



LREC 2022 Workshop
Language Resources and Evaluation Conference
20-25 June 2022

**Terminology in the 21st century:
many faces, many places
(Term21)**

PROCEEDINGS

Editors:
Rute Costa
Sara Carvalho
Ana Ostroški Anić
Anas Fahad Khan

**Proceedings of the LREC 2022 Workshop on
Terminology in the 21st century:
many faces, many places
(Term21)**

Edited by:

Rute Costa, Sara Carvalho, Ana Ostroški Anić, Anas Fahad Khan

ISBN: 979-10-95546-95-5

EAN: 979109554712

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

Preface

The Term21 (Terminology in the 21st century: many faces, many places) workshop aims to provide a discussion forum regarding the theoretical and methodological approaches that have characterised Terminology in recent years. In particular, it focuses on the connection to the Linguistic Linked (Open) Data paradigm and to Semantic Web technologies through the use of ontologies, as well as on how these approaches promote the creation of interoperable terminological resources in multiple domains and applications, such as Digital Humanities, lexicography, or eHealth. In addition, this workshop addresses current Terminology-related standards and their inherent advantages and challenges.

Following previous initiatives at LREC set up by members of the organising and programme committees (Lisbon, 2004, Genova, 2006, and Portoroz, 2016), Term21 is taking place within the framework of the NexusLinguarum COST Action – European network for Web-centred linguistic data science (CA 18209) and of the MORDigital project – Digitalização do Dicionário da Língua Portuguesa de António de Moraes Silva [PTDC/LLTLIN/6841/2020]. Term21's first edition is held in conjunction with LREC 2022 in Marseille, France, on June 20, 2022.

The workshop accepted eight papers, seven of which have been included in this volume. Four of the papers focus on Automatic Term Extraction (ATE), one on a lexicon-driven approach to Terminology, one on how knowledge organisation can contribute to language simplification, and one on the conversion to the TBX format.

As regards the topic of ATE, the paper by Banerjee et al., entitled *A Dataset for Term Extraction in Hindi*, introduces the first term-annotated dataset for Hindi, based on texts in the field of education, while also addressing the underlying challenges of annotation for under-resourced languages. Nazar and Lindemann, in the paper *Terminology extraction using co-occurrence patterns as predictors of semantic relevance*, propose a method for ATE in which term co-occurrence is used as a statistical measure, contributing to ranking term candidates according to their semantic relevance to a specific domain. This method is applied to a Spanish-English Linguistics corpus. In the paper *Evaluating Pre-Trained Language Models for Focused Terminology Extraction from Swedish Medical Records*, Jerdhaf et al. compare the performance of a generalist Swedish pre-trained language model with a domain-specific Swedish pre-trained model focusing on implant terms. The fourth paper related to ATE is authored by Rigouts Terryn et al. and is entitled *D-Terminer; Online Demo for Monolingual and Bilingual Automatic Term Extraction*. The paper presents an open access, online demo for monolingual and multilingual ATE from parallel corpora, as well as the updated version (1.5) of the Annotated Corpora for Term Extraction Research (ACTER) dataset.

In the paper *Lexicon-driven approach for Terminology: specialized resources on the environment in Brazilian Portuguese*, and following a lexicon-driven approach to terminology work, Arraes describes ongoing collaboration, especially in what concerns content in Brazilian Portuguese, for the development of DiCoEnviro (Dictionnaire Fondamental de Environnement – Fundamental Dictionary on the environment), a multilingual terminological resource developed by the Observatoire de Linguistique Sens Texte at the University of Montreal, Canada. The paper *Knowledge Representation and Language Simplification of Human Rights*, by Silecchia et al., addresses a very recent interdisciplinary project aiming at analysing both the conceptual and linguistic dimensions of human rights terminology, with the goal of developing a new knowledge-based multilingual terminological resource designed to meet the FAIR principles for Open Science. In the future, the authors intend to develop a prototype for the simplified rewriting of international legal texts relating to human rights, in order to facilitate their comprehension by non-experts.

Finally, in the paper *Converting from the Nordic Terminological Record Format to the TBX Format*, Skeppstedt et al. describe work carried out by the Institute for Language and Folklore within the Federated eTranslation TermBank Network Action and focus on the challenges of converting from the Nordic Terminological Record Format, as used in Rikstermbanken (Sweden's National Term Bank), to the TermBase eXchange (TBX) format.

Overall, these contributions address several of the current research topics in Terminology, namely the challenges underlying the role of Natural Language Processing in the automation of terminology-related tasks, the connection between the linguistic and the conceptual dimensions of terminology work, as well as data formats and interoperability. The diversity of domains explored in the papers (e.g. humanities and social sciences, medicine, law, the environment) also illustrates that, indeed, Terminology is increasingly developing in the confluence of many "faces" and "places".

We would like to thank the programme committee for their careful and constructive reviews, which have contributed to the quality of the event. We also would like to acknowledge the support from the MORDigital – Digitalização do Dicionário da Língua Portuguesa de António de Moraes Silva [PTDC/LLT-LIN/6841/2020] project, financed by the Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia.

Rute Costa
Sara Carvalho
Ana Ostroški Anić
Anas Fahad Khan

Organizers

Rute Costa - NOVA FCSH/NOVA CLUNL
Sara Carvalho - Universidade de Aveiro/NOVA CLUNL
Ana Ostroški Anić - Institute for Croatian Language and Linguistics
Anas Fahad Khan - Istituto di Linguistica Computazionale

Program Committee

Ana Ostroški Anić, Institute for Croatian Language and Linguistics (CROATIA)
Anas Fahad Khan, Istituto di Linguistica Computazionale (ITALY)
Bruno Almeida, ROSSIO Infrastructure / NOVA CLUNL (PORTUGAL)
Christian Chiarcos, Goethe University Frankfurt (GERMANY)
Federica Vezzani, Università di Padova (ITALY)
Giorgio Di Nunzio, Università di Padova (ITALY)
Laurent Romary, INRIA (FRANCE)
Maria José Bocorny Finatto, Universidade Federal do Rio Grande do Sul (BRAZIL)
Melania Cabezas García, University of Granada (SPAIN)
Pamela Faber, University of Granada (SPAIN)
Patrick Drouin, Montreal University (CANADA)
Penny Labropoulou, Institute of Language and Speech (GREECE)
Rute Costa, NOVA FCSH / NOVA CLUNL (PORTUGAL)
Sara Carvalho, Universidade de Aveiro / NOVA CLUNL (PORTUGAL)
Silvia Piccini, Istituto di Linguistica Computazionale (ITALY)
Sue Ellen Wright, Kent State University (USA)
Sylvie Desprès, Université Paris 13 (FRANCE)
Thierry Declerck, DFKI GmbH (GERMANY)
Ulrich Heid, University of Hildesheim (GERMANY)

Table of Contents

<i>Lexicon-driven approach for Terminology: specialized resources on the environment in Brazilian Portuguese</i>	
Flávia Lamberti Arraes	1
<i>Knowledge Representation and Language Simplification of Human Rights</i>	
Sara Silecchia, Federica Vezzani and Giorgio Maria Di Nunzio	8
<i>Converting from the Nordic Terminological Record Format to the TBX Format</i>	
Maria Skeppstedt, Marie Mattson, Magnus Ahltop and Rickard Domeij	13
<i>A Dataset for Term Extraction in Hindi</i>	
Shubhanker Banerjee, Bharathi Raja Chakravarthi and John Philip McCrae	19
<i>Terminology extraction using co-occurrence patterns as predictors of semantic relevance</i>	
Rogelio Nazar and David Lindemann	26
<i>Evaluating Pre-Trained Language Models for Focused Terminology Extraction from Swedish Medical Records</i>	
Oskar Jerdhaf, Marina Santini, Peter Lundberg, Tomas Bjerner, Yosef Al-Abasse, Arne Jonsson and Thomas Vakili	30
<i>D-Terminer: Online Demo for Monolingual and Bilingual Automatic Term Extraction</i>	
Ayla Rigouts Terryn, Veronique Hoste and Els Lefever	33

Workshop Program

Monday, June 20, 2022

14:00–14:10 *Opening session*

14:10–14:35 *Linking general and specialized semantic frames in AirFrame*
Ana Ostroški Anić

14:35–15:00 *Lexicon-driven approach for Terminology: specialized resources on the environment in Brazilian Portuguese*
Flávia Lamberti Arraes

15:00–15:25 *Knowledge Representation and Language Simplification of Human Rights*
Sara Silecchia, Federica Vezzani and Giorgio Maria Di Nunzio

15:25–15:50 *Converting from the Nordic Terminological Record Format to the TBX Format*
Maria Skeppstedt, Marie Mattson, Magnus Ahltop and Rickard Domeij

15:50–16:20 *Coffee break*

16:20–16:45 *A Dataset for Term Extraction in Hindi*
Shubhanker Banerjee, Bharathi Raja Chakravarthi and John Philip McCrae

16:45–17:10 *Terminology extraction using co-occurrence patterns as predictors of semantic relevance*
Rogelio Nazar and David Lindemann

17:10–17:35 *Evaluating Pre-Trained Language Models for Focused Terminology Extraction from Swedish Medical Records*
Oskar Jerdhaf, Marina Santini, Peter Lundberg, Tomas Bjerner, Yosef Al-Abasse, Arne Jonsson and Thomas Vakili

17:35–18:00 *D-Terminer: Online Demo for Monolingual and Bilingual Automatic Term Extraction*
Ayla Rigouts Terryn, Veronique Hoste and Els Lefever

18:00–18:10 *Discussion + Closing session*