

ABSTRAK

PENERAPAN *SEMANTIC CHUNKING* DAN *SEMANTIC RERANKING* UNTUK OPTIMALISASI PERFORMA *LARGE LANGUAGE MODEL* PADA SISTEM PEMBUATAN SOAL OTOMATIS BERBASIS *WEBSITE*

Daniel Abdillah Arif

H1D022055

Pengembangan sistem *Automatic Question Generation* (AQG) berbasis *Large Language Model* (LLM) menghadapi tantangan dalam mempertahankan relevansi konteks, terutama untuk menghasilkan soal kategori *Higher Order Thinking Skills* (HOTS). Masalah muncul ketika prapemrosesan data konvensional gagal menjaga hubungan semantik, yang memicu terjadinya halusinasi pada model akibat konteks yang terpotong. Penelitian ini bertujuan mengoptimalkan performa LLM pada sistem AQG berbasis *website* dengan menerapkan metode *semantic chunking* dan *semantic reranking* untuk meningkatkan kualitas dan konsistensi pertanyaan khususnya pada kategori HOTS. Pengujian yang dilakukan dibagi menjadi dua, yaitu pengujian fungsional menggunakan metode *blackbox testing* dan pengujian luaran model LLM menggunakan *framework* Deepeval. Hasil penelitian menunjukkan bahwa penerapan *semantic reranking* meningkatkan kualitas pengambilan informasi, dengan kenaikan skor *contextual precision* dari 0,604 menjadi 0,787 dan skor *contextual recall* dari 0,837 menjadi 0,850. Pada sisi luaran layanan AI, penerapan *semantic chunking* berhasil meningkatkan skor *faithfulness* dari 0,716 menjadi 0,855, skor *answer relevancy* dari 0,883 menjadi 0,961, serta skor *answer correctness* dari 0,777 menjadi 0,944. Secara keseluruhan, integrasi kedua metode semantik ini terbukti efektif meningkatkan ketepatan konteks dan validitas soal yang dihasilkan oleh sistem AQG.

Kata Kunci: *Automatic Question Generation, Large Language Model, Semantic Chunking, Semantic Reranking, Higher Order Thinking Skills, Deepeval.*

ABSTRACT

APPLICATION OF SEMANTIC CHUNKING AND SEMANTIC RERANKING FOR OPTIMIZING THE PERFORMANCE OF LARGE LANGUAGE MODELS IN WEBSITE-BASED AUTOMATED QUESTION GENERATION SYSTEMS

Daniel Abdillah Arif

H1D022055

The development of an Automatic Question Generation (AQG) system based on Large Language Models (LLMs) faces challenges in maintaining contextual relevance, especially when generating questions in the Higher Order Thinking Skills (HOTS) category. Problems arise when conventional data preprocessing fails to preserve semantic relationships, causing the model to hallucinate due to truncated context. This study aims to optimize LLM performance in website-based AQG systems by applying semantic chunking and semantic reranking methods to improve question quality and consistency, particularly in the HOTS category. The testing was divided into two parts, which is functional testing using the blackbox testing method and LLM model output testing using the Deepeval framework. The results showed that the application of semantic reranking improved the quality of information retrieval, with an increase in the contextual precision score from 0.604 to 0.787 and the contextual recall score from 0.837 to 0.850. On the AI service output side, the application of semantic chunking successfully increased the faithfulness score from 0.716 to 0.855, the answer relevance score from 0.883 to 0.961, and the answer correctness score from 0.777 to 0.944. Overall, the integration of these two semantic methods proved effective in improving the contextual accuracy and validity of questions generated by the AQG system.

Keywords: *Automatic Question Generation, Large Language Model, Semantic Chunking, Semantic Reranking, Higher Order Thinking Skills, Deepeval.*