GENRA: Enhancing Zero-shot Retrieval with Rank Aggregation

Georgios Katsimpras
Georgios Paliouras NCSR Demokritos, Athens, Greece gkatsibras, paliourg@iit.demokritos.gr

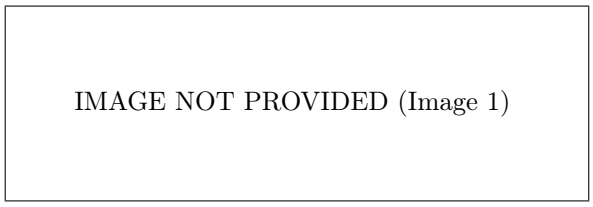

Figure 1: [Image 1 placeholder from source]

Abstract

Large Language Models (LLMs) have been shown to effectively perform zero-shot document retrieval, a process that typically consists of two steps: i) retrieving relevant documents, and ii) re-ranking them based on their relevance to the query. This paper presents GENRA, a new approach to zero-shot document retrieval that incorporates rank aggregation to improve retrieval effectiveness. Given a query, GENRA first utilizes LLMs to generate informative passages that capture the query's intent. These passages are then employed to guide the retrieval process, selecting similar documents from the corpus. Next, we use LLMs again for a second refinement step. This step can be configured for either direct relevance assessment of each retrieved document or for re-ranking the retrieved documents. Ultimately, both approaches ensure that only the most relevant documents are kept. Upon this filtered set of documents, we perform multi-document retrieval, generating individual rankings for each document. As a final step, GENRA leverages rank aggregation, combining the individual rankings to produce a single refined ranking. Extensive experiments on benchmark datasets demonstrate that GENRA improves existing approaches, highlighting the effect...

[MISSING]

[CONTENT BETWEEN ABSTRACT AND REFERENCES IS MISSING FROM SOURCE SNIPPETS]

References

[1] [MISSING AUTHORS]. 2003. ...s of the 2003 ACM symposium on Applied computing, pages 841–846.

[2] Stephen Robertson, Hugo Zaragoza, et al. 2009. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389.

[3] Devendra Singh Sachan, Mike Lewis, Dani Yogatama, Luke Zettlemoyer, Joelle Pineau, and Manzil Zaheer. 2023. Questions are all you need to train a dense passage retriever. *Transactions of the Association for Computational Linguistics*, 11:600–616.

[4] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

[5] Devendra Singh, Siva Reddy, Will Hamilton, Chris Dyer, and Dani Yogatama. 2021. End-to-end training of multi-document reader and retriever for open-domain question answering. *Advances in Neural Information Processing Systems*, 34:25968–25981.

[6] Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin, and Zhaochun Ren. 2023. Is ChatGPT Good at Search? Investigating...