



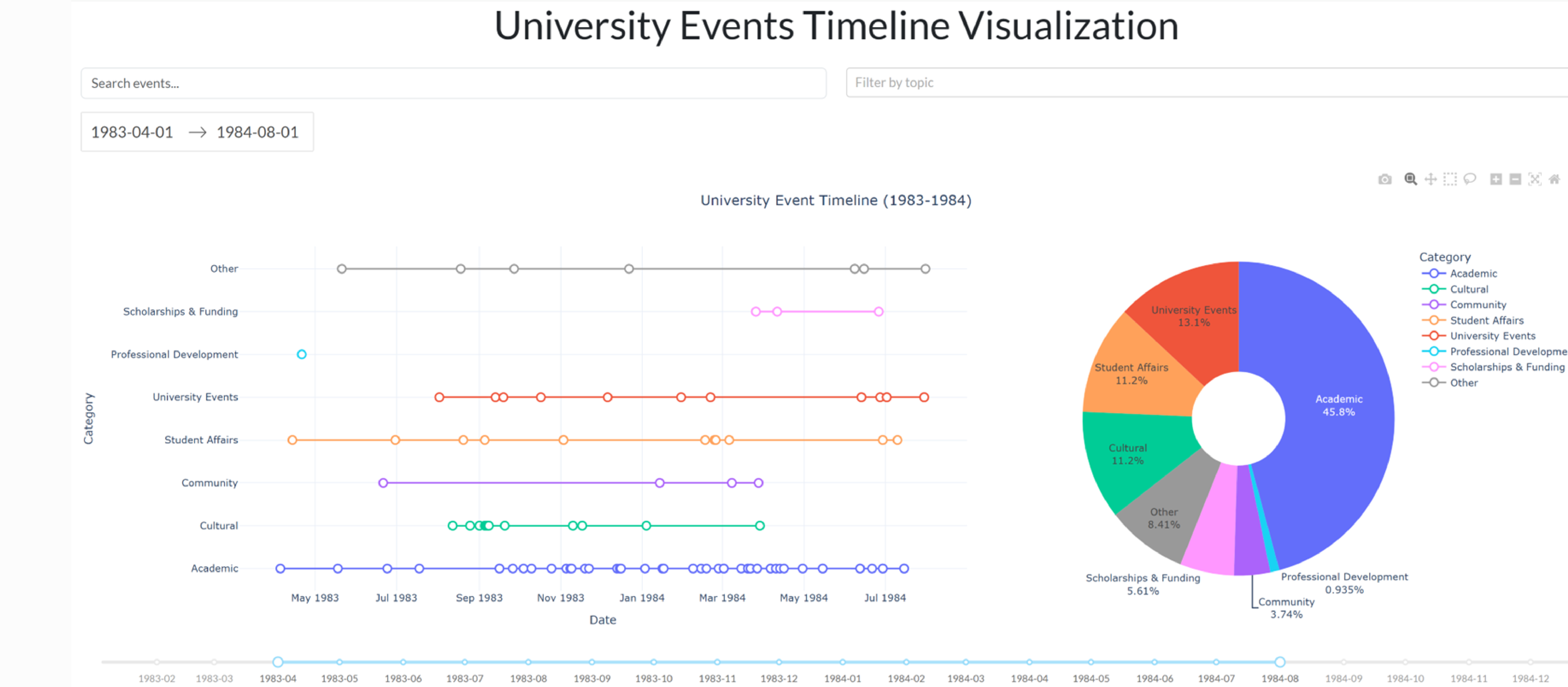
Streamlining Research: A Comprehensive Searching Platform for CUHK Press Releases

Introduction

This project aims to develop a searchable platform for **CUHK Press Releases**, enabling users to explore events by keyword, date, or tag and analyze event distributions over time. The CUHK Library collection spans **1983–2006**, with a focus on the **1983–1997** period, where many releases are handwritten or vertically printed—posing unique challenges for data analysis and visualization.

These documents record key events such as academic conferences, appointments, scholarships, and building inaugurations, offering valuable insight into CUHK's historical development.

Interactive Platform



This platform is an interactive University Events Timeline Visualization System for exploring CUHK Press Releases. It provides the following key features:

- **Calendar Date Selector:** Choose a custom date range.
- **Timeline Slider:** View events by date and category on a horizontal timeline.
- **Event Search:** Locate events using keywords.
- **Category Filter:** Narrow down results by topic.
- **Event Preview:** View event details on hover or click.
- **Pie Chart Summary:** Visualize category distribution.

This system supports efficient historical analysis through a clear and interactive interface.

Group members: HUANG Yanwen(MPhilMath/1), LU Xiaoting(BUDS/1)

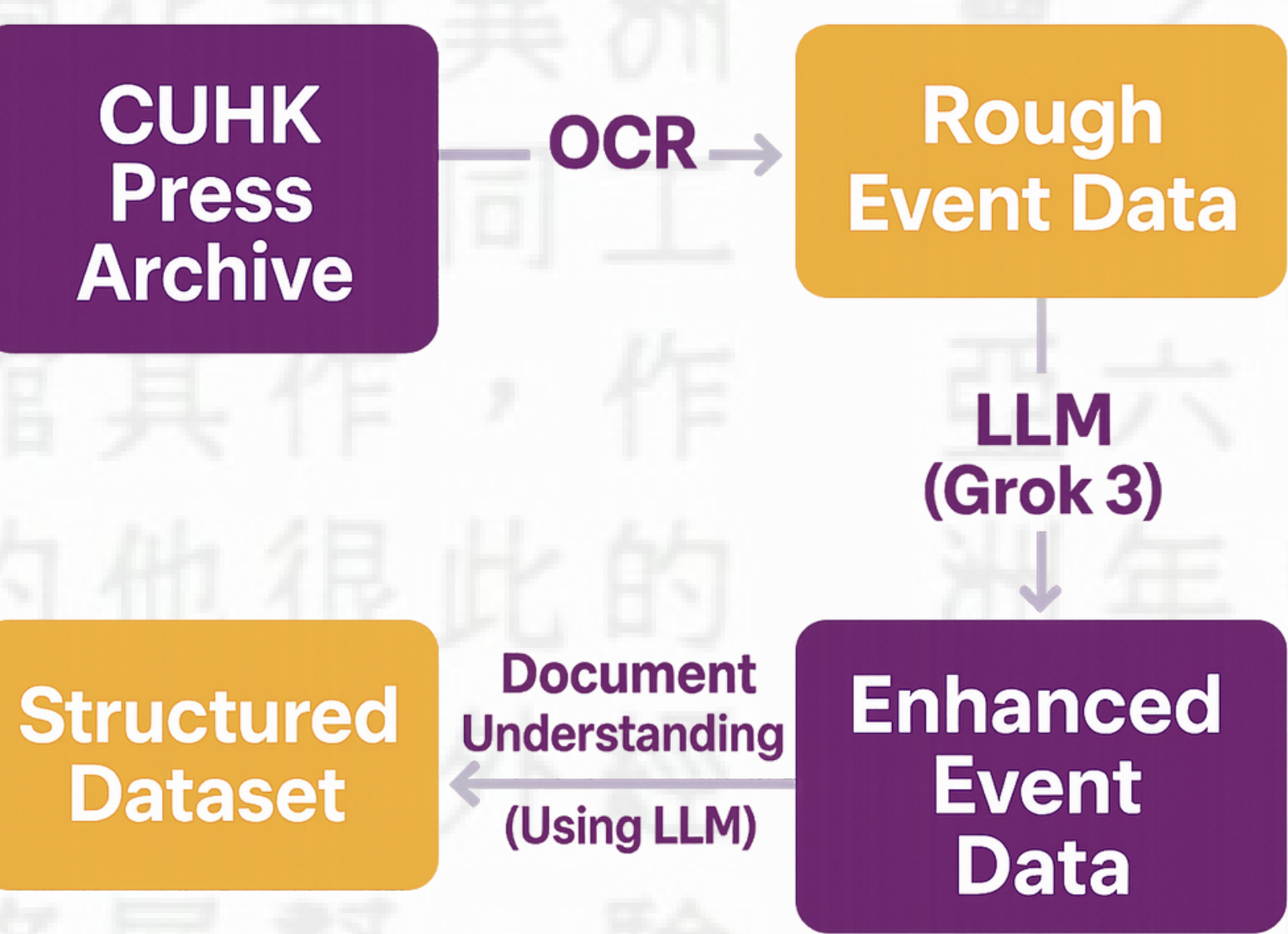
More about this project: <https://dsprojects.lib.cuhk.edu.hk/en/projects/cuhk-press-releases/introduction/>

If you would like to reuse the data, please contact data@cuhk.edu.hk



Data Processing

There are around 2500 documents in the CUHK Press Releases Archive. The Archive is digitized using OCR, extracting text from scanned documents: including challenging handwritten and vertical formats. A large language model (LLM) then refines this data, identifying key details such as dates, events, and entities. The result is a structured, searchable dataset ready for interactive exploration.



Challenges

- Redundant Content
- Mixed Formats
- Department Variations
- Category Decision