

CS410 Project Proposal

Group name: AlphaFold3

Group captain: Yang Duan (yangd4)

Group member: Yang Duan (yangd4), Chongyu Wang (chongyu4), Haoyu Su (haoyus2), Yulei Li (yuleili2), Jiaye Wang (jiayew3)

Chosen topic: Intelligent Browsing

Problem Introduction

The theme of our project is Google Chrome Extension for paper suggestions based on key-phrase from all mentioned papers in lectures of CS410, which is an educational chrome extension to help students in the CS410 course. When browsing the lecture slides, students might have problems with the conception in slides and want to learn more about them. To help students better understand the content in lecture slides, we design a google chrome extension, which shows the conceptions of terms in Wikipedia when users highlight the terms. Students are also provided with research paper recommendations from the recommendation system from the backend program. These recommendations are based on the relevance of the highlighted words or phrases when students browse web pages, and they are chosen from the offline database containing all research papers mentioned at the end of each lecture.

Method and Programming Languages

These algorithms introduced in the lecture slides are options for us to use. To find the most related papers and Wikipedia pages with given queries, we choose different methods, including BM25, PLSA, etc. We find the related topics or terms of each paper and Wikipedia page, then match them with user's highlight content based on the topics or terms we extract. Finally, we will evaluate the performance of our method using nDCG@5 evaluation.

We plan to use several programming languages, including Python, Java, HTML, CSS and Javascript.

Data and Expected Result

We plan to create the dataset of all research papers mentioned at the end of each lecture.

The input is the highlighted words or phrases when students browse web pages. The output of the Chrome Extension will show the relevant explanation from Wikipedia or other useful sources and the recommended research paper mentioned in lectures.

Workload Justification

We have the total work of at least $20 \times 5 = 100$ hours. Below, we list the major tasks completed and the expected amount of time required:

- Collect all research papers mentioned at the end of each lecture as the main concept of our database. Approximately 5 hours.
- Process the downloaded files of the papers into TXT files. Approximately 5 hours.
- Train the chosen algorithm. Approximately 20 hours.
- Evaluate the trained model on the test dataset. Approximately 15 hours.

- Create the database of inverted indexes. Approximately 20 hours.
- Implement the backend system. Approximately 40 hours.
- Implement the Chrome Extension. Approximately 30 hours.
- Fix miscellaneous errors and bugs. Approximately 20 hours.

In total, we plan to spend approximately $5 + 5 + 20 + 15 + 20 + 40 + 30 + 20 = 155$ hours