National University of Singapore
Institute of Systems Science

ISY5001
Lemon Academic Knowledge Platform–User Guidance

Group 10

Bian Weizhen A0285814W

1 Setup Instructions

Before you begin, ensure you meet the following requirements:

Prerequisites:

- Operating System: macOS (recommended).
- IDE: IntelliJ IDEA (recommended as it auto-identifies and installs project dependencies).
- Java: JDK 11.

Steps:

1. Install IntelliJ IDEA: Download and install from the official website.
2. Install Java JDK 11: Download JDK 11 from the Oracle official site and follow the on-screen installation instructions.
3. Clone the Repository:

```
git clone https://github.com/LuckyBian/ISY5001.git
```

After completing the above steps, your application should be up and running. If you encounter any problems, please double-check that all prerequisites are met.

2 Data Collection

For the functionality of the academic search engine, there are two primary ways to acquire the data:

- Using Pre-collected Data
- Web Scraping for New Data

Using Pre-collected Data

For those who wish to use the pre-collected dataset, here's what's included:

- `collectedData.csv`: It contains information on 1,000 academic articles, encompassing title, link, year, abstract, keywords, and more.
- `data_table.ser`: Stores the mapping between core terms and associated article information.
- **User Interactions:** `userActivity.xlsx` Records user click activities, `likeActivity.xlsx` Logs liked articles by the users, `starActivity.xlsx` Registers user's favorite or starred articles.

Web Scraping for New Data

If you want to collect more data or collect a new dataset, you can find `DataScraper.java` and adjust the parameters. For data crawling algorithms and a detailed explanation of parameters, please refer to `REPORT`. Note: The NUS campus network is unable to access some of the web pages and the data collection fails. Please connect to other networks such as cell phone hotspots for collection.

3 Running Application

If data collection is complete, ensure `data_table.ser` and `collectedData.csv` exist. Then, you can launch the application:

1. Find `SearchEngineApplication.java` in IntelliJ IDEA and click run.
2. Go to the following website:

```
http://localhost:8080/
```

3.0.1 Accessing the Application

My application is divided into visitor mode (not logged in) and user mode (logged in), which I'll describe next in relation to the interface for each.

Home Page:

Once you have successfully launched the app and accessed the web page, you will see the Home Page. As shown in the image below, you can enter keywords for a literature search, login, or register for an account. In addition, the platform will recommend some of the latest and hottest academic articles based on the FEATURES of the article (available even without login).

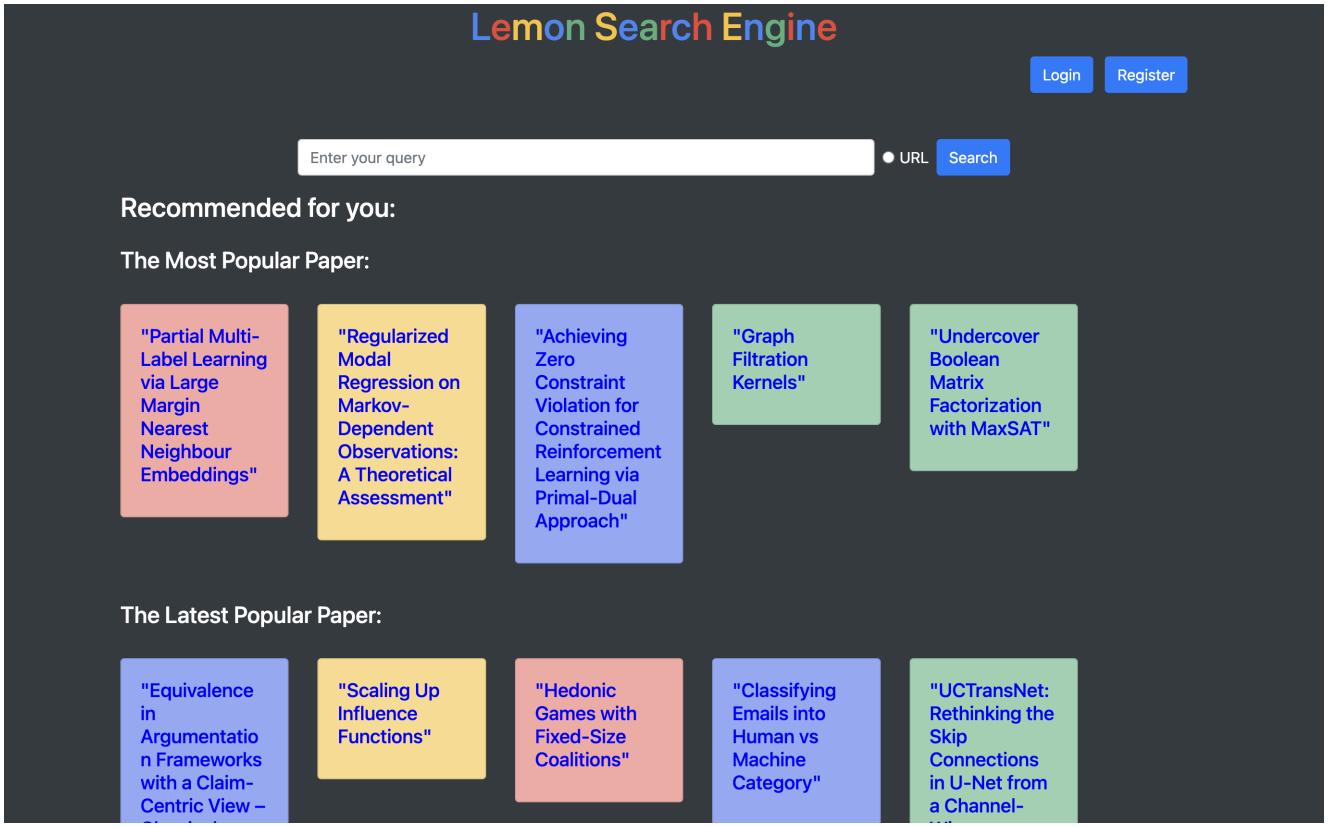


Figure 1: Home Page

Register:

If you don't have an account, you'll need to click 'Register' to sign up for a new account. The user id will be generated automatically and displayed in the popup window below. Please remember your user id as it will be used when logging in.

Figure 2: Register

Survey:

After clicking 'Register', you must complete a questionnaire to register your account successfully. As shown below, you need to select at least one article that interests you and submit it.

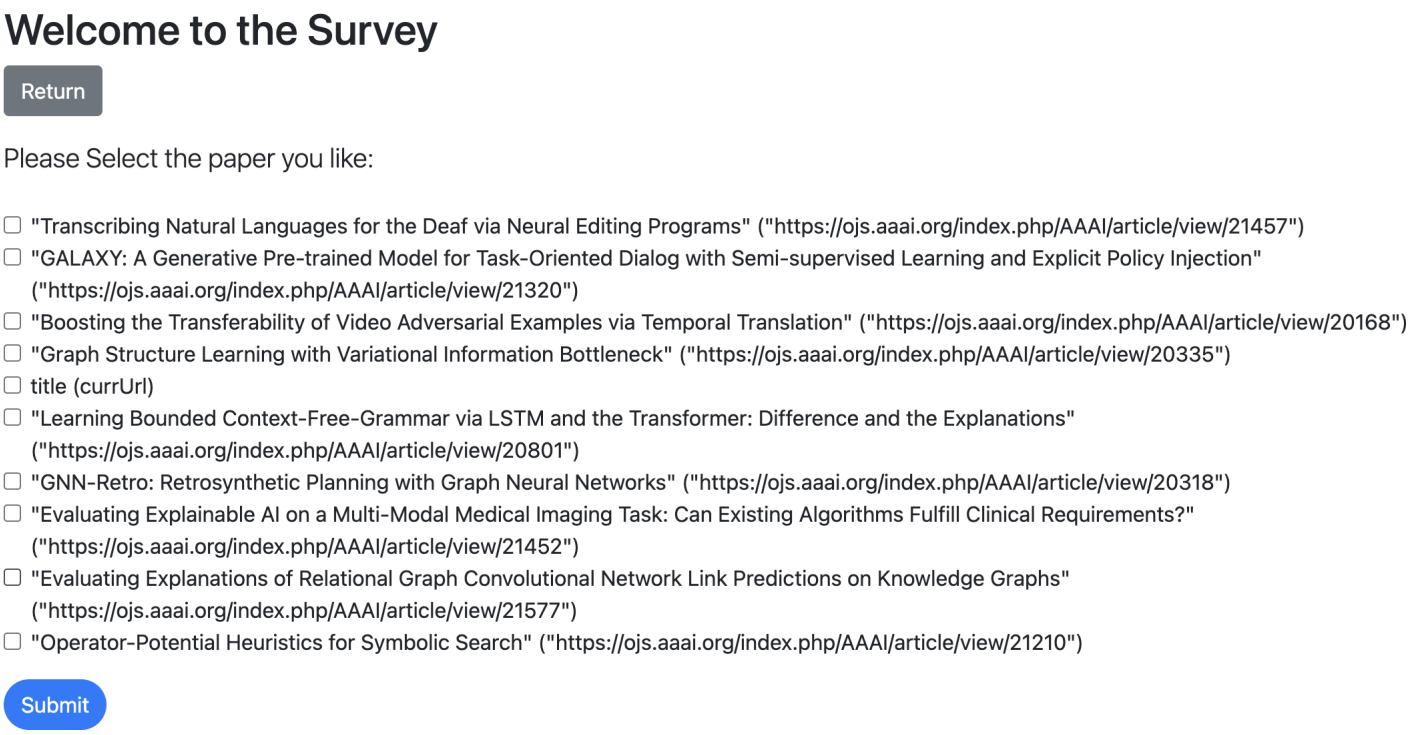


Figure 3: Survey

Search:

You can enter a keyword to search, if it matches the result, it will jump to the following page, if not, it will return to the Home Page. As shown in the figure below, the searched articles will be arranged according to the degree of match. The information about the article will be displayed on the left side of the screen. You can do some operations such as clicking on the page view, like (click on the heart icon), or favorite (click on the five-pointed star icon). All these activities will be recorded and analyzed to recommend academic articles for you. In addition, you can add a year range to re-filter the page information.

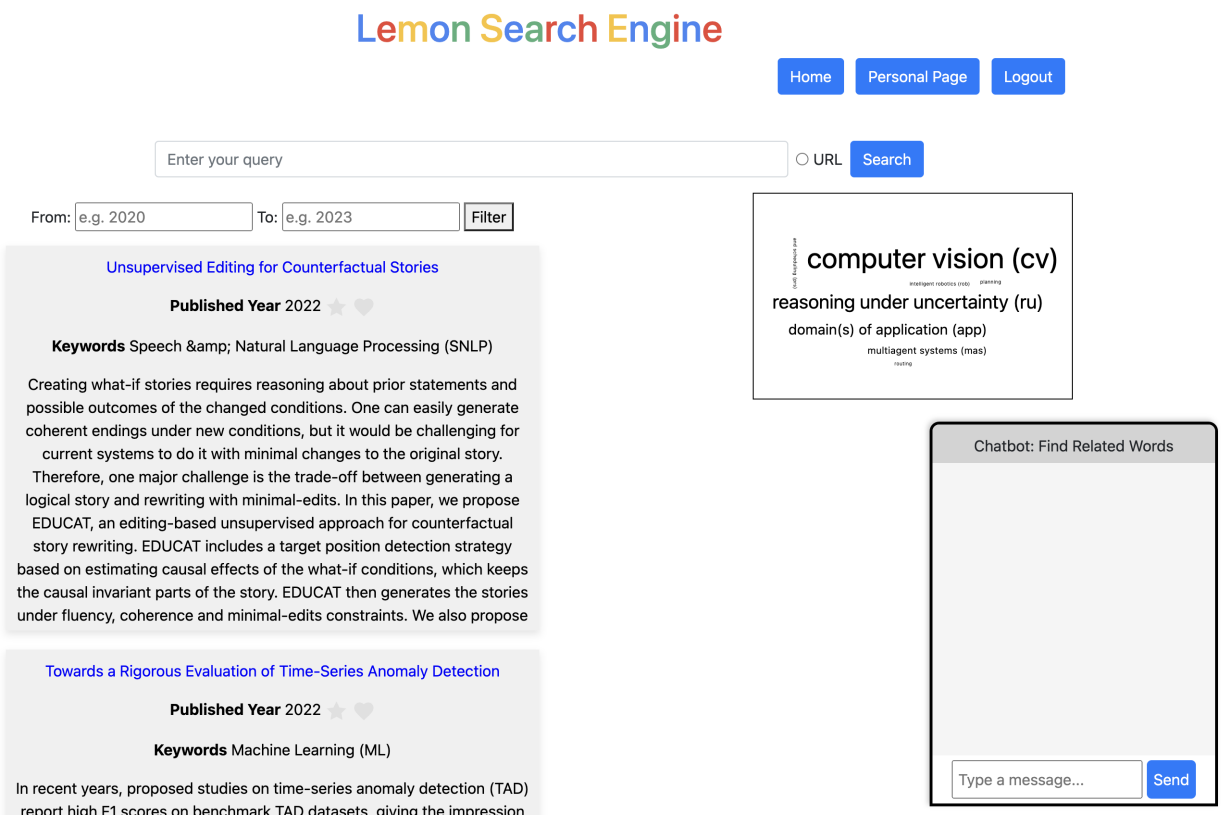


Figure 4: Search

Word Cloud Diagram:

As shown in the figure below, all the keywords in the search result articles will be organized, counted, and displayed on the word cloud map according to the frequency of occurrence (the more frequent, the larger the font size). Therefore, users can use the words in the word cloud to make a preliminary judgment whether the search terms they entered are accurate or not, and whether the matched articles are what they want or not.

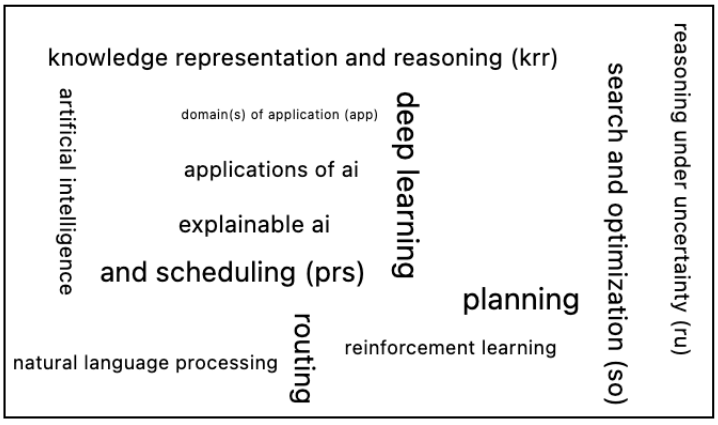


Figure 5: Word Cloud Diagram

ChatRobot:

Considering that some users are approaching an academic field for the first time, I've included the conversational bot shown below. The user can send a word to the bot and the bot will reply with 5

several words that are most relevant to the input word. With these hints, the user will find more precise search terms and thus optimize the terms.

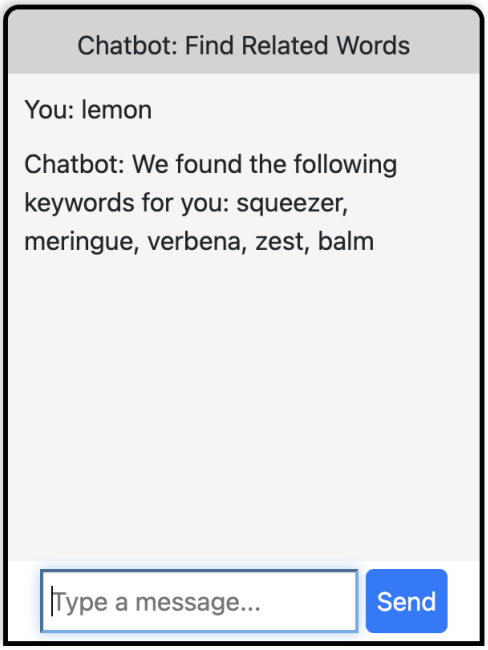
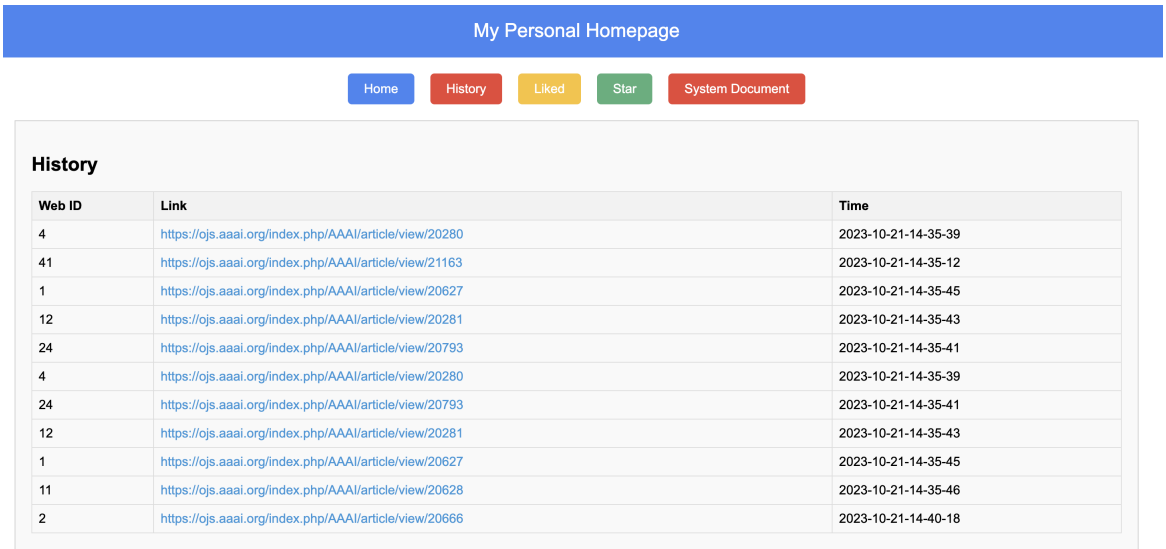


Figure 6: ChatRobot

Personal Page:

When a user has made a web page visit, liked an article or favorited an article, these records are handled differently and displayed in the 'Personal Page' (as shown below). The default display is the history and shows the time of the web page visit. Users can click on the icons above to switch between them to access Favorites, Likes, and User Guide.



Web ID	Link	Time
4	https://ojs.aaai.org/index.php/AAAI/article/view/20280	2023-10-21-14-35-39
41	https://ojs.aaai.org/index.php/AAAI/article/view/21163	2023-10-21-14-35-12
1	https://ojs.aaai.org/index.php/AAAI/article/view/20627	2023-10-21-14-35-45
12	https://ojs.aaai.org/index.php/AAAI/article/view/20281	2023-10-21-14-35-43
24	https://ojs.aaai.org/index.php/AAAI/article/view/20793	2023-10-21-14-35-41
4	https://ojs.aaai.org/index.php/AAAI/article/view/20280	2023-10-21-14-35-39
24	https://ojs.aaai.org/index.php/AAAI/article/view/20793	2023-10-21-14-35-41
12	https://ojs.aaai.org/index.php/AAAI/article/view/20281	2023-10-21-14-35-43
1	https://ojs.aaai.org/index.php/AAAI/article/view/20627	2023-10-21-14-35-45
11	https://ojs.aaai.org/index.php/AAAI/article/view/20628	2023-10-21-14-35-46
2	https://ojs.aaai.org/index.php/AAAI/article/view/20666	2023-10-21-14-40-18

Figure 7: Personal Page

Star:

It is worth noting that when the user clicks to collect, the system will further compress the article's abstract. As shown in the figure below, users can understand the general direction of the article just by reading one sentence. This saves users reading time and helps users quickly organize the article to determine the direction of interest.

Star

Web ID	Link	Summary
88	https://ojs.aaai.org/index.php/AAAI/article/view/21439	Compared with the pathological analysis and CT scans, the circulating tumor DNA methylation based approach is noninvasive and cost-effective, and thus is one of the most promising methods for early detection of lung cancer.]
53	https://ojs.aaai.org/index.php/AAAI/article/view/20637	Abstract Structural Equation/Causal Models are widely used in epidemiology and social sciences to identify and analyze the average causal effect and conditional ACE. Traditional causal effect estimation methods such as Inverse Probability Weighting and more recently Regression-With-Residuals are widely used - as they avoid the challenging task of identifying the SCM parameters - to estimate ACE and CACE. However, much work remains before traditional estimation methods can be used for counterfactual inference, and for the benefit of Personalized Public Policy Analysis in the social sciences.]
240	https://ojs.aaai.org/index.php/AAAI/article/view/20169	Abstract Vision transformers have demonstrated impressive performance on a series of computer vision tasks, yet they still suffer from adversarial examples.]

Figure 8: Star

Like:

The user's liking record is as follows.

Liked

Web ID	Link
11	https://ojs.aaai.org/index.php/AAAI/article/view/20628
18	https://ojs.aaai.org/index.php/AAAI/article/view/21285
43	https://ojs.aaai.org/index.php/AAAI/article/view/20795
88	https://ojs.aaai.org/index.php/AAAI/article/view/21439
53	https://ojs.aaai.org/index.php/AAAI/article/view/20637
205	https://ojs.aaai.org/index.php/AAAI/article/view/21186
240	https://ojs.aaai.org/index.php/AAAI/article/view/20169

Figure 9: Like

Recommendation

Finally, new recommendation methods will emerge when the user browses, likes, collects some websites, and returns to the Home Page. Beyond the previously mentioned recommendations rooted in article popularity and publication dates, we have three distinct recommendation strategies: User-based, Item-based, and mixed.

1. **User-based:** Our recommendation system adopts a user-centric collaborative filtering method, leveraging collective 'like' patterns to curate suggestions. This encompasses:
   (a) Analyzing users' 'like' actions to decipher preferences.
   (b) Forming a User-Item Matrix where rows represent users, columns denote articles, and matrix entries signify user-article interactions.

(c) Employing cosine similarity to compute the likeness in article preferences and gauge user similarity.

(d) Recognizing 'nearest neighbors' possessing the most parallel interaction patterns.

(e) Proposing articles admired by these neighbors but unseen by the target user, positing that similar users have overlapping article interests.

2. **Item-based:** Alongside the user-centric method, our platform also applies an item-driven recommendation approach. Its workings are as follows:

   (a) Extract the abstracts and keywords from articles a user expressed fondness for.

   (b) Execute a textual similarity analysis on these keywords and abstracts to pinpoint articles with aligned content.

   (c) Advocate for articles that attain the pinnacle of textual closeness, ensuring users get suggestions in line with prior preferences.

3. **Mix:** Our mechanism utilizes a hybrid recommendation modality, merging item-based filtering virtues with popularity indicators. Here's an in-depth breakdown:

   (a) Kick off using the item-centric method to discern potential articles of interest.

   (b) Classify these articles based on click frequency, prioritizing ones with heightened user involvement.

   (c) Fine-tune the recommendations by contrasting the resemblance between these sought-after articles and ones the user previously adored.

   (d) Suggest articles that echo the user's historical tastes and are also in vogue among the wider user community.

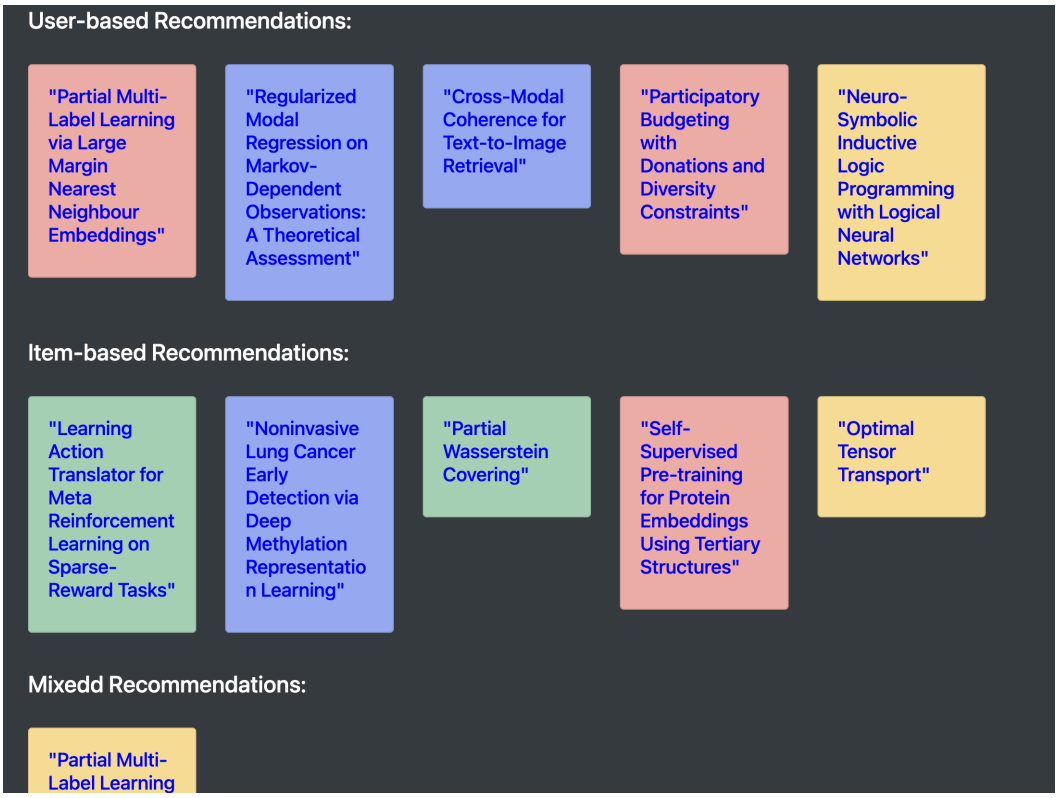


Figure 10: Recommendation