3th June, 2024 - Scatter Gather

Minutes of MeetingDate: June 3, 2024

Attendees: Aravind Narayanan, Prof. Javed Mostafa

Subject: Data Collection and Integration into PATTIE for Enhanced Digital Humanities Research

Agenda:

1. Update on data collection strategy using the "Text as Data" and "The Digital Humanities and Literary Studies" references.

2. Integration and enhancements to the PATTIE platform.

Minutes:

1. **Pause on Friends of Friends Feature:**

   - Temporary hold on the integration of the Friends of Friends data structure to prioritize other data collection strategies.

2. **Data Collection from "Text as Data":**

   - Data on one-on-one authors obtained from the references of the "Text as Data" book is stored in Notion and needs to be exported into Excel/CSV/Json formats.

   - For each of the 100 authors, aim to collect 30-40 articles through database searches or possibly through AI tools like ChatGPT. Explore databases such as Google Scholar, Scopus, and Web of Science for article retrieval.

- Investigate potential APIs from these databases for direct data acquisition.

- If full articles are inaccessible due to paywalls, collect at least the titles and abstracts. The organization of data does not need to be date-specific.

- Import the collected data (approximately 2,000 papers) into PATTIE and make it searchable.

- Develop a script to automate the data collection and integration process.

3. **Continued Use of PATTIE for Data Organization:**

- PATTIE will be utilized as the primary tool for organizing and analyzing the collected data.

- Enhance PATTIE's capability to generate networks and find topic clusters.

- Invest additional effort in improving the user interface (UI) of PATTIE:

  - Implement a UI feature that allows switching between topics and authors within topic clusters.

4. **Expansion of Author Database:**

- Collect data on an additional 50 authors from another key reference, "The Digital Humanities and Literary Studies" by Martin Paul Eve.

5. **UI Development Strategy:**

- Focus on developing a functional version of the search UI tailored to the project's needs rather than learning about general UI design.
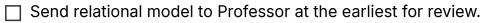
6. **UI and Label Improvement Using AI:**

- Explore the use of advanced AI models like Llama or Gemini or GPT to improve the labeling system, which will enhance the UI's effectiveness.

Action Items:

☐ Export author data from Notion to Excel/CSV/Json.

☐ Send relational model to Professor at the earliest for review.

☐ Develop and execute a script for automated data collection and integration into PATTIE.

☐ Enhance PATTIE's UI to support efficient data visualization and interaction.

☐ Begin collecting data on additional authors from the specified literature.