

[Click here to be redirected to the virtual room of the Project Expo.](#)

**Project abbreviation:** Text2TCS

**Project name:** Extracting Terminological Concept Systems from Natural Language Text

**Project coordinator:** Ass.-Prof. Dr. Dagmar Gromann, University of Vienna

**Funding:** Pilot project of the European Language Grid (ELG) project that has received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement no. 825627

**Project duration:** 15 July 2020 to 15 September 2021

**Main key words:** term extraction, relation extraction, multilingual information extraction, neural language models

**Background of the research topic:** Text2TCS is firmly rooted in terminology science, information extraction and in its method in computational linguistics. From terminology science, the idea of a Terminological Concept System (TCS) to organize domain-specific knowledge is taken. To extract a TCS from text requires extracting terms and relations, which makes this approach firmly rooted in multilingual information extraction. Finally, in terms of method Text2TCS relied on latest advances in computational linguistics and large multilingual neural language models.

**Goal of the project:** Text2TCS aimed at developing a practical terminological application that is publicly available to everyone in need of organized domain-specific knowledge. Such knowledge implicit in natural language text is explicated by extracting domain-specific terms, grouping them by synonymy to concepts, and extracting conceptual relations, a so-called Terminological Concept System (TCS). A TCS can be extracted from text in multiple natural languages and is provided as inline-annotation, TBX/XML output file, and conceptual graph on the European Language Grid ([ELG](#)).

**Project abstract:** Domain-specific knowledge is paramount to specialized communication settings, from corporate language to translation and crisis communication. To ensure consistency of communication, domain-specific knowledge needs to be organized by grouping terms to concepts and interrelating concepts with hierarchical and non-hierarchical relations, a Terminological Concept System (TCS). While there are many existing tools for term extraction that provide candidate lists, very few consider relations between terms and/or concepts at all. For instance, there are few tools that provide semantic search based on relations, which usually consider general hierarchical relations, such as *narrower* or *broader*. Instead, Extracting Terminological Concept Systems from Natural Language Text (Text2TCS) aimed at extracting terms, grouping them to concepts, and providing hierarchical and non-hierarchical terminological relations from causal and activity to property relations. [Text2TCS](#) is available on the European Language Grid (ELG).

**Publications:**

Lang C, Wachowiak L, Heinisch B, Gromann D. Transforming Term Extraction: Transformer-Based Approaches to Multilingual Term Extraction Across Domains. in Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Online: Association for Computational Linguistics (ACL). 2021. S. 3607-3620 doi.org/10.18653/v1/2021.findings-acl.316

Gromann D, Wachowiak L, Lang C, Heinisch B. Multilingual Extraction of Terminological Concept Systems. 2021. Beitrag in Workshop on Deep Learning and Neural Approaches for Linguistic Data, Nordmazedonien.



**Text2TCS**

[Click here to be redirected to the virtual room of the Project Expo.](#)

Wachowiak L, Lang C, Heinisch B, Gromann D. Towards Learning Terminological Concept Systems from Multilingual Natural Language Text. in 3rd Conference on Language, Data and Knowledge (LDK 2021). 2021 doi.org/10.4230/OASICS.LDK.2021.22

Wachowiak L, Lang C, Heinisch B, Gromann D. CogALex-VI Shared Task: Transrelation - A Robust Multilingual Language Model for Multilingual Relation Identification. in Proceedings of the Workshop on Cognitive Aspects of the Lexicon. Association for Computational Linguistics (ACL). 2020. S. 59–64. 2020.cogalex-1.7