

TermTrends: Trends in Terminology Generation and Modelling

Patricia Martín-Chozas^{1,*†}, Elena Montiel-Ponsoda^{1,†}, Sara Carvalho^{2,†} and Rute Costa^{3,†}

¹*Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

²*University of Aveiro, Portugal*

³*Universidade NOVA de Lisboa, Portugal*

Abstract

This document presents the objectives, content and organisation of the TermTrends tutorial within EKAW 2022. The tutorial intends to give an overview of current techniques and tools for terminology generation, as well as of standardisation approaches for terminological data. Thus, the first part of the tutorial is a theoretical block that includes an introduction to the terminological work, current standards for terminology modelling and two use cases on legal and medical terminology. The second part is a hands-on block that deals with terminological resources and tools. The tutorial is linked to the conference through a series of topics, such as knowledge acquisition and ontology engineering; and it is suitable for both an expert and non-expert audience.

Keywords

Terminology Generation, Terminology Modelling, Linguistic Linked Data, Semantic Web

1. Introduction

Undeniably, terminology plays an essential role in knowledge representation and organisation, as well as in natural language processing activities dealing with domain-specific knowledge [1] [2]. In its origins, terminology pursued the classification of terms to avoid vagueness and ambiguity, giving birth to initiatives towards the standardisation of terminologies and other language resources [3] [4]. These efforts evolved into ISO standards¹ such as LMF² [5] or TMF³. More recently, terminology work has been increasingly relying on W3C recommendations and standards⁴ like RDF and OWL, with the aim of promoting interoperability of terminological

EKAW-C 2022: Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management, September 26–29, 2022, Bolzano, Italy

*Corresponding author.

†These authors contributed equally.

✉ pmchozas@fi.upm.es (P. Martín-Chozas); emontiel@fi.upm.es (E. Montiel-Ponsoda); sara.carvalho@ua.pt (S. Carvalho); rute.costa@fcsh.unl.pt (R. Costa)

🆔 0000-0002-8922-7521 (P. Martín-Chozas); 0000-0003-3263-3403 (E. Montiel-Ponsoda); 0000-0002-7501-5405 (S. Carvalho); 0000-0002-3452-7228 (R. Costa)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.iso.org/ics/01.020/x/>

²<https://www.lexicalmarkupframework.org/>

³<https://www.iso.org/standard/56063.html>

⁴https://www.w3.org/2001/sw/wiki/Main_Page

resources and automatising the terminological work.

In the TermTrends tutorial⁵, we study the different standardisation approaches, ranging from the initially proposed standards to represent terminology within ISO, to models that represent linguistic data in the Semantic Web, including emerging vocabularies still under development. We also want to give an overview of the main features of the terminology work, exploring its evolution and new methods to speed up the terminology generation process. To complement this, we will present different use cases in which both new methods to generate terminologies and new ways to represent terminological knowledge are applied in specific domains, such as Law or Life Sciences. Additionally, half of the tutorial is planned to be a hands-on session, testing several tools with different purposes, such as the extraction and enrichment of terminologies and their representation as per different vocabularies.

2. Objectives and Content

The objectives of this tutorial can be grouped in three main ideas:

1. **Introducing terminology and terminology work:** its importance for several areas, namely knowledge discovery, knowledge engineering, ontology engineering, and how it has evolved over time, including the study of main terminological resources nowadays, some of which integrating the Linguistic Linked Open Data cloud⁶.
2. **Offering a complete view of the standards for language data in general**, and for terminology specifically, from initial standardisation approaches, going through XML-based formats, such as TBX⁷ [6], to new ISO standards related to terminology extraction and knowledge organisation. We will also take a closer look at those following Semantic Web and Linked Data principles [7], such as SKOS⁸ and Ontolex⁹, and their extensions still under development¹⁰.
3. **Testing different knowledge extraction and management tools** applied to terminology, such as TermitUp¹¹ [8], for terminology enrichment, and OpenRefine¹², Protégé¹³ or VocBench¹⁴ [9], to structure and manage terminologies in Semantic Web formats.

Consequently, the content of the tutorial is divided into two blocks: a theoretical and a practical one.

The theoretical part comprises:

- An introduction to Terminology, which includes a brief history of the terminological work, from its origins to the intersection of terminology and knowledge organisation.

⁵<https://termtrends.linkeddata.es/>

⁶<https://linguistic-lod.org>

⁷<https://www.tbxinfo.net/>

⁸<https://www.w3.org/TR/skos-reference/>

⁹<https://www.w3.org/2016/05/ontolex/>

¹⁰<https://www.w3.org/community/ontolex/wiki/Terminology>

¹¹<https://www.w3.org/2009/08/skos-reference/skos.html>

¹²<https://openrefine.org/>

¹³<https://protege.stanford.edu/>

¹⁴<http://vocbench.uniroma2.it/>

- An introduction to the current and the emerging standards to organise the data contained in terminological resources.
- An introduction to two different use cases: legal terminology and medical terminology. Both presentations include overviews of research projects, resources, repositories, standards, challenges, etc.

The practical part includes the following hands-on sessions:

- Exploration of Semantic Web vocabularies to model terminological data, such as SKOS and Ontolex, to identify challenges and limitations.
- Exploration and querying (SPARQL¹⁵) of terminological resources in RDF.
- Terminology tool testing, mainly: terminological knowledge acquisition (TermitUp) and terminological knowledge organisation and management (WebProtégé and VocBench).

Therefore, we identify three topics of the conference that can be related to the content of the tutorial:

1. **Methods for ontology engineering**, since we propose a review of existing vocabularies, standards and formats to model linguistic data, as well as new proposals specifically intended to model terminological data.
2. **Methods for knowledge acquisition and management**, since the hands-on session is focused on testing several tools for semi-automatic terminological knowledge acquisition, including term extraction, term enrichment and relation acquisition. We also propose the use of different tools for knowledge organisation and management over terminological data.
3. **Applications in specific domains**, during the theoretical session and the practical session we will be presenting several use cases in which different terminological generation and modelling approaches have been applied. These use cases match three different conference subtopics: 1) eGovernment and public administration, 2) Life sciences, health and medicine, and 3) Humanities and Social sciences.

3. Conclusions

This document presents the TermTrends tutorial within EKAW 2022, which focuses on the terminology work from its origins to the present, paying special attention to current techniques for terminology generation and management, and to state-of-the-art models to represent terminological data, specifically in the Semantic Web. The tutorial contains a theoretical and a practical part, and it is targeted at a general audience. With this tutorial we intend to raise awareness about the value of terminological resources specifically and language resources in general within the Semantic Web and Natural Language Processing communities, aiming at creating synergies amongst researchers in both fields.

¹⁵<https://www.w3.org/TR/rdf-sparql-query/>

Acknowledgments

This work is framed within the COST Action (European Cooperation in Science and Technology) through NexusLinguarum, the “European network for Web-centred linguistic data science” COST Action (CA18209)¹⁶.

References

- [1] E. Hovenga, Guideline and knowledge management in a digital world, in: Roadmap to Successful Digital Health Ecosystems, Elsevier, 2022, pp. 239–270.
- [2] K. Stefaniak, Terminology work in the european commission: Ensuring high-quality translation in a multilingual environment, *Quality aspects in institutional translation* 8 (2017) 109.
- [3] M. T. Cabré, *Terminology: Theory, methods, and applications*, volume 1, John Benjamins Publishing, 1999.
- [4] E. Wüster, *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, Romanist. Verlag, 1991.
- [5] G. Francopoulo, M. George, N. Calzolari, M. Monachini, N. Bel, M. Pet, C. Soria, Lexical markup framework (lmf), in: *International Conference on Language Resources and Evaluation-LREC 2006*, 2006.
- [6] A. Melby, Tbx: A terminology exchange format for the translation and localization industry, 201), *Handbook of Terminology* (2015) 393–424.
- [7] C. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: *Semantic services, interoperability and web applications: emerging concepts*, IGI global, 2011, pp. 205–227.
- [8] P. Martín-Chozas, K. Vázquez-Flores, P. Calleja, E. Montiel-Ponsoda, V. Rodríguez-Doncel, Termitup: Generation and enrichment of linked terminologies, *Semantic Web* (2022) 1–20.
- [9] A. Stellato, M. Fiorelli, A. Turbati, T. Lorenzetti, W. Van Gemert, D. Dechandon, C. Laaboudi-Spoiden, A. Gerencsér, A. Waniart, E. Costetchi, et al., Vocbench 3: A collaborative semantic web editor for ontologies, thesauri and lexicons, *Semantic Web* 11 (2020) 855–881.

¹⁶<https://nexuslinguarum.eu/>