





Clustering of labor market data and trend flow

Research Question:	How can we track the flow of topics (e.g., sustainability or artificial intelligence) in different clusters of labor market data?
Background:	It is not yet clear how we can reasonably cluster labor market data. Natural categories are occupations (KldB) or named entities that are recognized. However, exploring different clustering and cluster labeling strategies could help to track which trends flow, remain, or spread through these clusters.
Data:	<i>German texts!</i> Annotated Labor Market Data, e.g., <ul style="list-style-type: none">• Twitter data (2017-2023)• YouTube metadata (2027-2024)• (C)VET Regulations
Literature:	 T. Reiser, J. Dörpinghaus, P. Steiner, M. Tiemann. Towards a dataset of digitalized historical German VET and CVET regulations. <i>Data</i> , 9 (11):128 (2024).  T. Reiser, J. Dörpinghaus, P. Steiner. Analyzing Historical Legal Textcorpora: German VET and CVET regulations. <i>INFORMATIK 2024, Lecture Notes in Informatics (LNI)</i> , 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. <i>ACM Transactions on Computing Education</i> , 2024, 24 (3), 1-22.  J. Dörpinghaus, D. Samray, R. Helmrich. Challenges of Automated Identification of Access to Education and Training in Germany. <i>Information</i> 2023, 14 (10), 524
Requirements:	Coding (Python); Databases (SQL); Knowledge of the German Labor Market
Contact:	Dr. Jens Dörpinghaus (doerpinghaus@uni-koblenz.de)
