



SCIENTIFIC PAPER DISCOVERY

Using Rag | arXiv

-Tharun Ravi

Project link :

<https://github.com/Tharun-Rv/Scientific-paper-discovery>

Scientific Paper Discovery using RAG

S.No	Contents
1	Introduction
2	Scope of the Project
3	System Architecture
4	Modules Explanation
5	Sample Outputs
6	Conclusion
7	Reference

1. Introduction

Welcome to Scientific Paper Discovery (RAG + Evaluation)! 📄

This tool lets users enter any research topic and fetch relevant, real-time scientific papers from arXiv. It uses a Retrieval-Augmented Generation (RAG) pipeline to answer questions and summarize content from the papers.

Built with:

- 📄 ArXiv API for live paper fetching
- 📄 RAG using Transformers
- 📄 Evaluation of summary & retrieval quality
- 📄📄 Gradio UI for a friendly interface

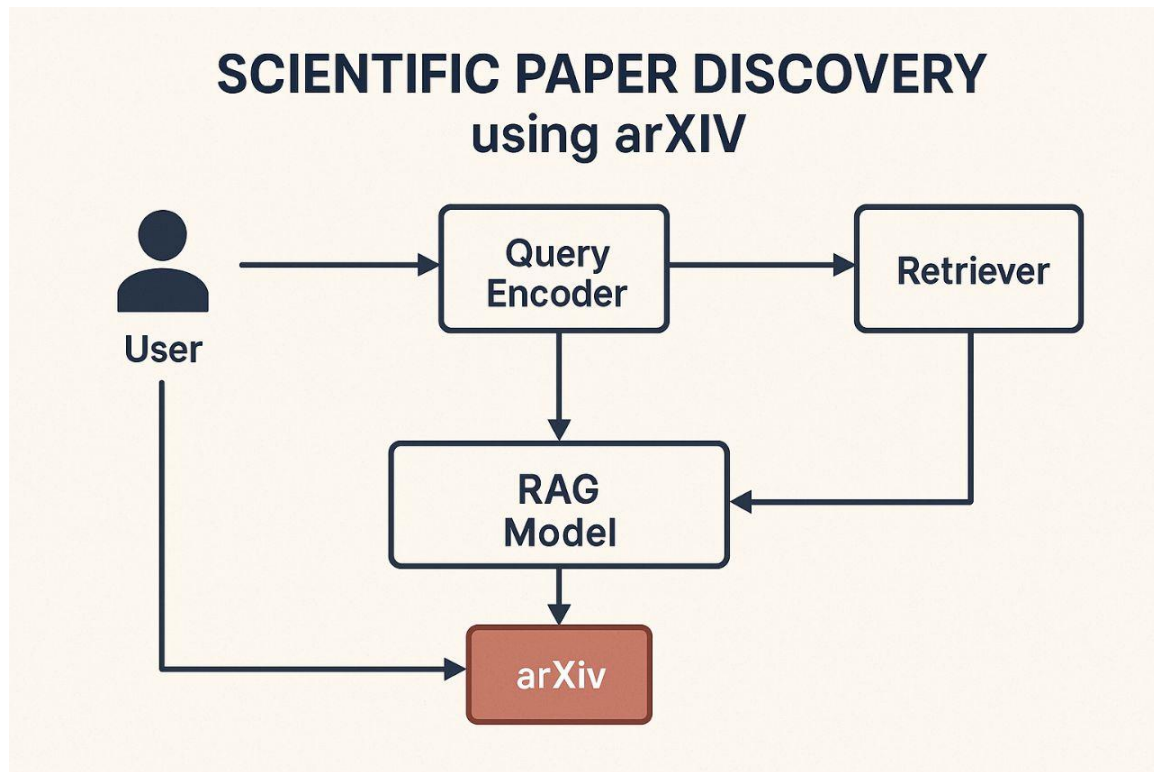
2. Scope of the Project

- Research Topic Exploration: Users enter a topic or keyword to fetch relevant arXiv papers.
- Real-Time Retrieval: Dynamically pulls the latest papers using the arXiv API.
- Summarization & Answer Generation: Summarizes insights from retrieved papers.
- Quality Evaluation: Evaluates summary relevance and accuracy.
- No-Code Interface: Gradio UI for easy use.
- Extensible: Can be expanded to other scholarly APIs or domain-specific filters.

This acts as a smart assistant for faster literature review and research discovery.

4. System Architecture Diagram

Below is a conceptual diagram of the system architecture used:



5. Modules Explanation

1. User Interface Module (Gradio): Accepts research topic input and displays results.
2. ArxivLoader: Fetches the most recent papers based on user query.
3. Embedding & Retrieval Module: Embeds abstracts using sentence-transformers, then finds relevant papers.
4. RAG Model Pipeline: Performs question answering and summarization from the retrieved data.
5. Evaluation Module: Uses similarity metrics to assess summary and retrieval quality.

6. Sample Outputs

Example: Entering query 'quantum computing' returns latest arXiv papers and generates a summary. It also answers custom questions like 'What are the latest methods in quantum computing?'

6.(a) Sample Output page [i]

The screenshot shows a web browser window with multiple tabs. The active tab is 'huggingface.co/spaces/TharunRavi/Scientific-paper-discovery'. The page title is 'Scientific Paper Discovery — RAG-Based QA & Evaluation'. Below the title, there is a prompt: 'Enter your research topic below to explore papers from arXiv and get summarized insights.' A text input field contains the text 'Retrieval-Augmented Generation'. Below the input field is a button labeled 'Search and Summarize'. Underneath the button, there are two sections: 'Retrieval Accuracy' and 'Summary Quality', each with a blue progress bar and a 'processing | 7.1/60.3s' status indicator. At the bottom of the page, there is a footer that says 'Use via API · Built with Gradio · Settings'.

6.(a) Sample Output page [ii]

The screenshot shows the same web browser window as the previous one, but now displaying the search results. The 'Search and Summarize' button is at the top. Below it, there are two sections, 'Paper 1' and 'Paper 2'.
Paper 1
Title: Generalized Baer and Generalized Quasi-Baer Rings of Skew Generalized Power Series
Abstract: Let R be a ring with identity, (S, \leq) an ordered monoid, $\omega: S \rightarrow \text{End}(R)$ a monoid homomorphism, and $A = R[\leftarrow[S, \omega]]$ the ring of skew generalized power series. The concepts of generalized Baer and generalized quasi-Baer rings are generalization of Baer and quasi-Baer rings, respectively. A ring R is called generalized right Baer (generalized right quasi-Baer) if for any non-empty subset S (right ideal I) of R , the right annihilator of S^n is generated by an idempotent for some positive integer n . Left cases may be defined analogously. A ring R is called generalized Baer (generalized quasi-Baer) if it is both generalized right and left Baer (generalized right and left quasi-Baer) ring. In this paper, we examine the behavior of a skew generalized power series ring over a generalized right Baer (generalized right quasi-Baer) ring and prove that, under specific conditions, the ring A is generalized right Baer (generalized right quasi-Baer) if and only if R is a generalized right Baer (generalized right quasi-Baer) ring.
Summarized Answer: R the ring of skew generalized power series
Link: <http://arxiv.org/abs/2405.03423v2>
Retrieval Accuracy: Low
Summary Quality: Good
Paper 2
Title: On generalized topological groups
Abstract: In this work, we will introduce the notion of generalized topological groups using generalized topological structure and generalized continuity defined by τ_A, C_S

6.(a) Sample Output page [iii]

Paper 5

Title: Generalized Lucas Numbers and Relations with Generalized Fibonacci Numbers

Abstract: In this paper, we present a new generalization of the Lucas numbers by matrix representation using Generalized Lucas Polynomials. We give some properties of this new generalization and some relations between the generalized order-k Lucas numbers and generalized order-k Fibonacci numbers. In addition, we obtain Binet formula and combinatorial representation for generalized order-k Lucas numbers by using properties of generalized Fibonacci numbers.

Summarized Answer: Generalized Lucas Polynomials

Link: <http://arxiv.org/abs/1111.2567v1>

Retrieval Accuracy: Low
Summary Quality: Good

Retrieval Accuracy

Multiple Papers Shown

Summary Quality

See Above

Use via API · Built with Gradio · Settings

7. Conclusion

This project delivers a powerful and user-friendly platform for discovering, retrieving, and understanding scientific papers from arXiv using Retrieval-Augmented Generation (RAG). With live fetching, summarization, and evaluation — it simplifies the literature review process and empowers researchers with quick, insightful overviews.

8. References

- arXiv API: <https://arxiv.org/help/api>
- Hugging Face Transformers: <https://huggingface.co/transformers>
- Gradio: <https://www.gradio.app>
- Sentence-Transformers: <https://www.sbert.net/>