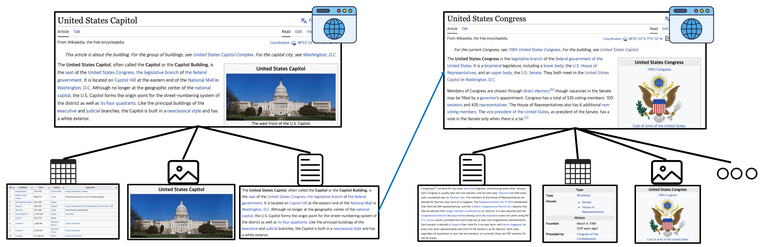
Failure is Feedback: History-Aware Backtracking for Agentic Traversal in Multimodal Graphs

Joohyung Yun [1], Doyup Lee [2], Wook-Shin Han [1]
POSTECH [1] DirectorLabs [2]

Open-Corpus Multimodal Retrieval

The task of retrieving query-relevant components (paragraphs, tables, images) from a large, interlinked multimodal document corpus.

Background: Layered Representation of Documents



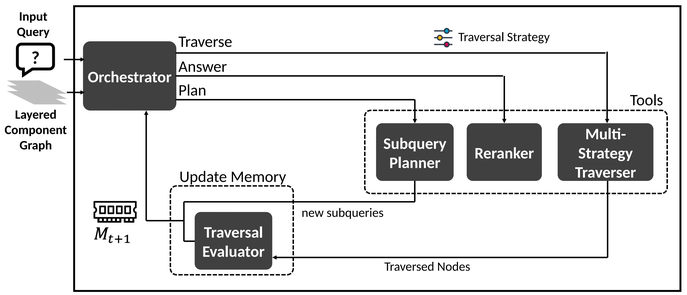
Recent works have represented documents as a **Layered Component Graph (LCG)**, hierarchically linking components and their subcomponents while connecting documents through navigational edges.

Motivation

Existing methods perform multihop traversal over a layered component graph—but they are too rigid to navigate it reliably.

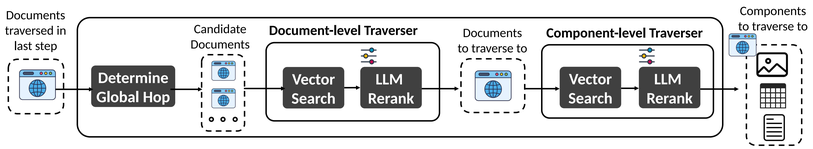
- **Hop-agnostic scoring:** rely on a uniform similarity metric across all hops
  → follows spurious cues to irrelevant components
- **Rigid, pre-defined plans:** execute a fixed retrieval procedure
  → cannot refine underspecified hops or adapt the route
- **No recovery from dead ends:** cannot turn failures into feedback
  → errors propagate and compound across hops

Failure is Feedback



We present an agentic retriever over the layered component graph, where failure is not a dead end but a **signal**: a misstep is diagnosed and turned into guidance.

Multi-Strategy Traverser



Executes a single retrieval hop over the layered component graph in two stages:
(1) **Document-level traverser** that shortlists candidate documents,
(2) **Component-level traverser** that pinpoints query-relevant components among them.
Configurable strategy engine driven by a tuple τ_t trades accuracy for cost — tuning hop scope, vector granularity, and LLM-reasoning — so it starts cheap and escalates only when a hop is ambiguous or fails.

Memory Structure

A structured memory logs the entire retrieval trajectory, including

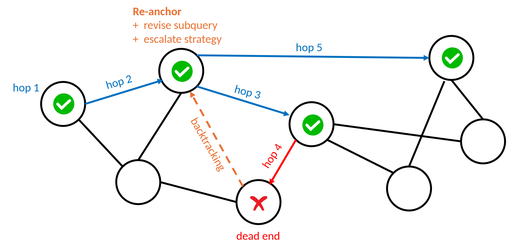
- **Subquery List (Q_t)**: the evolving retrieval plan

q_1 Solved	q_2 Solved	q_3 Deprecated	q_4 Active

- **Action History (H_t)**: an ordered log of each hop's strategy, anchor, retrieved evidence, and success/failure outcome

h_1 Plan	h_2 Search q_1 τ_1 ✓	h_3 Search q_2 τ_2 ✓	h_4 Search q_3 τ_3 ✗	h_5 Search q_3 τ_4 ✗	h_6 Plan

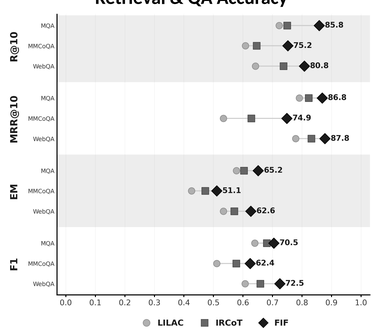
History-Aware Backtracking



Unlike standard backtracking that simply reverts and retries blindly, we treat each dead end as a **signal**: we piggyback on the failure trace to *re-anchor* to a promising prior document, *revise* the subquery with the constraints the failure revealed, and *escalate* to a stronger strategy.

Experiments

Retrieval & QA Accuracy



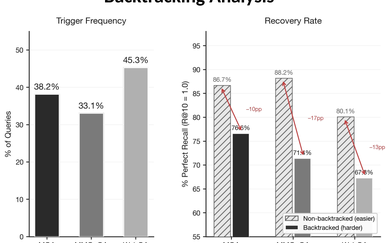
Efficiency

Dataset	Method	Time (s)	# LLM Calls	API Cost ($)
MQA	LILAC	19.53	1.00	0.0115
	IRCoT	95.74	5.44	0.0752
	FIF	**114.59**	**7.92**	**0.1109**
MMCoQA	LILAC	20.24	1.00	0.0107
	IRCoT	97.82	5.61	0.0753
	FIF	**105.36**	**7.31**	**0.1037**
WebQA	LILAC	19.19	1.00	0.0077
	IRCoT	173.82	6.94	0.1082
	FIF	**128.75**	**8.61**	**0.1278**

Ablation Study

Configuration	R@10 (%) MQA	R@10 (%) WebQA	API Cost ($) MQA	API Cost ($) WebQA
Full FIF	**85.66**	**80.92**	0.1083	0.1251
w/o Backtracking Orch.	79.12	77.61	0.1468	0.1618
w/o LLM Reasoning	77.68	78.52	0.0822	0.0997
w/o Global Hop	75.59	73.48	0.1048	0.1231
w/o Vector Granularity	83.92	79.11	0.1173	0.1326
w/o Subquery Planner	76.21	76.82	0.0717	0.0791

Backtracking Analysis



Budget Sensitivity

Budget	Method	R@10 (%) MQA	MMCoQA	WebQA	Avg
$0.05	IRCoT	71.27	62.14	69.32	67.58
	MoLoRAG	68.66	63.76	68.03	66.82
	FIF	**72.02**	**62.96**	**70.10**	**68.36**
$0.075	IRCoT	74.80	64.72	72.24	70.59
	MoLoRAG	70.73	65.20	69.88	68.60
	FIF	**78.15**	**67.86**	**73.90**	**73.30**
$0.10	IRCoT	75.30	65.32	73.10	71.24
	MoLoRAG	72.79	66.18	70.23	69.73
	FIF	**83.12**	**74.11**	**78.31**	**78.51**