Ontology-based Named Entity Recognition in Training Regulations

Research Question:	How can we use ontologies and taxonomies to recognize named entities or spans in German training regulations?
Background:	German Labor Market Ontology (GLMO) (and others) provide many terms of German Labor Market research. Automated recognition occupation names, skills, etc. in training regulations using Named Entity Recognition (NER) Tasks: • Systematic literature review on NER based on terms of large ontologies • Research of state of the art NER methods • Identification of key challenges in ontology-based NER (e.g. huge number of terms in ontology) • If feasible, train own NER model on the corpus of German training regulations **Note**: These texts are only available in German, so to train machine learning models for NER and label data on them, skills in reading German language are highly recommended
Data:	*German texts!* • German Labor Market Ontology (GLMO) • Extended Computer Science Ontology (CSO) • (C)VET Regulations
Literature:	T. Reiser, J. Dörpinghaus, P. Steiner, M. Tiemann. Towards a datatset of digitalized historical German VET and CVET regulations. *Data*, **9**(11):128 (2024). T. Reiser, J. Dörpinghaus, P. Steiner. Analyzing Historical Legal Textcorpora: German VET and CVET regulations. *INFORMATIK 2024*, *Lecture Notes in Informatics (LNI)*, Gesellschaft für Informatik, p. 2001-2012. J. Dörpinghaus, J. Binnewitt, D. Samray, K. Hein. Understanding Informatics in Continuing Vocational Education and Training Data in Germany. *ACM Transactions on Computing Education*, 2024, 24(3), 1-22. J. Dörpinghaus, D. Samray, R. Helmrich. Challenges of Automated Identification of Access to Education and Training in Germany. *Information* 2023, *14*(10), 524
Requirements:	Coding (Python); Databases (SQL); Knowledge of the German Labor Market
Contact:	Dr. Jens Dörpinghaus (doerpinghaus@uni-koblenz.de)