Graph RAG

EmergentAGI Inc. 08/08/24

Many RAG algorithms

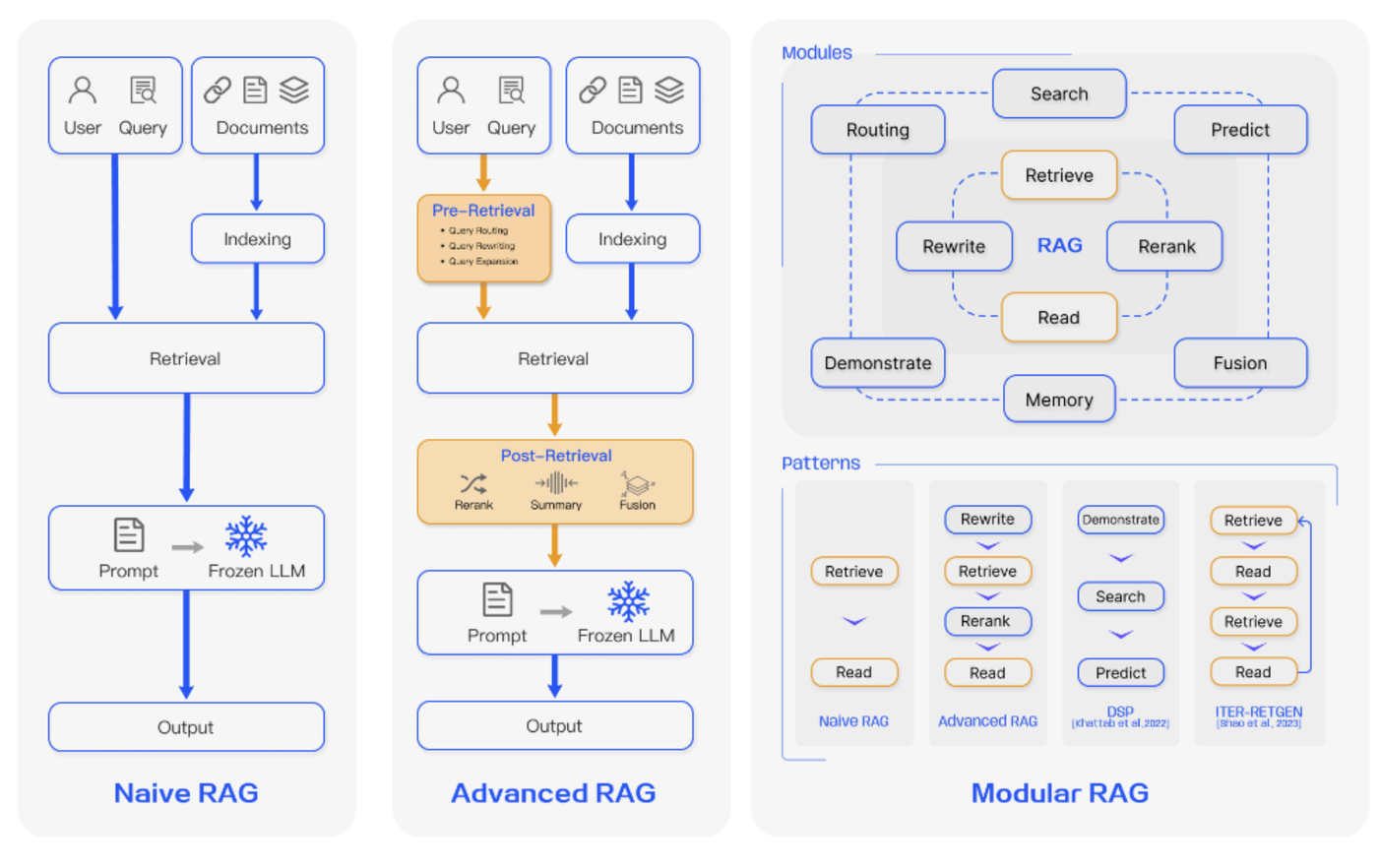


TABLE I
SUMMARY OF RAG METHODS

Method	Retrieval Source	Retrieval Data Type	Retrieval Granularity	Augmentation Stage	Retrieval process
CoG [29]	Wikipedia	Text	Phrase	Pre-training	Iterative
DenseX [30]	FactoidWiki	Text	Proposition	Inference	Once
EAR [31]	Dataset-base	Text	Sentence	Tuning	Once
UPRISE [20]	Dataset-base	Text	Sentence	Tuning	Once
RAST [32]	Dataset-base	Text	Sentence	Tuning	Once
Self-Mem [17]	Dataset-base	Text	Sentence	Tuning	Iterative
FLARE [24]	Search Engine,Wikipedia	Text	Sentence	Tuning	Adaptive
PGRA [33]	Wikipedia	Text	Sentence	Inference	Once
FILCO [34]	Wikipedia	Text	Sentence	Inference	Once
RADA [35]	Dataset-base	Text	Sentence	Inference	Once
Filter-rerank [36]	Synthesized dataset	Text	Sentence	Inference	Once
R-GQA [37]	Dataset-base	Text	Sentence Pair	Tuning	Once
LLM-R [38]	Dataset-base	Text	Sentence Pair	Inference	Iterative
TIGER [39]	Dataset-base	Text	Item-base	Pre-training	Once
LM-Indexer [40]	Dataset-base	Text	Item-base	Tuning	Once
BEQUE [9]	Dataset-base	Text	Item-base	Tuning	Once
CT-RAG [41]	Synthesized dataset	Text	Item-base	Tuning	Once
Atlas [42]	Wikipedia, Common Crawl	Text	Chunk	Pre-training	Iterative
RAVEN [43]	Wikipedia	Text	Chunk	Pre-training	Once
RETRO++ [44]	Pre-training Corpus	Text	Chunk	Pre-training	Iterative
INSTRUCTRETRO [45]	Pre-training corpus	Text	Chunk	Pre-training	Iterative
RRR [7]	Search Engine	Text	Chunk	Tuning	Once
RA-e2e [46]	Dataset-base	Text	Chunk	Tuning	Once
PROMPTAGATOR [21]	BEIR	Text	Chunk	Tuning	Once
AAR [47]	MSMARCO,Wikipedia	Text	Chunk	Tuning	Once
RA-DIT [27]	Common Crawl,Wikipedia	Text	Chunk	Tuning	Once
RAG-Robust [48]	Wikipedia	Text	Chunk	Tuning	Once
RA-Long-Form [49]	Dataset-base	Text	Chunk	Tuning	Once
CoN [50]	Wikipedia	Text	Chunk	Tuning	Once
Self-RAG [25]	Wikipedia	Text	Chunk	Tuning	Adaptive
BGM [26]	Wikipedia	Text	Chunk	Inference	Once
CoQ [51]	Wikipedia	Text	Chunk	Inference	Iterative
Token-Elimination [52]	Wikipedia	Text	Chunk	Inference	Once
PaperQA [53]	Arxiv,Online Database,PubMed	Text	Chunk	Inference	Iterative
NoiseRAG [54]	FactoidWiki	Text	Chunk	Inference	Once
IAG [55]	Search Engine,Wikipedia	Text	Chunk	Inference	Once
NoMIRACL [56]	Wikipedia	Text	Chunk	Inference	Once
ToC [57]	Search Engine,Wikipedia	Text	Chunk	Inference	Recursive
SKR [58]	Dataset-base,Wikipedia	Text	Chunk	Inference	Adaptive
ITRG [59]	Wikipedia	Text	Chunk	Inference	Iterative
RAG-LongContext [60]	Dataset-base	Text	Chunk	Inference	Once
ITER-RETGEN [14]	Wikipedia	Text	Chunk	Inference	Iterative
IRCoT [61]	Wikipedia	Text	Chunk	Inference	Recursive
LLM-Knowledge-Boundary [62]	Wikipedia	Text	Chunk	Inference	Once
RAPTOR [63]	Dataset-base	Text	Chunk	Inference	Recursive
RECITE [22]	LLMs	Text	Chunk	Inference	Once
ICRALM [64]	Pile,Wikipedia	Text	Chunk	Inference	Iterative
Retrieve-and-Sample [65]	Dataset-base	Text	Doc	Tuning	Once
Zemi [66]	C4	Text	Doc	Tuning	Once
CRAG [67]	Arxiv	Text	Doc	Inference	Once
1-PAGER [68]	Wikipedia	Text	Doc	Inference	Iterative
PRCA [69]	Dataset-base	Text	Doc	Inference	Once
QLM-Doc-ranking [70]	Dataset-base	Text	Doc	Inference	Once
Recomp [71]	Wikipedia	Text	Doc	Inference	Once
DSP [23]	Wikipedia	Text	Doc	Inference	Iterative
RePLUG [72]	Pile	Text	Doc	Inference	Once
ARM-RAG [73]	Dataset-base	Text	Doc	Inference	Iterative
GenRead [13]	LLMs	Text	Doc	Inference	Iterative
UniMS-RAG [74]	Dataset-base	Text	Multi	Tuning	Once
CREA-ICL [19]	Dataset-base	Crosslingual,Text	Sentence	Inference	Once
PKG [75]	LLM	Tabular,Text	Chunk	Inference	Once
SANTA [76]	Dataset-base	Code,Text	Item	Pre-training	Once
SURGE [77]	Freebase	KG	Sub-Graph	Tuning	Once
MK-ToD [78]	Dataset-base	KG	Entity	Tuning	Once
Dual-Feedback-ToD [79]	Dataset-base	KG	Entity Sequence	Tuning	Once
KnowledGPT [15]	Dataset-base	KG	Triplet	Inference	Muti-time
FABULA [80]	Dataset-base,Graph	KG	Entity	Inference	Once
HyKGE [81]	CMeKG	KG	Entity	Inference	Once
KALMV [82]	Wikipedia	KG	Triplet	Inference	Iterative
RoG [83]	Freebase	KG	Triplet	Inference	Iterative
G-Retriever [84]	Dataset-base	TextGraph	Sub-Graph	Inference	Once

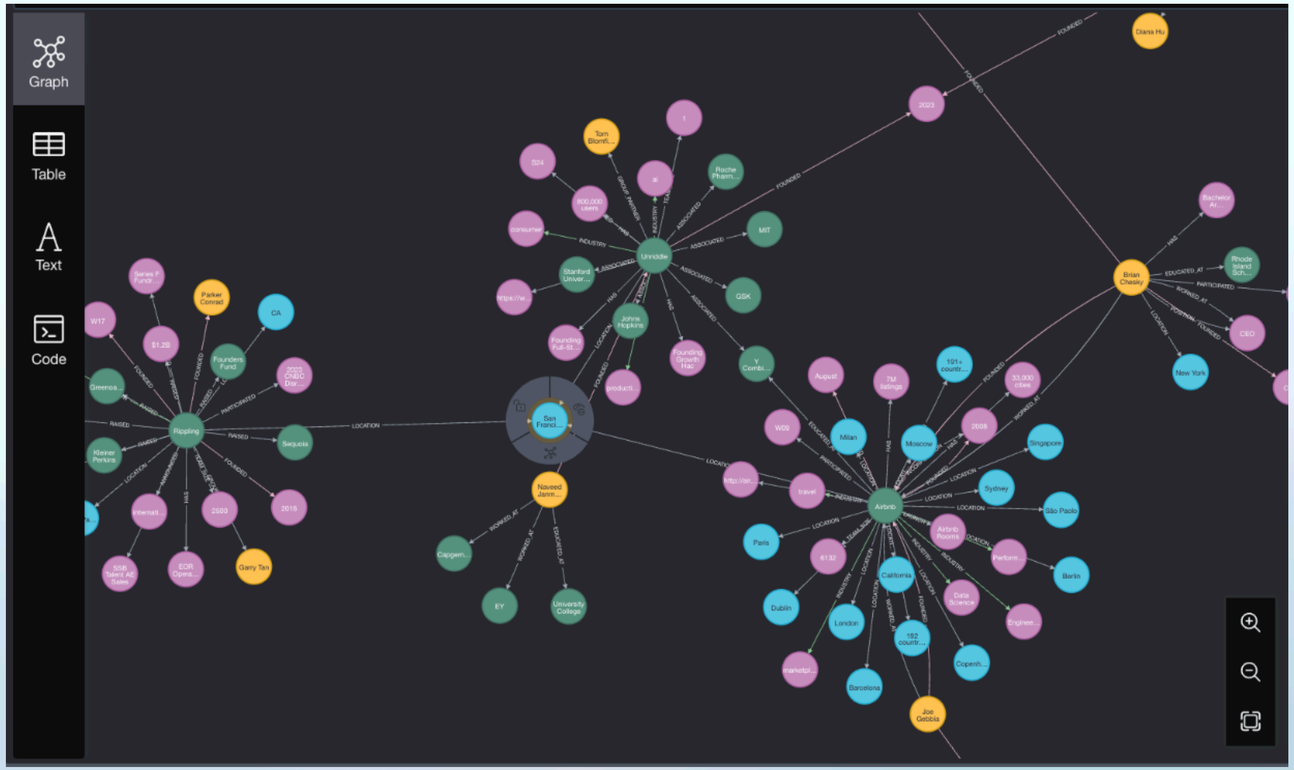
Global search

- Most existing algorithms struggle with population level queries such as
  - Who are the main competitors of company X?
  - List all people who have worked with person X
  - What does Paul graham generally write about?

- These queries require information from a large amount of chunks to be answered correctly
- More efficient ways of information representation are needed

GraphRAG

- Construct a graph on your dataset and query it using LLMs
- Natively supports population level queries, but construction is very expensive

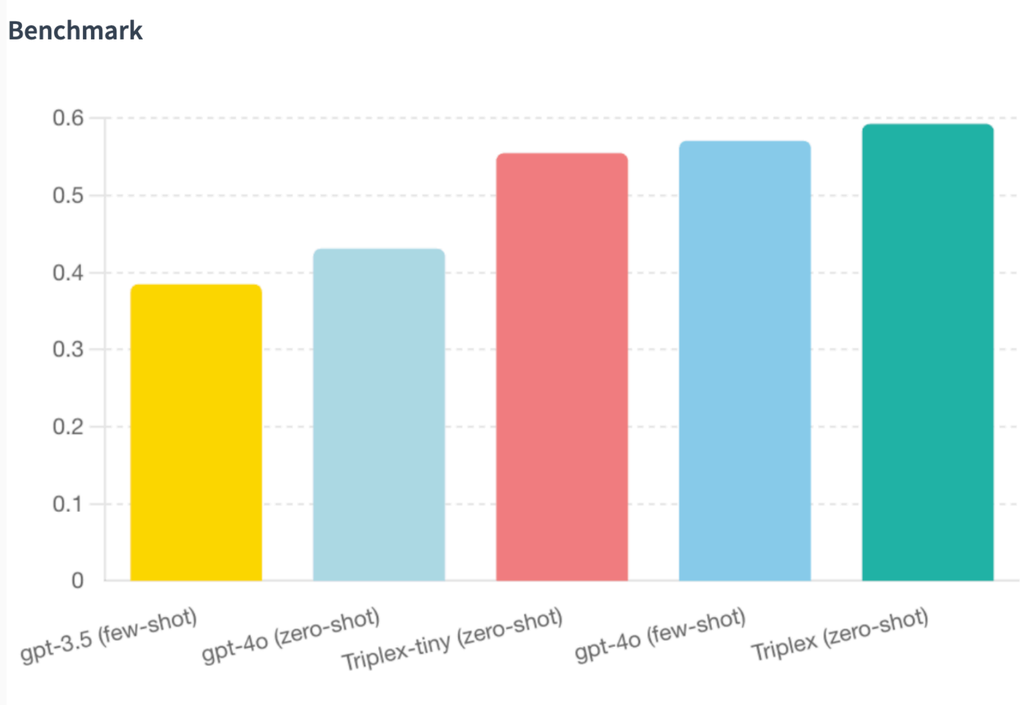


Triplex: SOTA LLM for Knowledge Graphs

Entity Types	PERSON, ORGANIZATION
RELATIONSHIPS	FOUNDED_BY
TEXT	Elon Musk is the founder of SpaceX

Triplex

```
{
  "entities_and_triples": [
    "[1], PERSON:Elon Musk",
    "[2], ORGANIZATION: SpaceX",
    "[2] FOUNDED_BY [1]"
  ]
}
```



- Better than GPT-4o and 97% cheaper
- Blog: https://www.sciphi.ai/blog/triplex
- Demo: https://kg.sciphi.ai

Example

Paul Graham Essays

- Q: How to create a successful startup according to Paul?

Creating a successful startup involves several critical components that must be addressed to ensure long-term success. Based on the provided data, the following key elements are essential:

User Feedback and Engagement

User feedback is essential for startup development. Startups must actively engage with users to understand their needs and challenges, which is crucial for creating successful products. This interaction leads to retaining offerings and fosters user loyalty and satisfaction, which are vital for long-term success [Data: Relationships (6edf42e4-f72c-4d5f-a440-4733661400c3)].

Experiential Learning and Advising

Experience gained from working with startups is crucial in enhancing the quality of advising provided to them. Advisors who have firsthand knowledge of the challenges faced by startups can offer more relevant and practical guidance. This relationship underscores the value of experiential learning in the advising process, as insights drawn from past experiences inform the strategies and recommendations given to new ventures [Data: Reports (18e345e3-545c-420b-9d2e-9a6e8837dda0)].

Technical Skills and Innovation

The interconnectedness of Lisp hacking and startups suggests that skills in programming can directly influence entrepreneurial success. Engaging with Lisp hacking not only improves technical capabilities but also fosters innovative thinking essential for startups. This interconnectedness implies that individuals or organizations proficient in Lisp may have a competitive edge in the startup landscape, potentially leading to more successful ventures [Data: Relationships (fcfc2022-477a-4be6-b957-f6720133b0d1)].

Funding and Financial Backing

Wealthy individuals play a crucial role in the startup ecosystem by providing essential funding that enables startups to grow and develop. Their investments are often pivotal for turning seeds to scale their operations, innovate, and bring new products to market. This financial backing can significantly influence the success trajectory of these startups [Data: Relationships (02a3d3e8-4c6b-49e3-8754-68b726dc7e31)]. Funding is identified as a crucial element for the establishment and growth of startups. It provides the necessary resources for operations and development, which are vital for any startup aiming to succeed in a competitive market [Data: Reports (b2a3c8e8-4c6b-49e3-8754-68b726dc7e31)].