Project Proposal - Developer Team

1.Team Member Information

- Our group consists of 4 people: Allan (shengqi) Huang (Shengqi5), Zengjie Tang (zengjie3), Danmeng Zheng(danmeng2), Chuching Ho (cch11)
- Allan Huang is the captain

2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

- For our project, we have chosen theme 1: Intelligent Browsing. We want to develop a browser extension for campuswire that retrieves the campuswire posts relevant to the query/keyword.
- It's a problem because the search function in campuswire only returns posts that match the keyword. The search results are not ranked by their relevance to the query. The results also don't have other information like number of likes or exact date of the posts, which makes it difficult for the users to judge the usability of the posts.
- Our browser extension helps with information retrieval and can make the campuswire function more intelligent and user-friendly. We are also going to implement the retrieval functions covered in course to recommend most useful posts to users.

3. Briefly describe any datasets, algorithms or techniques you plan to use:

- Our aim is to use Campuswire post as a dataset and develop a chrome extension through chrome extension API. We plan to integrate text retrieval techniques, primarily focusing on the algorithms BM25 and TF-IDF. The main objective is to enable the extension to display the top five most popular posts for a specific keyword input by the user.

4. How will you demonstrate that your approach will work as expected? :

- To validate our work, other than testing the extension ourselves, we will also use our personal connections to find friends who are also students of UIUC and also use Campuswire. And we will combine the feedback from them and make changes accordingly.

5. Which programming language do you plan to use?

- We are planning to use JavaScript, HTML and CSS and Python for the Chrome extension user interface and logic and Python for implementing the crawler logic and data processing.

6. Workload

- Since we have 4 members, the estimated completion time will roughly be 20 * 4 hours, we have divided the workload:
    - Understanding / Scraping Data (20 hours)

- Backend development (30 hours)
- Validating and Testing (20 hours)
- Progress Report, Create Final Presentation (10 hours)