GINGER
Grounded Information Nugget-Based Generation of Responses

Weronika Łajewska
Krisztian Balog

Information Access and Artificial Intelligence

University of Stavanger

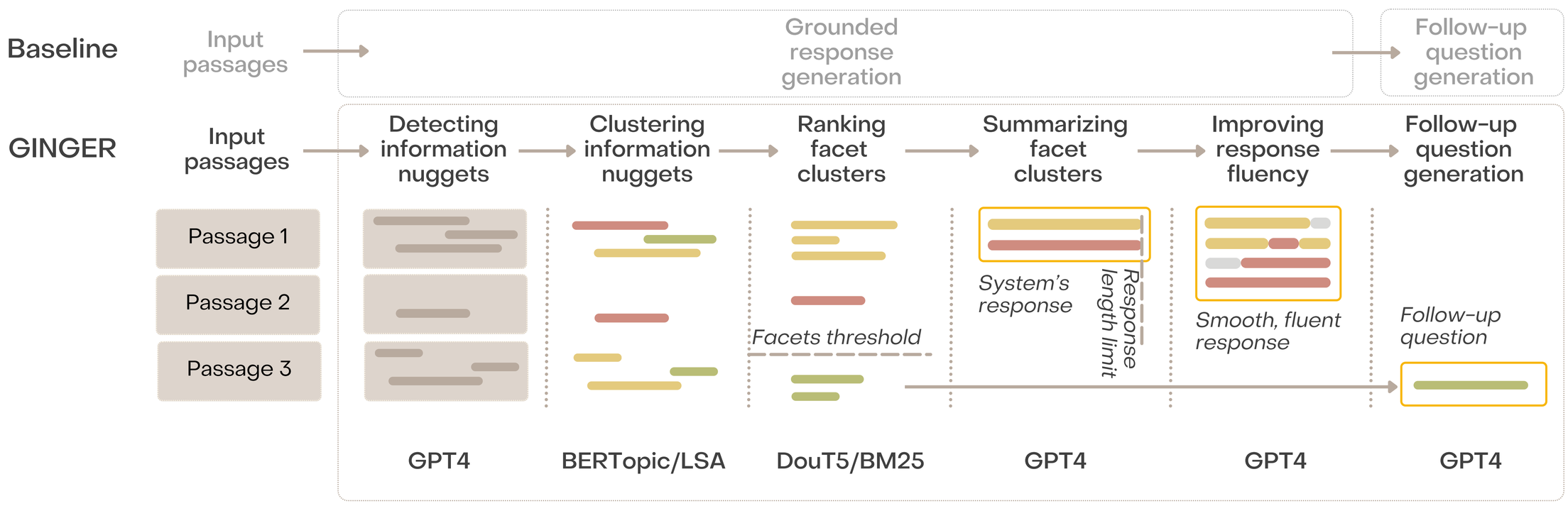
Paper:

Repository:

- A novel modular response generation pipeline that operates on information nuggets, enabling precise grounding and facilitating source verification
- Evaluated on the TREC RAG'24 augmented generation task and in the LiveRAG Challenge

KEY ADVANTAGES OF GINGER

- Maximizing information within response length limits
- Guiding users with relevant follow-up questions
- Providing source attribution for verifiability
- Allowing control over response completeness



GINGER FEATURES

- Employs context curation, allowing summarization to operate on a shorter but more relevant context
- Conditions follow-up question generation on information nuggets to ensure the relevance and answerability of the questions
- Operates on information nuggets in all pipeline components, ensuring the grounding of the final response in the source passages

RESULTS @ TREC RAG'2024

Method	V_strict	
	TREC	avg LLM
baseline-top5	0.247	0.442
baseline_CoT-top5	---	0.428
Webis	0.357	---
TREMA	0.261	---
GINGER-top20 wo/ rewriting	0.427	0.568
GINGER-top10 wo/ rewriting	0.369	0.502
GINGER-top5 wo/ rewriting	0.213	0.362
GINGER-top5	0.211	0.377

- The best-performing variant of our system outperforms both baseline approaches
- GINGER achieves state-of-the-art performance on TREC RAG'24
- Responses generated from more passages are of higher quality, indicating that the additional context is indeed utilized by GINGER
- LLMs can be used to refine response fluency without sacrificing quality or grounding
- GINGER's performance mainly stems from operating on information nuggets, not from tweaking individual components