

Dynamic Chunking and Concept Mapping

Ananth Yegavakota, Prajna Penmetsa

Saturday, Dec 6th

Abstract/Description

This research investigates "chunking" as a collaborative meta-learning principle by developing an interactive tool that facilitates dialogue between a Subject Matter Expert (SME) and a student to negotiate the breakdown of complex academic topics. The project integrates an LLM-driven chat interface with a visual canvas (using Mermaid.js or D3.js) to generate and modify concept maps that visualize these chunks and their connections. The study aims to evaluate whether this co-creative, interactive mapping process leads to better retention and understanding of conceptual relationships compared to traditional text-only or static summary methods.

Research Questions

1. Can an LLM-driven dialogue effectively guide a student to identify meaningful "chunks" of a complex academic topic?
2. Does the co-creation of a visual concept map lead to better retention and understanding of the relationships between concepts?

What progress has been made until now?

Following our initial literature review (specifically referencing the methodology in <https://arxiv.org/pdf/2306.17459.pdf>), we have successfully established the theoretical framework and the ground-truth dataset required for the project.

Theoretical Framework Expansion: While the reference paper primarily utilizes Bloom's Taxonomy, we have expanded the research scope to include two additional pedagogical frameworks:

- Bloom's Taxonomy: For assessing cognitive complexity
- ABCD Model: Analyzing Learning Objectives based on Audience, Behavior, Condition, and Degree.
- SMART Framework: Evaluating objectives for Specificity, Measurability, Achievability, Relevance, and Time-bound constraints.

Creation of Gold Standard Dataset: We have constructed a "Gold Standard" dataset derived from verified academic documents. This involved the extraction, cleaning, and structuring of data from approximately 30 distinct courses.

- Data Extraction: We extracted high-granularity data including Course Codes, Descriptions, Learning Objectives (LOs/COs), and Detailed Syllabi from PDF course catalogs.
- Domain Diversity: The dataset so far covers a spectrum of disciplines, including:
 - Foundational CS
 - Mathematics
 - Humanities/Linguistics
- Data Structuring: The raw PDF data was parsed and converted into a structured JSON format to facilitate the "backward generation" training pipeline (Input: LOs + Syllabus → Output: Description).

Tools Used:

- Data Curation: Manual verification and PDF-to-JSON parsing scripts.
- Schema Design: JSON formatting for mapping LOs and Syllabus units (e.g., mapping specific units like "Unit 3: Context Free Grammar" to their respective COs).

Result: A verified, machine-readable Gold Standard dataset of ~30 courses is now ready for the implementation of the multi-taxonomy classification and description generation models.

What are the next steps?

1. Data Expansion (Proof of Concept - IIIT Monsoon 2024)
 - a. Scale the dataset generation by automating the extraction of course names, descriptions, LOs, and detailed syllabi for all IIIT courses offered in Monsoon 2024.
 - b. Goal: Create a comprehensive internal dataset to test scalability.
2. Multi-Model Generation (Backward Synthesis)
 - a. Implement the "Backward Generation" pipeline: Input Learning Objectives + Detailed Syllabus → Output Course Description.
 - b. Execute this task using local and API-based LLMs (targeting Gemini, Qwen, and Grok) to generate candidate descriptions.
3. Development of Evaluation Metrics
 - a. Proxy Methods: Implement "LLM-as-a-Judge" to evaluate the generated descriptions against the Gold Standard. We will define specific criteria (coherence, hallucination rate, alignment with LOs) to ensure the evaluation is justifiable and not arbitrary.
 - b. Metric Selection: Explore additional reference-free metrics for linguistic quality.

4. Human Annotation Framework

- a. Design a protocol for human evaluation to validate the "LLM-as-a-Judge" findings, ensuring the automated metrics align with human perception of quality.