UNIVERSITAS
GADJAH MADA
Analisis Penggunaan Variasi Model Retrieval Pada Pengambilan Input Few-shot untuk NER Menggunakan SLM
Faiz Miftakhur Rozaqi, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Analisis Penggunaan Variasi Model Retrieval Pada Pengambilan Input Few-shot untuk NER Menggunakan SLM

Oleh

FAIZ MIFTAKHUR ROZAQI
24/548072/PPA/06915

Named Entity Recognition (NER) merupakan komponen fundamental dalam pengembangan sistem kecerdasan buatan dan *Knowledge Graph* (KG) yang berfungsi untuk mengidentifikasi serta mengekstraksi entitas dari teks tidak terstruktur. *Large Language Models* (LLMs) telah menunjukkan potensi signifikan dalam penerapan metode *few-shot learning* melalui pendekatan *in-context learning*, yang mengeliminasi kebutuhan proses *fine-tuning* yang intensif secara komputasi. Kapabilitas ini turut diadaptasi oleh *Small Language Models* (SLMs) yang memiliki keunggulan efisiensi sumber daya, namun kinerja *few-shot* NER pada arsitektur ini masih memerlukan optimalisasi, khususnya dalam aspek seleksi data demonstrasi.

Penelitian ini bertujuan untuk menganalisis efektivitas pemilihan *input few-shot* dengan membandingkan pendekatan *semantic search*, *lexical search*, dan *hybrid search*. Penelitian dilakukan menggunakan model Gemma-3 4B pada *dataset* NER-Grit yang bersumber dari *benchmark* IndoNLU. Hasil penelitian menunjukkan bahwa penerapan *few-shot learning* secara signifikan meningkatkan rata-rata *F1-score* dari 44,63% pada *zero-shot* menjadi kisaran 58%–64%. Secara spesifik, metode leksikal murni (BM25) dan *hybrid search* terbukti secara konsisten mengungguli metode *semantic search* murni (Cosine Similarity). Kinerja rata-rata terbaik dicapai oleh skema BM25 pada jumlah demonstrasi 9-*shot* dan *Hybrid Search* beralfa 0,3 (dominan leksikal) pada konfigurasi 4-*shot*. Temuan ini menegaskan bahwa penambahan jumlah *shot* tidak berbanding lurus secara linear terhadap performa; SLM mencapai tingkat efektivitas (*sweet spot*) pada rentang 4 hingga 9 *shot*. Pasokan maksimal hingga 10 *shot* terbukti justru memicu saturasi konteks dan menyebabkan model terdistraksi oleh *noise*. Hal ini membuktikan bahwa, untuk menangkap konteks tugas ekstraksi entitas pada SLM, presisi pencocokan kata kunci eksak jauh lebih esensial dibandingkan proyeksi kemiripan vektor semantikal.

Kata-kata kunci : *Named Entity Recognition*, *Small Language Models*, *Few-shot Learning*, Gemma-3 4B.

UNIVERSITAS
GADJAH MADA
Analisis Penggunaan Variasi Model Retrieval Pada Pengambilan Input Few-shot untuk NER Menggunakan SLM
Faiz Miftakhur Rozaqi, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Analysis of Variation Retrieval Model on Few-shot Input for NER Using SLM

By

FAIZ MIFTAKHUR ROZAQI
24/548072/PPA/06915

Named Entity Recognition (NER) serves as a fundamental component in the development of artificial intelligence systems and Knowledge Graphs (KG), designed to identify and extract entities from unstructured text. Large Language Models (LLMs) have demonstrated significant potential in applying few-shot learning methods through in-context learning, eliminating the need for computationally intensive fine-tuning. This capability extends to Small Language Models (SLMs), which offer resource efficiency advantages; however, few-shot NER performance in these architectures remains an area requiring optimization, particularly regarding demonstration data selection.

This study aims to analyze the effectiveness of few-shot input selection by comparing semantic search, lexical search, and hybrid search approaches. Research was conducted using the Gemma-3 4B model on the NERGrit dataset sourced from the IndoNLU benchmark. The results demonstrate that applying few-shot learning significantly improved the average F1-score from 44.63% in zero-shot scenarios to a range of 58%–64%. Specifically, both pure lexical (BM25) and hybrid search methods consistently outperformed the pure semantic search (Cosine Similarity) approach. The highest average performance was achieved by the BM25 scheme at 9-shot and Hybrid Search with an alpha of 0.3 (lexical dominant) at 4-shot. These findings confirm a non-linear correlation in few-shot scaling; SLMs achieve optimal effectiveness (a ”sweet spot”) between 4 and 9 shots. Providing up to 10 shots was proven to trigger context saturation, causing the model to become distracted by noise. This proves that integrating exact keyword matching precision is far more essential for capturing entity extraction contexts in SLMs than mere semantic vector similarities.

Keywords: Named Entity Recognition, Small Language Models, Few-shot Learning, Gemma-3 4B.