UNIVERSITAS
GADJAH MADA
Adaptive Latent Entity Expansion For RAG-Based Question Answering System for Physical Activity Recommendations
ADIKA FARULIA ZHARFAN, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

ADAPTIVE LATENT ENTITY EXPANSION UNTUK SISTEM TANYA JAWAB BERBASIS RAG UNTUK REKOMENDASI AKTIVITAS FISIK

Oleh

Adika Farulia Zharfan
24/548239/PPA/06922

World Health Organization (WHO) menekankan bahwa kurangnya aktivitas fisik adalah salah satu penyebab utama meningkatnya risiko kematian akibat penyakit tidak menular, yang memicu banyaknya penelitian di bidang aktivitas fisik; namun, diperlukan suatu teknologi sebagai sarana untuk menyebarkan informasi tersebut agar bermanfaat secara luas, salah satunya adalah sistem Tanya Jawab. Meskipun salah satu peningkatan yang diketahui untuk sistem Tanya Jawab adalah pendekatan *retrieval-augmented generative* (RAG), salah satu kelemahan metode tersebut adalah kerentanannya terhadap tidak adanya data yang relevan di dalam konteks yang diambil karena masalah penempatan; untuk mengatasi masalah tersebut, penelitian ini mengusulkan *Adaptive Latent Entity Expansion* (ALEE) sebagai peningkatan untuk sistem QA berbasis RAG.

Proses pengambilan dokumen bertahap dijalankan dalam sebuah *pipeline* iteratif yang menggabungkan ekspansi kueri secara statistik dan semantik, serta model pemeringkatan ulang. Entitas yang diekstrak dan indeks istilah digunakan untuk memperluas kueri yang digunakan di iterasi berikutnya. Sementara itu, lapisan pemeringkatan ulang mengurangi jumlah dokumen teratas menjadi 20 pada akhir proses ALEE. Selain itu, *groupwise scoring* diimplementasikan pada periode pasca-sidang untuk mengatasi masalah pada mekanisme penilaian *default* (i.e., *pointwise scoring*) dari model pemeringkatan ulang.

Hasil penelitian menunjukkan perbedaan mencolok antara kedua metode penilaian tersebut. Sistem yang diusulkan mencapai hasil di bawah standar pada fase ablasi, validasi, dan pengujian ketika menggunakan mekanisme pemeringkatan ulang *default*, sementara penerapan *groupwise scoring* menunjukkan hasil yang berlawanan, dengan salah satu contohnya adalah tingginya skor *precision* (0,8287), *recall* (0,8972), dan skor F1 (0,8613) pada tahap validasi, yang didukung oleh hasil uji *Wilcoxon signed-rank* yang konsisten. Selain itu, pengamatan terhadap 5 data pertama yang diambil dari satu kueri pada data uji menunjukkan bahwa *groupwise inference* berhasil mencegah dokumen yang sama muncul di posisi awal dari sekumpulan dokumen yang terdiri dari 20 data. Hal tersebut tidak terjadi ketika menggunakan *pointwise scoring*.

Kata Kunci: Natural Language Processing, RAG, Question Answering, Adaptive Latent Entity Expansion, Hugging Face, Qdrant, BERTScore, BioBERT

UNIVERSITAS
GADJAH MADA
Adaptive Latent Entity Expansion For RAG-Based Question Answering System for Physical Activity Recommendations
ADIKA FARULIA ZHARFAN, Yunita Sari, S.Kom., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

ADAPTIVE LATENT ENTITY EXPANSION FOR RAG-BASED QUESTION ANSWERING SYSTEM FOR PHYSICAL ACTIVITY RECOMMENDATIONS

By

Adika Farulia Zharfan
24/548239/PPA/06922

The World Health Organization highlighted that sedentary behavior plays a major role in increased risk of fatality from non-communicable diseases, which fuels the vast number of studies in the field of physical activity; yet, a form of technology is needed as a medium to transfer such information to be widely useful, with one being Question Answering system. While one of the known improvements for QA system is the retrieval-augmented generative (RAG) approach, one weak link of the said method is on its vulnerability to not having the relevant data inside the retrieved context due to placement issues; to overcome the described problem, this study proposes Adaptive Latent Entity Expansion (ALEE) as an advancement for RAG-based QA system.

A multi-step document retrieval process is operated inside an iterative pipeline, which incorporates statistical and semantical query expansion, and a reranking model. Extracted entities and terms index were used to expand the query that will used in the upcoming iteration, while the reranking layer narrowed down the top documents to 20 at the end of the ALEE pipeline. Additionally, groupwise scoring was introduced in post-defense period to address the issue surrounding the reranker's default scoring mechanism (i.e., pointwise scoring).

The result shows a stark difference between the two scoring mechanisms. The proposed system achieved subpar result across ablation, validation, and test phase when the default reranking mechanism is active, while implementing inference of groupwise scoring shows near-polar opposite result, with one example being the high precision (0.8287), recall (0.8972), and F1 (0.8613) score in the validation stage, which was further supported by the consistent result of Wilcoxon signed-rank test. Moreover, a closer look at the first 5 passed context of a single query from the test data revealed the groupwise inference managed to prevent duplicates from populating earlier spots of a series of 20 documents, which was not the case when the said data was composed while pointwise scoring is in effect.

Keywords: Natural Language Processing, RAG, Question Answering, Adaptive Latent Entity Expansion, Hugging Face, Qdrant, BERTScore, BioBERT