

Workshop: NLP and Computational Linguistics in Economics

Benjamin W. Arold, ETH Zurich

Location: Louis A. Simpson International Building, Rm 271, Princeton University

Time: March 2024

Syllabus Summary:

A vast portion of human knowledge resides in unstructured formats, particularly within written texts. This workshop provides an overview of recent methods from Natural Language Processing (NLP) and Computational Linguistics to process and analyze text data. We also explore how information extracted from text can be integrated into economic analysis. Topics covered include dictionary-based methods, tokenization, measuring document distance, machine learning with text, word and sentence embeddings, linguistic parsing, as well as transformers and large-language models. The workshop concludes with the presentation of a current research project that employs several of these methods to extract information from the content of collective bargaining agreements (Benjamin W. Arold, Elliott Ash, W. Bentley MacLeod, Suresh Naidu, mimeo: “When Words Matter: The Value of Worker Rights in Collective Bargaining Agreements”).