**Progress on Milestone 1**

* Milestone 1 (April 15, 2024)

1. To collect the data

Web of Science will be the venue to collect the data from. As comparing the disciplinary variation is one aim of this project, the search will not be restricted to any discipline. In addition, the year of publication will also be empty for the search. An Excel file will be downloaded and converted into CSV format.

**Progress: This has been done.**

**“information literacy” in TOPIC was searched on Web of Science;**

**I limited the data to Article;**

**The search was made on 10:41pm, April 9, 2024;**

**This results in 5,900 entries.**

1. To do data cleaning

As Web of Science only allow the users to download 1000 entries of bibliometric information in one search, the bibliometric information will need to be downloaded multiple times and collated together manually. The Excel file will contain many columns and the data is structured and relatively tidied. The columns of interest in this project include times being cited, year of publication, discipline, author name(s), title, and doi. However, other columns may be kept for the exploration of other potential analyses.

**Progress: This has been done. More specifically, 637 non-English publications and 253 + 108 (book chapters and conference papers) non-journal articles were deleted were deleted. After checking the dataset, it does not contain too many irrelevant columns. So, I did not delete any of them. The dataset for analysis contains 4,901 cases or journal articles.**

1. To get familiar with the dataset

As I get familiar with the dataset, I may refine the columns that can be included in this study.

**Progress: I included “Source Title” to get an idea on the journals that publish information literacy research. I also included “WoS Categories” to know the disciplines that publish information literacy research. “Authors” were included to know the “big names” in this subject. “Times Cited” were included to understand the most influential research in this field. By looking at their contents later, I can know what topics are most cared about or receive the most attention in this field.**

**I think I still need to see what has been included for other bibliometric analysis articles.**

* Milestone 2 (April 22, 2024)

1. To try different methods to plot the descriptive information of these articles. The primary grouping variables will be the (1) discipline and (2) year of publication. Different graphs will be tried to see the number of publications in different disciplines within each five years. This can help me to see if the research of information literacy has become increasingly popular in different disciplines. For this part, I may try pie chart, line chart, and bar chart. I may also try stacked bar chart. I will evaluate how these different charts fulfill my goals.   
   **Progress: I tried bar chart and line chart and found that bar chart is the best, because it is easy to click on.**  
   I am also considering to plot the journals where these articles were published and the major scholars in this field. This may be categorized by year of publication and discipline, but I will see if they publish a substantial number of articles in each time slot. I may also find some scholars publish in different disciplines.  
   **Progress: I have done the descriptive analysis of discipline to know the number of articles in each disciplines. I did analysis on scholars, but did not across disciplines, because I don’t think it is that meaningful to plot scholars across disciplines. There may only be a few top authors in disciplines outside of Library and Information Science.**
2. To draw a bubble plot with x-axis being year of publication and y-axis being times being cited. Each bubble represents each individual article. However, this may not be the only suitable plots for this project. I may try other methods or different ways to present the bubble plot, for example, if I need to set a threshold for the times being cited, because many articles may only be cited one or two times. In addition, the times being cited can be influenced by their year of publication. The more recent articles may have fewer citations. I will also consider set a threshold for different time slots, maybe the bottom line for the first quantile.
3. To draw separate bubble plots of different time periods.
4. To add buttons that categorize the bubbles in accordance with different research fields.  
   **Progress (for 2, 3, & 4): I did the bubble plot, but did not with different disciplines and time periods, because it is more efficient to have the buttons to select on the disciplines. For separate plots for different periods, I cannot see clear trend in only 5-year period, so I only plot in the whole time periods. For the year not ending in 2024, I will fix in the upcoming week. I may consider collapse some related categories into one.**
5. For those more recent articles, I will also try with the times being use (downloaded). This can be an alternative way to know the influence of published articles.

**Progress: I have done this.**

* Milestone 3 (April 29, 2024)

1. To add a time slider with which readers can select year of publication.
2. To try different ways to plot the keywords and most frequent authors. This can be organized by discipline but if the number of publications written by one author is not substantial. I may only plot the author data throughout the whole period.
3. To make the different visualizations more aesthetic, including size and color of each component.

**Progress: I have not done this.**

***I may consider using word cloud for keywords. I tried this and it should work for this project.***