CS410 Final Project Proposal

Annotation Extension For Educational Websites

About

- Team Name: Night Crawler
- Members:
  Jiawei Yuan(jiaweiy3) - captain
  Ziqi Xu(ziqixu3)
  Ziyao Zhang(zhang416)

Overview

Our project falls under the intelligent browsing theme listed in the project guide. We aim to build a chrome extension that enables students to link Q&A on Campuswire discussion forum to coursera lecture video contents. More specifically, by using the extension, students will be able to select any keyword on Coursera web pages (lecture video transcript, homework description, etc) and easily retrieve a collection of related discussions from Campuswire. The tool will integrate learning materials from both websites in a more organized way and improve the efficiency of learning, as students no longer need to log into Campuswire and switch back and forth to manually search for Q&A regarding a topic on Coursera. They can simply highlight keywords, and bookmark / annotate those that are useful from a list of related discussion topics being fetched.

The topic is also related to the course because for the development of the tool we will be leveraging the use of text retrieval and document ranking techniques, as well as tuning the model to achieve optimal accuracy. We will also scrape and index the web pages through an implementation of an inverted index.

Techniques

To make our course experience more intelligent, we use posts on Campuswire and transcripts on Coursera as the datasets. With techniques like text retrieval and chrome extension APIs, we can implement an extension to improve efficiency. The main algorithms include web search-related document ranking, BM25, and web crawlers. We plan to use Python for data processing and JavaScript for our extension frontend and logic.

Validation

To make sure our approach will work, we will conduct user research with users who will be using this tool. We will find several students to check if the relevant result we provide matches their expectations. If they find the information not relevant, they will be

requested to select the keywords they're looking for and the corresponding data will be updated to ensure better results in the future.

Estimate of work

Required: 60+ hours
Estimate: 82 hours

- Scripting and scraping (15 hours)

- Data processing (35 hours)
    - Preprocessing: tokenization, document collection, build inverted index (10 hours)
    - Classification: categorize the questions on Campuswire according to tags and topic of content (15 hours)
    - Develop and test scoring function (10 hours)

- User Interface (12 hours)
    - Extension Popup Interface (6 hours)
    - Annotation System Interface (6 hours)

- Validation + modification (10 hours)
    - Solicit volunteers (2 hours)
    - Collect User Feedbacks (3 hours)
    - Evaluate and Improve Models (5 hours)

- Progress Report (5 hours)

- Final Code with Documentation and Presentation (5 hours)
    - Documentation and Tutorials (3 hours)
    - Presentation (2 hours)