



Contribution ID: 31

Type: **Presentation**

Mining Meaning: How SMU Libraries Use NLP and AI Tools to Uncover Strategic Insights

Monday, October 13, 2025 3:25 PM (11 minutes)

Academic institutions often hold large volumes of unstructured text data—such as chat transcripts, research publications, and strategic documents—but may lack accessible methods to analyze and interpret these resources effectively. This presentation shares how Singapore Management University (SMU) Libraries leveraged BERTopic, an AI-driven topic modeling tool for text clustering, along with generative AI tools like ChatGPT and Deepseek, to extract meaningful insights from institutional data—supporting both service enhancement and strategic planning, and illustrating the growing potential of generative AI in supporting institutional goals and strategies.

In the first case, we applied BERTopic to anonymized library chat transcripts to identify recurring topics in user queries. This approach allowed us to efficiently analyze thousands of transcripts and uncover common types of enquiries received through the library's chat service. The findings provide insights that may inform future service improvements, staff training, and chatbot development—areas of growing interest for many academic libraries. By applying modern topic modeling to a traditionally underused dataset, we demonstrated how unstructured service data can support evidence-based decision-making.

In the second case, we used BERTopic to cluster and analyze a collection of publications by university faculty. The goal was to identify thematic groupings that reflect our university's research strengths and to respond to senior leadership inquiries about research trends and institutional output. To aid interpretation of the keyword-based topic representations generated by BERTopic, we used generative AI tools such as ChatGPT and Claude to produce easily understandable summaries.

Additionally, we will briefly share a small-scale application of Deepseek API for zero-shot classification to categorize faculty publications by the university's strategic priorities—demonstrating the value of generative AI for institutional insights.

Across these cases, we reflect on the strengths and limitations of using BERTopic and generative AI tools. Challenges included interpreting noisy topic groups, tuning model parameters to improve results, and balancing automation with human judgment. We discuss how data preparation, prompt refinement, and iterative experimentation influenced the outcomes. Importantly, these tools are now significantly more accessible than in the past, making it feasible for staff without deep technical expertise to conduct advanced text analysis. We also note the privacy advantages of using BERTopic, as it can be deployed locally without transmitting sensitive data to external servers. These reflections aim to provide a grounded view of how AI-based tools can be responsibly and effectively applied in institutional settings, while remaining mindful of their constraints.

These projects illustrate how librarians and institutional research staff, equipped with accessible NLP and AI tools, can contribute meaningfully to both operational and strategic initiatives. We highlight the evolving role of libraries and research support units in advancing institutional data intelligence, improving workflows, and enhancing decision-making in the age of AI.

Primary author: DONG, Danping (Singapore Management University)

Co-authors: Mr TAY, Aaron (Singapore Management University); Ms YEO, Pin Pin (Singapore Management University); Ms RATMELIA, Bella (Singapore Management University)

Presenter: DONG, Danping (Singapore Management University)

Session Classification: Presentations Session 2: Data and Research & Data Science and Data Analysis

Track Classification: SciDataCon Persistent Themes: Data Science and Data Analysis