Using Publicly Accessible Tools to Create an Open Access Index Database on Periodicals published during the Late Qing and Republican Eras

Yan He, Ka Hang Ngau, Ann James, Sophie Muro
George Washington University

March 21, 2022
Council on East Asian Libraries Annual Conference
Overview

1. Overview of Index Database
2. Data Collecting Using SFM
3. Data Extraction Using Python
4. Data Clean-Up Using Excel
5. Data Visualizations by Tableau
6. Limitations and Future Plans
1. Overview of the Index Database

- Data source – tweets
- Index database on periodicals published during the Late Qing and Republican Eras in East Asia
- The digitized content - Internet Archives
- Workflow: Social Feed Manager (collect tweets), Python (extract data), Excel (clean up data), Tableau (data visualizations).
2. Data Collecting Using SFM

- Social Feed Manager
- Capture tweets through following individual account or hashtag/keyword
- The tweets fields we used
3. Data Extraction Using Python

- Data Extraction by Python
  - Using string manipulation/text analysis (Pandas)
4. Data Clean-up Using Excel
5. Data Visualizations Using Tableau

- Dashboard/Table/Story map
- Lists, Maps, Location/Time Period Filter, language, number of issues
- Interactive
- Sharable embedded html code
- Link to the digitized copies
6. Limitations and Future Plans

- Limitations:
  - We are not content provider and content owner.
  - The digitized materials are subject to change.

- Future plans:
  - We will need to find a way to track the change of accessibility for each title.
  - We can create similar index databases using information from tweets or other data sources.
  - We would like to share our tips on how to use SFM, Python, Excel and Tableau if there is interest.
Acknowledgement

- Professor Peter Bol (project advisor) from Harvard University
- Libraries and Academic Innovation, GWU
- Columbian College of Arts and Sciences, GWU
- Anonymous Twitter users who provided information about the digital periodicals
- Anonymous content providers via Internet Archives
- Internet Archives that host digital content