

LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

**6th International Workshop on Computational Terminology
(COMPUTERM 2020)**

PROCEEDINGS

Batrice Daille, Kyo Kageura, Ayla Rigouts Terry (eds.)

Proceedings of the LREC 2020
6th International Workshop on Computational Terminology
(COMPUTERM 2020)

Edited by: Béatrice Daille, Kyo Kageura, and Ayla Rigouts Terryn

ISBN: 979-10-95546-57-3

EAN: 9791095546573

For more information:

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: lrec@elda.org

© European Language Resources Association (ELRA)

These Workshop Proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License

Introduction

The aim of this sixth Computerm workshop is to bring together Natural Language Processing (NLP) and Human Language Technology researchers as well as terminology researchers and practitioners to discuss recent advances in computational terminology and its impact within automatic and human applications. This time we will also host a special session for the shared task TermEval, which uses the large manually annotated ACTER (Annotated Corpora for Term Extraction Research) dataset that covers multiple domains and languages.

Terminology has a unique status in language and communication. Theoretically, it is situated within the tension between the flexibility of natural language and the rigidity of artificial sign systems. Reflecting this theoretical status, terms are treated in a specific way in human language practice. In the technical translation pipeline, terms are not “translated” but relevant target language terms are looked up and used, as “mistranslation” can cause grave consequences. Many organisations, including such public institutions as the EU, WIPO and the NLM, and private LSPs, construct and maintain terminologies. Translation quality assurance schemes identify terminology-related issues as one of the focal checking points. Terminologies also provide important resources for education and knowledge transfer.

Although a substantial number of terms are linguistically categorized as so-called multiword expressions, the requirements and desiderata for handling terms as well as their status in language practice pipelines are different from most other multiword expressions such as idioms.

Computational terminology, if it is to make an *in vivo* contribution to the human communication ecosystem, needs to take into account this uniqueness of terminology at every stage of research, from defining problems to be solved and determining methods to be adopted, to developing evaluation schemes to be used.

In the four years since the 5th Workshop on Computational Terminology (Computerm 2016) was held, advancements in distributional representations and deep learning have changed, at least on the surface, the major NLP scene. What about terminology processing? This issue has yet to be fully explored or discussed. For instance, while Neural Machine Translation (NMT) has greatly improved target language fluency, it is sometimes reported that the quality of NMT is on a par with Statistical Machine Translation when it comes to the translation of terminology. Given the unique status of technical terms in communication and language practice, there is much for computational terminology to examine and explore in the face of the recent development of deep learning based NLP technologies, which may not necessarily be in the same line with most NLP tasks.

This workshop thus aims to investigate what deep learning has brought to computational terminology, its impact within human applications, and the new questions that it raises within the scope of terminology. With this in mind, Prof. Dr. Sabine Schulte im Walde (University of Stuttgart) was invited to highlight the new results achieved in modelling noun compound meaning in general and domain-specific language using such statistical methods.

We received 20 submissions, of which 15 are for the general session (9 long papers, 6 short papers) and 5 are for the shared Task TermEval (3 long papers, 2 short papers). We retained 15 papers: 6 long papers for oral presentation (acceptance rate: 30%), of which 4 belong to the general session and 2 to TermEval, and 9 papers for poster presentation (4 long papers and 5 short papers), of which 7 belong to the general session (4 long papers and 3 short papers) and 2 to TermEval (2 short papers).

The 6 long papers retained for oral presentations are the following:

Automatic Term Extraction from Newspaper Corpora: Making the Most of Specificity and Common Features Authors: Patrick Drouin, Jean-Benoît Morel and Marie-Claude L’Homme

TermPortal: A Workbench for Automatic Term Extraction from Icelandic Texts Authors: Steinþór Steingrímsson, Ágústa Þorbergsdóttir, Hjalti Danielsson and Gunnar Thor Ornlósson

Translating Knowledge Representations with Monolingual Word Embeddings: the Case of a Thesaurus on Corporate Non-Financial Reporting Authors: Martín Quesada Zaragoza, Lianet Sepúlveda Torres and Jérôme Basdevant

Which Dependency Parser to Use for Distributional Semantics in a Specialized Domain? Authors: Pauline Brunet, Olivier Ferret and Ludovic Tanguy

TermEval 2020: Shared Task on Automatic Term Extraction Using the Annotated Corpora for Term Extraction Research (ACTER) Dataset Authors: Ayla Rigouts Terryn, Veronique Hoste, Patrick Drouin and Els Lefever

TermEval 2020: TALN-LS2N System for Automatic Term Extraction Authors: Amir Hazem, Mérieme Bouhandi, Florian Boudin and Beatrice Daille

While these workshop proceedings have been published as planned, the workshop itself could not take place due to the current global pandemic. It is currently postponed indefinitely and any updates about this situation will be posted on the workshop website: <https://sites.google.com/view/computerm2020>.

B. Daille, K. Kageura, A. Rigouts Terryn Computerm 2020 organizers

COMPUTERM Workshop Organizers:

Béatrice Daille, LS2N, University of Nantes, France

Kyo Kageura, Library and Information Science Laboratory, The University of Tokyo, Japan

Ayla Rigouts Terryn, LT3 Language and Translation Technology Team, Ghent University, Belgium

TermEval Shared Task Organizers:

Ayla Rigouts Terryn, LT3 Language and Translation Technology Team, Ghent University, Belgium

Véronique Hoste, Ghent University, Belgium

Patrick Drouin, OLST Observatoire de Linguistique Sens-Texte, Université de Montréal, Canada

Els Lefever, LT3 Language and Translation Technology Team, Ghent University, Belgium

Program Committee:

Ehsan Amjadian, University of Ottawa, Canada

Lynne Bowker, University of Ottawa, Canada

Patrick Drouin, University of Montréal, Canada

Natalia Grabar, CNRS UMR 8163 STL, France

Gregory Grefenstette, IHMC, USA

Thierry Hamon, LIMSI-CNRS I& Université Paris 13, France

Rejwanul Haque, Dublin City University, Ireland

Geert Heyman, Nokia Bell Labs, Belgium

Amir Hazem, LS2N, Université de Nantes, France

Olga Kanishcheva, Kharkiv Polytechnic Institute, Ukraine

Koen Kerremans, Vrije Universiteit Brussel, Belgium

Marie-Claude L'Homme, University of Montréal, Canada

Philippe Langlais, RALI, University of Montreal, Canada

Thomas Lavergne, Université Paris-Saclay, CNRS, LIMSI, France

Els Lefever, Ghent University, Belgium

Lieve Macken, Ghent University, Belgium

Veronique Malaise, Elsevier BV, the Netherlands

Ruslan Mitkov, University of Wolverhampton, UK

Emmanuel Morin, LS2N, Université de Nantes, France

Fleur Mougín, University Bordeaux, France

Agnieszka Mykowiecka, IPI PAN, Poland

Rogelio Nazar, University Pompeu Fabra, Spain

Goran Nenadic, University of Manchester, UK

Carlos Periñán-Pascual, Universitat Politècnica de València, Spain

Selja Seppälä, University College Cork, Ireland

Min Song, Yonsei University, Korea

Koichi Takeuchi, Okayama University, Japan

Jorge Vivaldi Palatresi, University Pompeu Fabra, Spain

Leo Wanner, University Pompeu Fabra, Spain

Xiangqing Wei, Nanjing University, China

Pierre Zweigenbaum, LIMSI, France

Table of Contents

<i>Automatic Term Extraction from Newspaper Corpora: Making the Most of Specificity and Common Features</i>	
Patrick Drouin, Jean-Benoît Morel and Marie-Claude L' Homme	1
<i>TermPortal: A Workbench for Automatic Term Extraction from Icelandic Texts</i>	
Steinþór Steingrímsson, Ágústa Þorbergsdóttir, Hjalti Danielsson and Gunnar Thor Ornlósson	8
<i>Translating Knowledge Representations with Monolingual Word Embeddings: the Case of a Thesaurus on Corporate Non-Financial Reporting</i>	
Martín Quesada Zaragoza, Lianet Sepúlveda Torres and Jérôme Basdevant	17
<i>Which Dependency Parser to Use for Distributional Semantics in a Specialized Domain?</i>	
Pauline Brunet, Olivier Ferret and Ludovic Tanguy	26
<i>Leveraging the Inherent Hierarchy of Vacancy Titles for Automated Job Ontology Expansion</i>	
Jeroen Van Haute, Vincent Schelstraete and Mikaël Wornoo	37
<i>Terminology in Written Medical Reports: A Proposal of Text Enrichment to Favour its Comprehension by the Patient</i>	
Rosa Estopà, Alejandra López-Fuentes and Jorge M. Porrás-Garzon	43
<i>A study of semantic projection from single word terms to multi-word terms in the environment domain</i>	
Yizhe WANG, Beatrice Daille and Nabil Hathout	50
<i>The NetViz terminology visualization tool and the use cases in karstology domain modeling</i>	
Senja Pollak, Vid Podpečan, Dragana Miljkovic, Uroš Stepišnik and Špela Vintar	55
<i>Towards Automatic Thesaurus Construction and Enrichment.</i>	
Amir Hazem, Beatrice Daille and Lanza Claudia	62
<i>Supporting terminology extraction with dependency parses</i>	
Malgorzata Marciniak, Piotr Rychlik and Agnieszka Mykowiecka	72
<i>Computational Aspects of Frame-based Meaning Representation in Terminology</i>	
Laura Giacomini and Johannes Schäfer	80
<i>TermEval 2020: Shared Task on Automatic Term Extraction Using the Annotated Corpora for Term Extraction Research (ACTER) Dataset</i>	
Ayla Rigouts Terryn, Veronique Hoste, Patrick Drouin and Els Lefever	85
<i>TermEval 2020: TALN-LS2N System for Automatic Term Extraction</i>	
Amir Hazem, Mérieme Bouhandi, Florian Boudin and Beatrice Daille	95
<i>TermEval 2020: RACAI's automatic term extraction system</i>	
Vasile Pais and Radu Ion	101
<i>TermEval 2020: Using TSR Filtering Method to Improve Automatic Term Extraction</i>	
Antoni Oliver and Mercè Vázquez	106