

CS410 project proposal

Topic: Chrome extension for Google Scholar

Project Team member:

Rongrong Lin, NetID: rlin15

Goals:

This tool benefits a researcher/scholar who has a google scholar page.

On a Google scholar user/scholar page, there is information about the scholar's list of publications, the citation number for each publication and etc. Based on these text contents, with the implementation of a chrome extension, I will extract further information such as:

1. From the articles that cite a publication, find the most relevant articles to the publication, make it a recommended reading list to the scholar.
This uses the search engine technique and ranker function (such as BM25) in text retrieval.
This functionality is useful for scholars who wants to pick some others' work that's related to his/her own research for reading and studying.
2. Make a list of the authors who cited your publications, with their Google Scholar links (for their contact info and research interests).
This uses the technique of web scraping from HTML texts.
This functionality is useful to help you with building connections with other scholars who share the same research interest with you and value your work (by citing them). For example, sometimes you may need recommendation letters from people who know your work but are not a colleague of you.
3. The Google Scholar page provides the citation count for each of your publication, but regardless of whether it's a self-citation (cited by yourself and your coauthors) or an independent citation. I will add a function to count the independent citations.
This uses the technique of web scraping from HTML texts.
This functionality is useful for scholars when they need a detailed analysis of impacts of their work.

Algorithms and techniques to be used:

Text retrieval, search engine, ranking function (BM25), web crawler, HTML scraping

Programing languages to be used:

Python, JavaScript

Deliverables:

- Source codes of the chrome extension
- A demo showing how the tool works and how to use the tool

Workload estimation:

task	Hours needed
Chrome extension interface design, make .css, .html, .js files	6hrs
Add functionality of citation counting	5hrs
Add functionality of relevant reading, with Metapy toolkit from Python	7hrs
Add functionality of making author list	8hrs