UNIVERSITÄT ZU KÖLN

Introduction to Python for Social Sciences Featuring Text Analysis Examples

Christopher Klamm, M.Sc., M.A.

Modul: SM Applied Theories and Methods Political Science II
ID: 14335.0202
Duration: WS 2025/2026 ***(Tuesday 14:00-17:30)***
Target Audience: Social science students with no programming experience
E-Mail: klamm@wiso.uni-koeln.de
Office Hours: online appointment via link on Ilias

Course Description

Python is a prominent programming language in machine learning and is increasingly relevant for social sciences. It allows researchers to utilize state-of-the-art techniques and analyze extensive corpora to study phenomena and constructs at a large scale. This hands-on course introduces social science researchers to Python programming through a practical focus on text analysis. Students will learn to process, analyze, and visualize textual data from sources such as parliamentary speeches, news articles, books and other textual resources. The course emphasizes practical applications relevant to social science research, including sentiment analysis, word frequency analysis, and basic text analysis techniques with Python.
Learning Outcomes

Upon completion, students will be able to:

- Write your first code with Python
- Write Python programs to read, process, and analyze text files
- Structure code using functions and dataclasses for better organization
- Perform text cleaning and preprocessing using NLTK
- Perform basic functions like word frequency analysis on real-world datasets.
- Create visualizations to communicate text analysis findings in Python
- Apply computational methods with Python to their own research questions

Course Schedule

Week 1: Python Fundamentals and the (group) projects

- **Session 1** (1.5h): Introduction to Python (“hello world”) and Google Colab setup, variables, strings, lists, basic operations
- **Session 2** (1,5h): Control structures (if/for), functions, string manipulation, building a word counter

Week 2: External Libraries, File handling and Text Analysis Basics with Python

- **Session 3** (1.5h): File handling, dictionaries, text cleaning (punctuation, case), analyzing interview transcripts
- **Session 4** (1.5h): Introduction to pandas, datasets (HuggingFace), regular expressions to extract patterns (emails, hashtags, dates)

Week 3: Text Analysis with NLTK/SpaCy

- **Session 5** (1.5h): NLTK basics, tokenization, stop words removal, word frequency distributions
- **Session 6** (1.5h): Sentiment analysis with TextBlob, polarity and subjectivity analysis, analyzing social media data

Week 4: Project documentation and key phrase analysis (online)

- **Session 7** (1.5h): Introduction to project documentation and overview of standards for reporting a data science project as part of the final submission.
- **Session 8** (1.5h): N-grams and collocations, finding key phrases, basic topic identification

Week 5: Python becoming structured with dataclasses, and functions

- **Session 9:** Presenting project ideas and opening the floor for questions with testing a small task of the project in the session together.
- **Session 10:** Structured programming with functions, introduction to dataclasses, building a text analysis toolkit.

Week 6: Visualization and Application

- **Session 11** (1.5h): Creating word clouds, frequency charts, and time-series visualizations with matplotlib/altair
- **Sessions 12** (1.5h): Outlook to advanced topics preview

Week 7: Project presentations

- **Session 13** (1.5h): Project presentations for feedback
- **Session 14** (1.5h): Q&A and next steps for the final project

Assessment

- **Final Project** (100%): Analysis of a student-chosen text corpus presented as a project report, focused on a provided dataset and your own research question.

Required Software

- Google Colab account (free)
- Python libraries: NLTK, pandas, textblob, matplotlib/altair, wordcloud
- Hardware: Laptop

Prerequisites

Basic computer literacy and interest in text-based research methods. No programming experience required.

Literature:

- Bird, Steven/ Klein, Ewan/ Loper, Edward (2009): Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly and Associates.
- Hovy, Dirk (2021): Text Analysis in Python for Social Scientists: Discovery and Exploration. Cambridge University Press.
- Grimmer, Justin/ Roberts, Margaret/ Stewart, Brandon (2022): Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press.

For updates and news about Python, a great resource channel:
https://www.youtube.com/@ArjanCodes/videos