

Data Fellowship Project 2022

UKDS - Natural Language Processing with Qualitative Data

Organisation and Team

UK Data Service

Selection Method

Virtual interview

Application Criteria

Essential:

- Experience with quantitative and qualitative data
- Desire to learn data analysis tooling
- Interest in learning programming languages (Python, R, or similar)
- Familiarity with data visualisation (experience with creating simple graphs)
- Interest in data science techniques – cleaning and manipulating data

Desirable:

- Interest in learning about Natural Language Processing
- Experience with data analysis tooling

Project Outline

The project will involve the application of NLP to analyse the UKDS dataset on ‘Health and Social Consequences of the Foot and Mouth Epidemic’ (2001-2003).

The candidate will be tasked with implementing NLP techniques using a programming language.

By making use of the diary entries, there will be an opportunity to implement sentiment analysis, word segmentation, stemming, lemmatization, and other NLP methods.

Tasks will include quantitative descriptions, NLP (qualitative) analysis and visualisations of data set. If time permits, candidate will be taught to scrape tweets on Corona-19 as a comparison case.

Data Fellowship Project 2022

Practical Considerations

As this project is somewhat open ended, we would expect the candidate to work full time (flexibly) for the entire allowed duration of the project.

The time needed may be adjusted based on programming skills or experience of the candidate.

Supporting Information

Info on the dataset:

<https://ukdataservice.ac.uk/app/uploads/guidehealthandsocialconsequences.pdf>

Info on NLP packages:

<https://spacy.io>

<https://www.nltk.org/>

Support and Training

Initial training focuses on the necessary concepts and tools for the project e.g. what NLP is, how to use notebook software, how to configure computational environments, etc.

The candidate will then apply the concepts and tools to the UKDS data set.

After this, the candidate will be encouraged to:

- identify a research question using the data set and NLP,
- explore NLP libraries, packages and processes,
- create a report on the data set, the research question, the method and the results.

The candidate will meet frequently (~2-3 X weekly) for training, guidance, feedback and advice.

The candidate will also need to produce a blog post for us about their experiences of learning NLP in the context of an original research project.