Neo4j, Natrual Language Processing and Open AI at The Ditchley Foundation

Joseph Cronshaw | School of Social Science

Overview of the Data Fellowship

At my Data Fellowship with the Ditchley Foundation I worked on three separate projects in three separate teams. The first project was researching people given to us who were important to the Ditchley Foundation and adding them to the database. This was done in a team but was largely independent work unless someone got stuck. The second was finding important organizations and people within the education sector for Ditchley's partnership with the Lumina foundation which focused on educational reform in the UK and US. This the most team centred project, our team looked specifically at AI in education and how the curriculum needs to adapt. Finally, my favourite project was the quantitative work we did, it began with learning the required software, Neo4j, and then we were given free rein to find or do something interesting with the foundations graph database. We were in a team, but again it was mostly for troubleshooting.

Data Analysis

In my Quantitative strand of work, I began by learning Cypher the coding language of Neo4j. We then went on to work with mock datasets, as seen below, where we learnt NeoDash and Bloom tools used to present findings in graph databases more clearly.

A NeoDash dashboard I created while learning the software.

After that my team were given the task of analysing the recommendation nodes within Ditchley's actual graph database. This involved more complex coding and culminated to another dashboard. Finally, we were given free rein to research what we like so I used Open AI and Natural Language Processing to generate new subthemes for the database to more accurately categorize people, as many people's subthemes were defined as "other."

A picture of all the interns at Ditchley for the celebration event

Findings

In my group project where we researched the recommendation nodes in the Ditchley database we had three key findings. Ditchley events with a higher proportion of recommended individuals tend to be more successful. Recommendations within certain sectors were more reliable than others. Finally, cross gender recommendations, i.e. a man recommending a woman or vica versa were the most prevalent in successful events.

My Natural Language Processing project aimed to create new subthemes to better categorize individuals withing the graph database. I achieved this by first using natural language processing and Open AI and found strong evidence that Public Policy and International Relations should be subthemes and that Agriculture, Political figure/association, Women's rights, Drugs/Narcotics and addiction and AI could also be useful additions.

Key Skills Learnt

Throughout my time at Ditchley I did just under 20 presentations which culminated with me presenting my work in person at a Ditchley event. For me presentation is the best skills I have developed, I'm by no means an excellent public speaker now but I'm now comfortable and confident presenting. There was of course lots of teamwork and time management involved and sharpening those skills is always good.

The more technical skills I developed were centred around coding, I used Python and Cypher throughout my projects and of course learnt a lot about graph databases (Neo4j). Unexpectedly I learnt a lot about Open AI's features and a lot about Natural language processors ehich proved to be fascinating and will likely feature in my dissertation. We also did a lot of open-source research which is an invaluable skill for university.