Contriever and DPR Retrieval Accuracy on TriviaQA with Extended Context Windows

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does the retrieval accuracy of Contriever and DPR encoders compare on the TriviaQA benchmark when the context window size is increased to 4096 tokens. The advent of contextualised language models has brought gains in search effectiveness, not just when applied for re-ranking the output of classical weighting models such as BM25, but also when used directly for passage indexing and retrieval, a technique which is called dense. 17 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: On Single and Multiple Representations in Dense Passage Retrieval. Research question: How does the retrieval accuracy of Contriever and DPR encoders compare on the TriviaQA benchmark when the context window size is increased to 4096 tokens?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.2/10.

3 Results

14 papers retrieved. 17 claims extracted; 0 independently verified. Quality review score: 2.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The MSMARCO passage ranking dataset contains 8.8M passages.	×	0.05
The ANCE implementation and the ColBERT implementation were adapted using integrations with PyTerrier.	×	0.02
The ANCE model for the MSMARCO passage ranking dataset was provided by the authors.	×	0.05
ColBERT was trained using the MSMARCO passage ranking training triples file for 44,500 batches.	×	0.04
ColBERT fine-tunes the bert-base-uncased BERT model.	×	0.05
ANCE fine-tunes a RoBERTa model (specifically roberta-base).	×	0.05
A RoBERTa-based ColBERT model was trained for 300k batches and had relative performance 25% less than the BERT-based Col	×	0.05
The ANCE document index is stored in FAISS using the uncompressed IndexFlatIP format.	×	0.02
The ColBERT document index is stored in FAISS using the compressed and quantised IndexIVFPQ format.	×	0.02
The ColBERT document index is trained on a random 5% sample of the document embeddings.	×	0.05
Mean response times for ANCE is 211ms and for ColBERT is 635ms.	×	0.04
The publicly available query sets with relevance assessments include 5000 queries sampled from the MSMARCO Dev set and t	×	0.04
The MSMARCO Dev set contains on average 1.1 judgements per query.	×	0.01
The TREC 2019 query set contains 43 queries with an average of 215.3 judgements per query.	×	0.04
There are three major types of dysarthria in cerebral palsy: spastic, dyskinetic (athetosis) and ataxic.	×	0.01
The main event that led the US to entering WW2 was Japan bombing Pearl Harbor.	×	0.02
The U.S entered WW1 for several reasons, including the Germans declaring unlimited German submarine warfare and the Zimm	×	0.00

References

- http://arxiv.org/abs/2509.17829v1
- http://arxiv.org/abs/2402.13753v1
- http://arxiv.org/abs/2108.06279v2