Topic 16:

Interactive AI Assistant for Scientific Literature Review

Supervisor:

Dr. XU Lingling

Type of Project:

Application

Description:

1. Problem:

The volume of scientific literature is growing exponentially, making it nearly impossible for researchers, students, and practitioners to keep up with the latest developments in their fields. Traditional literature review processes are manual, time-consuming, and can be a significant barrier to conducting new research.

2. Objective:

This project aims to design and build an interactive web application that serves as an AI assistant to accelerate the initial stages of literature review. The system will ingest a collection of research papers on a user-defined topic, automatically organize the papers, and provide generative AI tools for summarization and synthesis, enabling users to quickly understand the key topics and papers in the field.

3. Core Question:

Can we leverage modern NLP and generative AI techniques to build a practical tool that helps scientists and students quickly navigate and understand large volumes of research literature?

4. Key Features and Workflow:

The final application will guide users through the following workflow:

- **Search:** Users enter research keywords (e.g., "Parameter-Efficient Fine-Tuning").
- **Ingest:** The system queries scientific database APIs (e.g., arXiv) to retrieve abstracts and metadata for hundreds of relevant papers.
- **Cluster:** The system uses sentence embeddings and advanced clustering techniques (BERTopic) to automatically group papers into distinct and coherent topics (e.g., "Sentiment Analysis," "Text Summarization," and "Recommender System").
- **Explore:** Users will see an interactive dashboard of identified topics and the papers they belong to.
- **Summarize & Synthesize:** The user can perform two main actions within the application's interface

✓ **Topic Synthesis:** The user can select an entire topic cluster generated by the system and click a "Synthesize Topic" button. This will generate a high-level report outlining the common methodologies, themes, and findings across all papers in that cluster.

✓ **Individual Paper Summarization:** After exploring a topic cluster, the user will see a list of papers belonging to that cluster. By clicking on any paper's title from this list, they can instantly generate a concise, easy-to-read summary of its abstract.

Reference:

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

3. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.