Acquiring Terminological Relations with Neural Models for Multilingual LLOD Resources

Dagmar Gromann, University of Vienna, Austria, dagmar.gromann@univie.ac.at

Abstract

Specialized communication strongly benefits from the availability of structured and consistent domain-specific knowledge in LLOD language resources. Manually curating such language resources is cumbersome and time-intensive. Thus, automated approaches for extracting terms, concepts, and their interrelations are required. Recent advances in computational linguistics have enabled the training of highly multilingual neural language models, such as GPT-3 or XLM-R, that can successfully be adapted to various downstream tasks, from sentiment classification and text completion to information extraction. Furthermore, several approaches exist to extract and explore lexico-semantic relations by means of these language models, however, only few focus on curating, representing, and interchanging domain-specific language resources in the LLOD cloud.

In this talk, we will explore recent work on creative methods to acquire LLOD resources by utilizing pre-trained language models with a particular focus on lexico-semantic and terminological relations. Methods to creatively utilize pre-trained multilingual neural language models, such as GPT-3 or XLM-R, to acquire and extract such relations range from sequence classification to question answering. On the other hand, this talk will discuss how existing LLOD resources can contribute to the task of neural relation acquisition. Any such methods require considerations of scalability and adaptability to domains and languages not considered during training. One objective of the LLOD community is to provide highly multilingual fully interoperable and interlinked language resources. Low-resource languages represent a particular challenge in this regard. Thus, we will further discuss how the proposed methodological combination of neural language models and LLOD technologies can support the acquisition, publication, and interoperability low-resource language data and resources.. In addition, their ease of use for and readiness of uptake by different communities, such as language experts, terminologists, domain experts, and the LLOD community, will strongly influence their final success. The question in this regard is how easy it is for, e.g. linguists, to apply these methods to languages and data of their choice and to evaluate the reliability of the results. To this end, some examples of easy to use technologies to acquire relations will be presented.