Computational Terminology in NLP and Translation Studies (ConTeNTS)

Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC)

associated with

The 14th International Conference on
Recent Advances in Natural Language Processing'2023

P R O C E E D I N G S

Edited by:
Amal Haddad Haddad, Ayla Rigouts Terryn and Ruslan Mitkov

Computational Terminology in NLP and Translation Studies (ConTeNTS)
Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC)
Associated with the International Conference
Recent Advances in Natural Language Processing'2023

PROCEEDINGS

7 September, 2023
https://contents2023.kulak.kuleuven.be/
https://comparable.limsi.fr/bucc2023/

Preface

The 1st workshop on Computational Terminology in NLP and Translation Studies (ConTeNTs) was held in Varna, Bulgaria on the 7th of September 2023. This workshop was one of the events co-located with the RANLP 2023 conference (Recent Advances in Natural Language Processing) and incorporated the 16th Workshop on Building and Using Comparable Corpora (BUCC).

Computational Terminology, i.e., research on the automatic collection, management, and analysis of terminology, has attracted the interest of scholars with a diverse range of multidisciplinary backgrounds and motivations. This encompasses a broad spectrum of domains in Natural Language Processing (NLP) such as information retrieval, terminology extraction, question-answering systems, ontology building, machine translation, computer-aided translation, automatic or semi-automatic abstracting, text generation, etc. The field greatly benefits from insights from these different perspectives.

As terms contain a lot of specialised and domain-specific information, they are essential for knowledge mining from texts. Quick evolutions and new developments in specialised domains require efficient and systematic automatic term management. New terms need to be coined and translated to ensure the equitable development of domains in all languages. During the last decade, deep learning and neural methods have become the state of the art for most NLP applications. Those applications were shown to outperform previous methods on various tasks, including automatic term extraction, language mining, assessment of quality in machine translation, accessibility of terminology, etc.

Cross-lingual terminology research is an especially interesting field for both translators and interpreters, who often spend a lot of time and effort on terminology and can benefit from improved tools, and for computational linguists, for whom this is a challenging and interesting field that can offer insights into the latest (neural) techniques. Therefore, it made sense to incorporate the BUCC workshop, which focuses on the use of multilingual comparable corpora (more readily available than parallel corpora), and which hosted a shared task specifically on bilingual term alignment in specialised comparable corpora.

The aim of the workshop ConTeNTS 2023 is to promote new insights into the ongoing and forthcoming developments in computational terminology by bringing together NLP experts, as well as terminologists and translators. By uniting researchers with such diverse profiles, we hope to bridge some of the gaps between these disciplines and inspire a dialogue between various parties, thus paving the way to more artificial intelligence applications based on mutual collaboration between language and technology.

Every submission to the workshop was evaluated by at least two reviewers who were members of the Programme Committee.

The conference contributions were authored by a total of 12 scholars from 8 different countries: Algeria, Argentine, Bangladesh, Belgium, Bulgaria, Chile, Italy, Serbia, Turkey and United States. These figures attest to the international nature of the workshop.

We would like to thank all the colleagues who submitted papers to ConTeNTs 2023 and to BUCC 2023, and who travelled to Varna to attend the event, or presented their work online. We are also grateful to all members of the Programme Committee for providing constructive feedback on each paper. A special thanks goes to Reinhard Rapp, and to the invited Keynote speakers, namely Mo El-Haj from the Lancaster University and Sida I. Wang from Facebook AI Research (FAIR).

September 2023

<div align="right">

Amal Haddad Haddad
Ayla Rigouts Terryn
Ruslan Mitkov

</div>

ConTeNTS:

Organising Committee & Workshop Chairs:

    Amal Haddad Haddad (Universidad de Granada, Spain)
    Ayla Rigouts Terryn (Katholieke Universiteit Leuven (KULAK) Belgium)
    Ruslan Mitkov (Lancaster University, UK)

Programme Committee:

    Sophia Ananiadou (University of Manchester)
    Maria Andreeva Todorova (Bulgarian Academy of Sciences)
    Silvia Bernardini (University of Bologna)
    Melania Cabezas García (Universidad de Granada)
    Esther Castillo Pérez (Universidad de Granada)
    Rute Costa (Universidade Nova de Lisboa)
    Patrick Drouin (Université de Montréal)
    Pamela Faber (Universidad de Granada)
    Mercedes García de Quesada (Universidad de Granada)
    Dagmar Gromann (Centre for Translation Studies – University of Vienna)
    Tran Thi Hong Hanh (L3i Laboratory, University of La Rochelle)
    Rejwanul Haque (National College of Ireland)
    Amir Hazem (Nantes University)
    Milos Jakubicek (Lexical Computing)
    Kyo Kageura (University of Tokyo)
    Barbara Karsch (NYU)
    Dorothy Kenny (Dublin City University)
    Hendrik Kockaert (KU Leuven)
    Philipp Koehn (Johns Hopkins University)
    Maria Kunilovskaya (Saarland University)
    Marie-Claude L'Homme (Université de Montréal)
    Hélène Ledouble (Université de Toulon)
    Pilar León-Araúz (Universidad de Granada)
    Rodolfo Maslias (former Head of TermCoord, European Parliament)
    Silvia Montero Martínez (Universidad de Granada)
    Emmanuel Morin (LS2N-TALN)

Rogelio Nazar (Pontificia Universidad Católica de Valparaíso)
Sandrine Peraldi (University College Dublin)
Silvia Piccini (Italian National Research Council)
Thierry Poibeau (CNRS)
Senja Pollak (Jožef Stefan Institute)
Maria Pozzi Pardo (El Colegio de México)
Tharindu Ranasinghe (Aston University)
Arianne Reimerink (Universidad de Granada)
Andres Repar (Jožef Stefan Institute)
Christophe Roche (Université Savoie Mont-Blanc)
Antonio San Martín Pizarro (Université du Québec à Trois-Rivières)
Beatriz Sánchez Cárdenas (Universidad de Granada)
Vilelmini Sosoni (Ionian University)
Irena Spasic (Cardiff University)
Elena Isabelle Tamba (Romanian Academy, Iași Branch)
Rita Temmerman (Vrije Universiteit Brussel)
Jorge Vivaldi Palatresi (Universitat Pompeu Fabra)

BUCC:

Workshop Chairs:

Reinhard Rapp (Magdeburg-Stendal University of Applied Sciences and University of Mainz, Germany) Pierre Zweigenbaum, (Université Paris-Saclay, CNRS, LISN, Orsay, France) Serge Sharoff (University of Leeds, UK)

Programme Committee:

Ebrahim Ansari (Institute for Advanced Studies in Basic Sciences, Iran)
Thierry Etchegoyhen (Vicomtech, Spain)
Philippe Langlais (Université de Montréal, Canada)
Yves Lepage (Waseda University, Japan)
Shervin Malmasi (Amazon, USA)
Emmanuel Morin (Nantes Université, France)
Dragos Stefan Munteanu (Language Weaver, Inc., USA)
Reinhard Rapp (University of Mainz and Magdeburg-Stendal University of Applied Sciences, Germany)
Nasredine Semmar (CEA LIST, Paris, France)
Serge Sharoff (University of Leeds, UK)
Richard Sproat (OGI School of Science and Technology, USA)
Tim Van de Cruys (KU Leuven, Belgium)
Pierre Zweigenbaum (Université Paris-Saclay, CNRS, LISN, Orsay, France)

Table of Contents