

Google Scholars Data Visualisation

Adi Bozzhanov

November 1, 2022

1 Introduction

The aim of this document is to brainstorm about my Final Year Project topic and try to structure my initial thoughts about it. The vague initial problem statement is as follows:

”Parse data from google scholars and visualise it, potentially as a network of articles”.

As a starting point, I will try to answer following questions before going into implementation and planning details:

- What kind of data am I working with?
- What are potential ways of visualising this data?
- What is the potential development Stack?
- What do I expect the final result to look like?

2 What am I working with?

This section is to just summarise what kind of information I can even get from google scholar. The google scholar landing page is just a simple search bar where one can enter any prompt and get a list of relevant scholarly articles as a result. Each article has the following information that will be useful for network visualisation.

- Authors [Article - Author]
- Cited by [Article - Article]
- Refers to [Article - Article]
- Number of citations

I believe this is everything that could be useful???

3 What can I visualise?

I could come up with those few useful ways to visualise the following data.

- Authors network [nodes: Authors, edges: articles, undirected graph]
- Citations networks [nodes: Articles, edges: citations/references, directed graph]

4 Potential Stack

Considering that there already is a very nice scholarly python library for accessing the google scholar in a pretty easy fashion, I would like to implement the majority of the project in python. Since it very likely will be a web application, Django + React could be just the right combination for the job.

5 Expected result

So far the plan is fairly simple: Make a web application with a python back-end which will provide the following features at the very least.

- Authors network visualisation and traversal
- Publications network visualisation and traversal
- Provide multiple ways of searching for a publication and rendering it's network (Integrating default google scholar search results generation)
- Data filtering, sorting and basically search configuration tools.

The plan from this point would be to convert these thoughts into a technical specification, mock ups, the architecture plan, and come up with a development road map for the rest of the year.