

Perencanaan akademik memerlukan dukungan sistem bimbingan akademik yang efektif, tetapi dalam praktiknya peran dosen pembimbing akademik kerap terkendala oleh beban kerja, keterbatasan waktu, dan tersebarnya informasi akademik. Pemanfaatan *chatbot* berbasis *Large Language Model* (LLM) menawarkan solusi otomatisasi, tetapi masih rentan terhadap halusinasi pada domain spesifik berbasis kurikulum Program Studi (prodi) tertentu. Untuk memastikan jawaban selaras dengan struktur kurikulum, pengetahuan akademik dapat dimodelkan sebagai *Knowledge Graph* (KG) dan diakses menggunakan Cypher, yaitu bahasa *query* deklaratif untuk basis data graf yang digunakan untuk mengekspresikan pola relasi antarentitas dan mengambil data secara terstruktur. Agar pengguna tetap dapat bertanya dalam bahasa alami, LLM perlu menerjemahkan pertanyaan tersebut menjadi *query* Cypher yang tepat, suatu tugas yang dikenal sebagai *Text-to-Cypher*. Namun, integrasi LLM dan KG masih menghadapi tantangan pada tugas tersebut karena kinerjanya dipengaruhi oleh teknik *prompt engineering* dan format representasi skema KG. Oleh karena itu, penelitian ini melakukan evaluasi empiris pengaruh kombinasi kedua faktor tersebut terhadap tugas *Text-to-Cypher* berbasis LLM.

Penelitian ini menggunakan pendekatan eksperimental kuantitatif dengan membandingkan kombinasi tiga teknik *prompt engineering* (*zero-shot prompting*, *few-shot prompting*, dan *chain-of-thought prompting* (CoT)) dengan tiga format representasi skema KG (*full schema*, *nodes and paths*, dan *only paths*). Evaluasi kinerja mengadaptasi kerangka kerja Metric-Based LLM Evaluation yang menilai Cypher *query* oleh LLM berdasarkan validitas sintaksis Cypher dan kepatuhan skema KG, kemiripan terhadap *ground truth*, serta kesesuaian hasil eksekusi. Kinerja tiap konfigurasi diukur menggunakan metrik LLMetric (tingkat konfigurasi) dan LLMetric-Q (tingkat pertanyaan). Eksperimen dilakukan dengan memanfaatkan KG berisi Kurikulum 2021 Prodi Teknologi Informasi FT UGM dan model Qwen2.5 Coder 32B Instruct.

Hasil penelitian menunjukkan bahwa kombinasi CoT dengan format *only paths* memberikan kinerja terbaik dengan nilai LLMetric sebesar 71,40% dan rata-rata LLMetric-Q sebesar 71,62%. Uji Friedman mengonfirmasi adanya perbedaan signifikan antarteknik *prompt engineering*. Uji Wilcoxon menunjukkan bahwa CoT dan *few-shot prompting* sama-sama unggul signifikan dibanding *zero-shot prompting*, sementara CoT dan *few-shot prompting* tidak berbeda signifikan satu sama lain. Sementara itu, format representasi skema KG tidak menunjukkan perbedaan signifikan secara statistik. Kombinasi CoT dengan format *only paths* unggul karena CoT membantu LLM menalar jawaban langkah demi langkah, sementara format *only paths* menyajikan relasi KG yang lengkap dan valid, tetapi tetap ringkas sehingga mempermudah penyusunan Cypher *query* yang tepat. Temuan ini menegaskan bahwa teknik *prompt engineering* merupakan faktor utama dalam keberhasilan tugas *Text-to-Cypher*, sementara representasi skema KG berperan sebagai penguat ketika dipadukan dengan teknik yang tepat.

Kata kunci: *Large Language Model*, *Knowledge Graph*, *Text-to-Cypher*, *Prompt Engineering*, Skema *Knowledge Graph*

Academic planning requires effective academic advising systems; however, in practice, the role of academic advisors is often constrained by heavy workloads, limited time availability, and the dispersion of academic information. The use of chatbots powered by large language models (LLMs) offers an automation solution, but they remain prone to hallucinations in program-specific, curriculum-based domains. To ensure that answers are aligned with the curriculum structure, academic knowledge can be modeled as a knowledge graph (KG) and accessed using Cypher, a declarative query language for graph databases that is used to express relationship patterns among entities and to retrieve data in a structured manner. To allow users to keep asking questions in natural language, the LLM must translate those questions into appropriate Cypher queries, a task known as Text-to-Cypher. However, the integration of LLMs and KGs still faces challenges on this task because its performance is influenced by prompt engineering techniques and the format of KG schema representation. Therefore, this study conducts an empirical evaluation of the effect of combining these two factors on LLM-based Text-to-Cypher tasks.

This research adopts a quantitative experimental approach by comparing combinations of three prompt engineering techniques (zero-shot prompting, few-shot prompting, and chain-of-thought prompting (CoT)) with three KG schema representation formats (full schema, nodes and paths, and only paths). The performance evaluation adapts the Metric-Based LLM Evaluation framework, which assesses Cypher queries produced by the LLM based on Cypher syntactic validity and compliance with the KG schema, similarity to the ground truth, and consistency of execution results. The performance of each configuration is measured using the LLMetric (configuration level) and LLMetric-Q (question level) metrics. The experiments are conducted using a KG containing the 2021 curriculum of the Information Technology undergraduate program, Faculty of Engineering, Universitas Gadjah Mada, and the Qwen2.5 Coder 32B Instruct model.

The results show that the combination of CoT with the only paths format provides the best performance, with an LLMetric score of 71.40% and an average LLMetric-Q of 71.62%. The Friedman test confirms significant differences among prompt engineering techniques. The Wilcoxon test shows that both CoT and few-shot prompting significantly outperform zero-shot prompting, while no significant difference is observed between CoT and few-shot prompting. Meanwhile, the KG schema representation formats do not exhibit statistically significant differences. The superiority of the CoT and only paths combination arises because CoT helps the LLM reason through answers step by step, whereas the only paths format presents complete and valid KG relations in a concise manner, thereby facilitating the construction of accurate Cypher queries. These findings confirm that prompt engineering techniques are the primary factor in the success of Text-to-Cypher tasks, while KG schema representation serves as a supporting factor when paired with an appropriate technique.

Keywords: *Large Language Model, Knowledge Graph, Text-to-Cypher, Prompt Engineering, Knowledge Graph Schema*