality, since terminological expressions are among the most common sources of translation errors.

Workshop organizers
Organizers

Workshop Chairs

Mihael Arcan  
Insight Centre for Data Analytics, National University of Ireland Galway, Ireland

Marco Turchi  
Fondazione Bruno Kessler (FBK), Italy

Jinhua Du  
Investment AI, AIG, UK

Dimitar Shterionov  
ADAPT Centre, Dublin City University, Ireland

Daniel Torregrosa  
Insight Centre for Data Analytics, National University of Ireland Galway, Ireland
Program Committee

Eneko Agirre University of the Basque Country
Juan Alberto Alonso Lucy Software Ibérica SL
Alexandra Birch The University of Edinburgh
Bram Bulte Vrije Universiteit Brussel / VUB Centre for Linguistics (CLIN)
Christian Chiarcos Goethe-Universität Frankfurt
Marta Ruiz Costa-jussà Universitat Politècnica de Catalunya (UPC)
Béatrice Daille Laboratoire d’Informatique Nantes Atlantique (LINA)
Thierry Declerck Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH)
Mauro Dragoni Fondazione Bruno Kessler (FBK)
Tomaž Erjavec Jožef Stefan Institute
Cristina España-Bonet Saarland University / Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH)
Christian Federmann Microsoft
Natalia Grabar STL CNRS Université Lille 3
Barry Haddow The University of Edinburgh
Jochen Hummels ESTeam
Miloš Jakubíček Lexical Computing
Michael Jellinghaus European Commission, DGT
Ilan Kernerman K Dictionaries
Els Lefever University College Ghent / Ghent University
Lieve Macken Ghent University
Mirjam S. Maučec University of Maribor
John P. McCrae National University of Ireland, Galway
Johanna Monti L’Orientale” University of Naples
Diego Moussallem University of Paderborn
Roberto Navigli Sapienza University of Rome
Matteo Negri Fondazione Bruno Kessler (FBK)
Mariana Neves German Federal Institute for Risk Assessment
Axel-Cyrille Ngonga Ngomo Agile Knowledge Engineering and Semantic Web (AKSW)
Tony O’Dowd KantanMT
Constantin Orasan University of Wolverhampton
Márccis Pinnis Tilde
Juan Pino Facebook
Georg Rehm Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH)
Jean Senellart SYSTRAN
Khalil Sima’an Institute for Logic, Language and Computation, University of Amsterdam
Arda Tezcan Ghent University
Josef van Genabith Saarland University / Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH)
Vincent Vandeghinste Katholieke Universiteit Leuven
Špela Vintar University of Ljubljana
Michał Ziemski WIPO Translate
Contents

WordNet Gloss Translation for Under-resourced Languages using Multilingual Neural Machine Translation
   Bharathi Raja Chakravarthi, Mihael Arcan and John P. McCrae  1

Leveraging SNOMED CT terms and relations for machine translation of clinical texts from Basque to Spanish
   Xabier Soto, Olatz Perez-De-Viñaspre, Maite Oronoz and Gorka Labaka  8