# Personalized Bookmark Search Engine

**CS 410 Text Information Systems (Fall 2021)**
**Prof. ChengXiang Zhai**

**Gazi Muhammad Samiul Hoque*** **Yuheng Xie†** **Grace, Mu-Hui Yu‡** **Ying-Chen Lee§**

## 1 Introduction

Most, if not all modern internet browsers such as Google Chrome offer a bookmark feature that allows users to retain website URLs for future reference. This makes it very convenient for users to access their favourite websites. However, as a bookmark list grows larger over time, users will have a hard time searching through it to find the relevant contents they want. Moreover, there isn't a way to search into the contents of those bookmarks as a bookmark only contains the URL and title of a particular web page. Thus, we are proposing a personalized bookmark search engine to address these shortcomings.

## 2 Our Project

For easy installation, we will develop the personalized bookmark search engine as a Google Chrome browser extension. When the user bookmarks a website using our extension, it will send the current web page address to our back-end server. There it will parse and index the page for the user. A search box will also be implemented in the widget that allows the user to enter a query term and it will fetch relevant contents from the index through our back-end API.

The theme of our project is **Intelligent Browsing**, where we implement the "bookmark-and-search" function on top of Google Chrome to allow users to browse intelligently through the web. We'll apply the web crawling, indexing, and searching techniques we have learnt in this class for this project.

## 3 System

**Algorithms** BM25 will be used as the search algorithm.

**Programming Language** On the front-end, Vue.js, Javascript, HTML and CSS will be used to create the Google Chrome extension and the web application to display search results. Python and Flask will be used to create the back-end application which serves the API for search, indexing and functions to retrieve user information.

**Primary Tasks and Workload** Below is the task breakdown for this project:

**Page Crawler & Indexing** Implementing a page crawler to parse the current web page. $\approx$ **10 hours**

**Manage Bookmark** Implementing functions for indexing of bookmarks and their contents in the server. $\approx$ **10 hours**

*ghoque2@illinois.edu
†yuhengx2@illinois.edu
‡muhuiyu2@illinois.edu
§yclee6@illinois.edu

**Searching** Implementing the BM25 search function and other related utility functions. ≈ **10 hours**

**User Management** Implementing functions to associate bookmarks and their contents with the current user. ≈ **15 hours**

**Back-end** Developing a back-end web application which serves the above functionalities on separate routes for the front-end application to call. ≈ **25 hours**

**Front-end** Developing a "pop-up" front-end web application for the extension that displays the search results and other relevant contents. ≈ **30 hours**

**Integration & Deployment** Integrating all the front-end components together as an install-able Google Chrome extension, and integrate the server app that can be easily deploy-able to any server. ≈ **20 hours**

**Testing** Testing and fixing of any potential bugs ≈ **10 hours**

Total time required: ≈ **130 hours** (tentative). The workload will be evenly spread among all team members.

# 4 Demonstration

We'll first demonstrate that our approach works by showcasing that the bookmarking and search functions are working as intended from a user's perspective. To demonstrate that our search algorithm is reasonably accurate, we'll compare its accuracy with other systems on some predefined queries. Finally, we can also perform user testing by asking our friends and relatives to test out the extension and gathering their feedback for evaluation afterwards.

# 5 Team Information

Our group name is **PBSE**, and we are a four-member team. Our details are as follows:

- Gazi Muhammad Samiul Hoque (NetID: `ghoque2`) - Captain
- Yuheng Xie (NetID: `yuhengx2`)
- Grace, Mu-Hui Yu (NetID: `muhuiyu2`)
- Ying-Chen Lee (NetID: `yclee6`)

# 6 Future of this Application

A possible future extension to this project is to create a user account management service that allows users to sync and persist their bookmarks across multiple devices and browsers. After logging into the service, a user can search over all his bookmarks and retrieve his desired contents anytime, anywhere.

# 7 Conclusion

Our project supports intelligent browsing by allowing users to perform search on their bookmarked contents, which increases their productivity when browsing the web.