Impact of Document Chunking Strategies on Blended RAG Performance in TriviaQA with Llama-3-8B

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the impact of different document chunking strategies on the accuracy and retrieval efficiency of Blended RAG when applied to the TriviaQA benchmark with Llama-3-8B. The performance of Retrieval-Augmented Generation (RAG) systems in information retrieval is significantly influenced by the characteristics of the documents being processed. In this study, the structured nature of textbooks, the conciseness of articles, and the narrative. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Exploring Information Retrieval Landscapes: An Investigation of a Novel Evaluation Techniques and Comparative Document Splitting Methods. Research question: What is the impact of different document chunking strategies on the accuracy and retrieval efficiency of Blended RAG when applied to the TriviaQA benchmark with Llama-3-8B?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

16 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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 evaluation framework incorporates four primary metrics: SequenceMatcher, BLEU, METEOR, and BERT Score.	×	0.15
SequenceMatcher is assigned a weight of 0.30 in the multi-metric scoring approach.	×	0.05
BLEU is assigned a weight of 0.30 in the multi-metric scoring approach.	×	0.05
METEOR is assigned a weight of 0.20 in the multi-metric scoring approach.	×	0.05
BERT Score is assigned a weight of 0.20 in the multi-metric scoring approach.	×	0.07
The ANOVA test for LMS Document Type yielded an F-Value of 12.655763 and a p-Value of 0.000005.	×	0.02
The ANOVA test for OAI Document Type yielded an F-Value of 14.179976 and a p-Value of 0.000001.	×	0.01
The pairwise comparison between Article and Novel document types resulted in a mean difference of -0.0117 with an adjust	×	0.03
The pairwise comparison between Novel and Textbook document types resulted in a mean difference of 0.0005 with an adjust	×	0.03
The Recursive Character Splitter achieved an LMS Final Avg Score of 0.208 based on 147 samples.	×	0.08
The Token Text Splitter achieved an OAI Final Avg Score of 0.223 based on 148 samples.	×	0.04
The ANOVA test for Splitting Method on LMS scores yielded an F-value of 62.91572 and a p-value of 4.587429e-14.	×	0.02
The ANOVA test for Splitting Method on OAI scores yielded an F-value of 46.976114 and a p-value of 4.247097e-11.	×	0.02

References

- http://arxiv.org/abs/2409.08479v2
- http://arxiv.org/abs/2602.22225v1

- http://arxiv.org/abs/2404.07220v2