

Name=shuvransu Halder(MSR)

MRR score: (KG+chunking)

Mean Reciprocal Rank (MRR) measures how well your retriever ranks the correct or relevant context among all retrieved contexts.

For my rag it is: 0.87(with 400 token,5 context)

0.89(with 800 token,5 contexts)

0.94(800 tokens,10 contexts)

0.92(800 tokens,7 context)

, which is pretty good.

RAGAS benchmarking

Context Recall

- Definition: Measures how well the RAG system retrieves relevant documents or context that actually contain the information needed to answer the query.

Faithfulness

- Definition: Measures how much the generated answer stays true to the retrieved context.
- High faithfulness → model only generates information supported by the retrieved documents.

Factual Correctness

- Definition: Measures how factually accurate the generated answer is with respect to ground truth, regardless of whether it matches the retrieved context.

Initial scores: KG+chunking(400 token,5 context)

context_recall	0.4201
faithfulness	0.8413
factual_correctness	0.6252

KG+chunking(800 tokens,5 context)

context_recall	0.4726
faithfulness	0.8484
factual_correctness	0.7838

KG+chunking(800 tokens,6 context)

context_recall	0.5528
faithfulness	0.8571
factual_correctness	0.7617

Pretty good faithfulness and factual correctness.

Context tuning paper:top 5 document

context_recall	0.3542
faithfulness	0.7282
factual_correctness	0.5629

Context tuning paper: 10 documents,chunk size=400

context_recall	0.3746
faithfulness	0.7602
factual_correctness	0.6624

Context tuning paper: 10 documents,chunk size=200

context_recall	0.3516
faithfulness	0.7186
factual_correctness	0.5810

Context tuning paper: 10 documents,chunk size=800

context_recall	0.3937
faithfulness	0.7892
factual_correctness	0.7594