languages

MDPI

Editorial

Introducing the Special Issue Terminology in the Digital World

Giorgio Maria Di Nunzio [1,*], Rute Costa [2] and Federica Vezzani [3]

[1] Department of Information Engineering, University of Padova, 35131 Padova, Italy
[2] Linguistics Research Centre, NOVA University Lisbon, 1069-061 Lisbon, Portugal; rute.costa@fcsh.unl.pt
[3] Department of Linguistic and Literary Studies, University of Padova, 35137 Padova, Italy; federica.vezzani@unipd.it
* Correspondence: giorgiomaria.dinunzio@unipd.it; Tel.: +39-049-8277613

This Special Issue is dedicated to the 2nd International Conference on "Multilingual digital terminology today: Design, representation formats, and management systems" (MDTT 2023) (Di Nunzio et al. 2023), which took place in Lisbon, Portugal, from 29 to 30 June, 2023.[1] The original idea behind the organization of this conference on multilingual digital terminology, now in its third edition,[2] stemmed from the desire to give voice to all the interdisciplinary components of this active research community, to create a forum for the discussion and sharing of innovative ideas, and to foster open collaboration among researchers. In this context, this Special Issue serves as a platform for scholars and practitioners across various disciplines, including terminology, terminography, specialized lexicography, applied linguistics, corpus linguistics, and theoretical linguistics, to converge and exchange insights on the landscape of digital terminology resources.

At the heart of this Special Issue lies a collective effort to explore and evaluate novel methodologies in the design, representation, and management of digital terminology resources (Vezzani 2022). Particular attention was given to the interoperability and reusability initiatives of the FAIR approach to digital resources (Wilkinson et al. 2016) and the FAIR terminology paradigm[3] as the first initiative aiming at providing a set of guidelines for the optimal organization of terminological data, metadata, and infrastructure compliant with the FAIR principles (Vezzani 2021; Vezzani and Di Nunzio 2020a, 2020b; Vezzani et al. 2023).

By fostering interdisciplinary dialogue and innovative approaches, this Special Issue supports the advances and understanding of terminology management systems. In this context, the terminologist of the 'digital world' is not only concerned with the conceptual and linguistic representation of terminology (Costa 2013; Santos and Costa 2015), but is also open to dialogue with the structural requirements necessary to preserve and manage a digital resource.

Through different perspectives, we invited contributors to focus on crucial aspects such as assessing the information needs of future users, analyzing structural design approaches for terminological data collection, examining terminological metadata and representation formats, and elucidating methods for validating resource ergonomics.

The four papers appearing in this Special Issue were selected from seven submissions after an invited call for papers from the eighteen contributions to MDTT 2023. These papers were selected according to their adherence to the aims and scope of the MDPI journal *Languages* as well as their quality and their relevance to the topic of Terminology in the Digital World. The heterogeneous research topics covered by this Special Issue range from explanatory combinatorial lexicology and lexicographic definition to corpus linguistics and data visualization, automatic term extraction, and specialized languages.

In the paper entitled 'Toward Non-Taxonomic Structuring of Scientific Notions: The Case of the Language of Chemistry and the Environment', Gotkova et al. (2024) present the problem of structuring scientific notions and challenge the dominance of taxonomies in this field. In particular, the authors propose alternative non-taxonomic structures derived from systematic lexicographic definitions of terminological lexical units. This approach suggests employing 'defined_by' relations rather than 'is_a' relations as a structuring principle,

check for updates

Citation: Di Nunzio, Giorgio Maria, Rute Costa, and Federica Vezzani. 2024. Introducing the Special Issue Terminology in the Digital World. *Languages* 9: 284. https://doi.org/10.3390/languages9090284

Received: 26 June 2024
Accepted: 5 August 2024
Published: 23 August 2024

embedded within a broader model of the general lexicon. The core methodology involves the lexicographic construction of network models of natural language lexicons.

The paper by Reimerink et al. (2024), 'Phraseology and Culture in Terminological Knowledge Bases: The Case of Pollution and Environmental Law', examines how conceptual differences between environmental science and its subdomain, environmental law, are linguistically conveyed, particularly focusing on phraseological combinations. The analysis carried out by the authors centers on verb collocations in environmental law to enrich the phraseology module of EcoLexicon. This study highlights differences in linguistic expressions related to the 'POLLUTION' frame between environmental law and environmental science. It proposes a representation method for this phraseology in a terminological knowledge base, categorizing verbs and their arguments to provide end users, such as translators and technical writers, with essential knowledge for selecting appropriate phraseological options.

The work entitled 'Quartz: A Template for Quantitative Corpus Data Visualization Tools' authored by Isaacs et al. (2024) addresses the issue of interoperability in corpus linguistics software and the consequent difficulty in the development of advanced visualization techniques. The lack of universal standards leads to fragmented tools, necessitating ad hoc solutions for visualization tasks. To tackle this problem, the article introduces Quartz, an open-source visualization tool designed as a template web application, facilitating the development of new visualization features. Quartz aims to encourage the collaborative improvement and sharing of methods. Specifically tailored for the Humanitarian Encyclopedia, it enhances concept analysis tasks by visualizing frequency data and providing concordances for humanitarian concepts.

Finally, the contribution 'Isolating Terminology Layers in Complex Linguistic Environments: A Study about Waste Management' by Cirillo (2024) introduces and experimentally evaluates a novel automatic term extraction (ATE) technique named Domain Concept Relatedness (DCR). Unlike traditional ATE tools that extract terms indiscriminately from specialized corpora, DCR leverages word embeddings to isolate terms related to specific subject fields. It focuses on extracting terms semantically linked to known terms from a glossary or thesaurus, organizing them into Term-Linking Relations (TLRs). This study's main contributions include organizing specialized corpus terminology into TLRs, proposing the DCR technique, and creating focus and contrastive corpora on waste management legislation.

By addressing critical questions, proposing new methodologies, and tackling real-world applications across different domains, this collection of papers exemplifies the collaborative approach necessary for advancing the field. In particular, future research in terminology may address the application of non-taxonomic structures that use alternative ways to organize and relate entities based on different criteria compared to traditional hierarchical organization which classify entities based on a parent–child relationship (e.g., "is a" relationships). The refinement of automatic term extraction (ATE) techniques is also the object of current and future research. As also shown in international activities such as SimpleText,[4] these research activities require a well-defined procedure for testing ATE effectiveness across different subject fields and exploring how innovative machine learning approaches, such as large language models, may enhance accuracy and relevance (Ermakova et al. 2024). Interdisciplinary collaboration should be encouraged to address complex challenges in digital terminology, facilitating dialogue between terminologists, linguists, and technologists to create more robust systems. Furthermore, comprehensive assessments of user information needs and the development of methods to validate and improve the ergonomics of digital terminology resources are critical. Finally, the continued development and implementation of the FAIR principles in terminological data management, along with the exploration of new guidelines and best practices, will enhance the FAIRness of terminological resources.

In conclusion, this Special Issue offers a multifaceted exploration of innovative approaches and challenges in the realm of multilingual digital terminology. From collaborative

data visualization projects to novel automatic term extraction techniques (Di Nunzio et al. 2023), each contribution has significantly added to the discourse surrounding terminology design, representation formats, and management systems in the computer-assisted framework (Pulizzotto et al. 2018).

Acknowledgments: This research has been partially supported by the National Agency for Research—FCT (Foundation for Science and Technology)—within the framework of projects UIDB/LIN/03213/2020 and UIDP/LIN/03213/2020 of the Centro de Linguística da Universidade NOVA de Lisboa (CLUNL) and the Centro Studi in Terminologia Computazionale (CENTRICO) of the University of Padua, and in the research directions of the Italian Common Language Resources and Technology Infrastructure CLARIN-IT.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

1 http://mdtt2023.dei.unipd.it/en/ (accessed on 22 August 2024).
2 http://mdtt2024.dei.unipd.it/en/ (accessed on 22 August 2024).
3 https://purl.org/fairterm (accessed on 22 August 2024).
4 https://simpletext-project.com/ (accessed on 22 August 2024)

References

Cirillo, Nicola. 2024. Isolating Terminology Layers in Complex Linguistic Environments: A Study about Waste Management. *Languages* 9: 68. [CrossRef]

Costa, Rute. 2013. Terminology and Specialised Lexicography: Two complementary domains. *Lexicographica* 29: 29–42. [CrossRef]

Di Nunzio, Giorgio Maria, Rute Costa, and Federica Vezzani. 2023. *Proceedings of the 2nd International Conference on Multilingual Digital Terminology Today (MDTT 2023)*. Edited by Giorgio Maria Di Nunzio, Rute Costa and Federica Vezzani. Lisbon: CEUR, Volume 3427 of CEUR Workshop Proceedings. ISSN 1613-0073. Available online: https://ceur-ws.org/Vol-3427/preface.pdf (accessed on 25 June 2024).

Di Nunzio, Giorgio Maria, Stefano Marchesin, and Gianmaria Silvello. 2023. A systematic review of Automatic Term Extraction: What happened in 2022? *Digital Scholarship in the Humanities* 38: i41–i47. [CrossRef]

Ermakova, Liana, Eric SanJuan, Stéphane Huet, Hosein Azarbonyad, Giorgio Maria Di Nunzio, Federica Vezzani, Jennifer D'Souza, and Jaap Kamps. 2024. Overview of the CLEF 2024 SimpleText track: Improving access to scientific texts for everyone. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024)*. Edited by Lorraine Goeuriot, Philippe Mulhem, Georges Quénot, Didier Schwab, Laure Soulier, Giorgio Maria Di Nunzio, Petra Galuscakova, Alba García Seco de Herrera, Guglielmo Faggioli and Nicola Ferro. Lecture Notes in Computer Science. Cham: Springer.

Gotkova, Tomara, Francesca Ingrosso, Polina Mikhel, and Alain Polguère. 2024. Toward Non-Taxonomic Structuring of Scientific Notions: The Case of the Language of Chemistry and the Environment. *Languages* 9: 95. [CrossRef]

Isaacs, Loryn, Alex Odlum, and Pilar León-Araúz. 2024. Quartz: A Template for Quantitative Corpus Data Visualization Tools. *Languages* 9: 81. [CrossRef]

Pulizzotto, Davide, Jean-François Chartier, Francis Lareau, Jean-Guy Meunier, and Louis Chartrand. 2018. Conceptual Analysis in a computer-assisted framework: Mind in Peirce. *Umanistica Digitale* 2018: 185–205. [CrossRef]

Reimerink, Arianne, Pilar León-Araúz, and Melania Cabezas-García. 2024. Phraseology and Culture in Terminological Knowledge Bases: The Case of Pollution and Environmental Law. *Languages* 9: 84. [CrossRef]

Santos, Claudia, and Rute Costa. 2015. Semasiological and onomasiological knowledge representation: Domain specificity. In *Handbook of Terminology*. Edited by Hendrik J. Kockaert and Frida Steurs. Amsterdam: John Benjamins Publishing Company, vol. 1, pp. 153–79. [CrossRef]

Vezzani, Federica. 2021. La ressource FAIRterm: Entre pratique pédagogique et professionnalisation en traduction spécialisée. *Synergies Italie* 17: 51–64.

Vezzani, Federica. 2022. *Terminologie Numérique: Conception, Représentation et Gestion*. Lausanne: Peter Lang International Academic Publishers. [CrossRef]

Vezzani, Federica, and Giorgio Maria Di Nunzio. 2020a. Methodology for the standardization of terminological resources: Design of TriMED database to support multi-register medical communication. *Terminology* 26: 265–97. [CrossRef]

Vezzani, Federica, and Giorgio Maria Di Nunzio. 2020b. On the Formal Standardization of Terminology Resources: The Case Study of TriMED. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani and et al. Marseille: European Language Resources Association, pp. 4903–10. Available online: https://aclanthology.org/2020.lrec-1.603/ (accessed on 25 June 2024).

Vezzani, Federica, Giorgio Maria Di Nunzio, and Rute Costa. 2023. ISO standards for terminology resources management: Are they FAIR enough? *Digital Translation* 10: 233–52. [CrossRef]

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, and et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. [CrossRef]