

Proposal Draft

Hanyi Jiang

April,03,2023

- Data

Fine-scale data: Title, abstract, money amount, lead institution/person

Broad data: What each research does?

Especially UK data, we can download all data from UK research and innovation using the API.

May need to use HPC. (But I am trying to focus on one or two countries, so... may not need it)

- Method

Mallet in bash (Machine learning for language Toolkit) (PPT has the link of the introduction for it)

The modeling type is unsupervised. We have to decide how many topics we need. (LDA)

If I meet some problem with natural language procession, I can also read the link from ppt. (realpython.com) (Initially use Python's NLTK and Gensim package.)

For this project, if possible, I would like to use the UK data to analyze how the focus on different areas in the UK has changed over time. For example, in my home country China, 10 years ago, the infrastructure and real estate popularity brought civil engineering-related majors a popular profession. However, with the country's development model shift, the social demand for civil engineering graduates has shrunk. Then, around 2018, when I applied to universities to decide on a major, computer science majors became the new hot direction due to the booming development of domestic Internet enterprises, which led to a computer science boom. I think maybe it's a similar situation in the UK. As particular policies or economies develop, different areas receive extra attention at different times, which will be reflected in the amount of funding received.