

AHRC/Exeter Summer School Course Outline - Text Analysis Using Python

Day 1: Introduction to Python and its utility in digital humanities research. Learn how to use Google Colab for analysis, set up Python with Google Colab, and analyze data frames using Pandas.

Day 2: Pandas, NER, Regex. Dive into the world of NLP libraries and tools, and learn how to manipulate text using regular expressions, lemmatization, stemming, and named entity recognition.

Day 3: Text Representation. Explore techniques for representing text, such as TF-IDF, Bag-of-Words, and Count-Matrix, and learn how to find structure in text data using pointwise mutual information, n-grams, and cosine similarity.

Day 4: Sentiment Analysis and Visualization. Discover how to perform sentiment analysis using lexicon-based techniques (AFINN, Bing, NRC) and visualize your results using Matplotlib, and Plotly.

Day 5: Language Models. Learn about word2vec, doc2vec, topic modeling, and LDA.

This course is designed to give you hands-on experience with the latest tools and techniques in text analysis and will equip you with the skills you need to make sense of text data and extract valuable insights.