PROJECT REPORT

Title of the Project

AI System to Automatically Review and Summarize Research Papers - Group 2

Submitted By

Name: Tanuj Kumar
Email: tanujkumar99995@gmail.com

Batch Time: 7:30PM to 8:30

Abstract

This project presents an AI-based system that automates the process of reviewing academic research papers. The system retrieves scholarly articles, extracts their content, generates AI-driven summaries, performs cross-paper comparison, and produces structured research outputs. The solution minimizes manual effort in literature review and enhances research productivity using modern natural language processing models and a graph-based workflow architecture.

---

1. Introduction

The rapid growth of scientific publications has made traditional manual literature review time-consuming and inefficient. This project addresses this challenge by developing an intelligent research automation system capable of retrieving, analyzing, summarizing, and comparing research papers automatically. The proposed system enables researchers to obtain meaningful insights from multiple academic sources with minimal human intervention.

---

2. Problem Statement

Manual analysis of multiple research papers is slow, error-prone, and difficult to scale. Researchers require an automated system that can efficiently extract relevant information, compare multiple studies, and generate structured research outputs.

---

3. Objectives

- Automate academic paper retrieval.
- Generate structured metadata for research papers.
- Perform AI-based summarization and cross-paper analysis.
- Implement a graph-based research workflow.
- Produce refined research reports.

---

4. System Architecture

The system follows a modular and graph-based architecture:

```
User Input
   ↓
Paper Retrieval → Metadata Generation → AI Analysis → Cross-Paper Comparison
   ↓
Workflow Graph Control → Review & Refinement → Final Output
```

Each module operates independently while being coordinated through a directed workflow graph.

---

5. Virtual Environment Setup

To ensure consistent execution and dependency management, a Python virtual environment was configured as follows:

```
D:
cd infosys_springboard
python -m venv venv
venv\Scripts\activate
pip install pandas scikit-learn nltk transformers torch sentencepiece
pip install pdfplumber pytesseract
pip freeze > requirements.txt
```

---

6. Technology Stack

Component	Technology
Programming Language	Python 3
Paper Retrieval/search	Semantic Scholar API
PDF Processing	PyMuPDF4LLM
AI Models	Hugging Face /Open Api Transformers
Validate and structure AI output	Pydantic
Workflow Control/ Architecture graph/ workflow engine	LangGraph / Grandalf
Data Storage	CSV
Environment Management	python-dotenv

---

7. Project Structure

infosys_springboard/

|

|-> app.py

|->config.py

|-> .env

|

| -> core/

| |-> researcher.py

| |-> pdf_loader.py

| |-> analyzer.py

| |-> comparator.py

| |-> writer.py

| |-> reviewer.py

| |-> metadata_writer.py

| |-> graph.py

|

|-> research_data.csv

|-> comparison_report.txt

---

8. Module Description

8.1 Paper Retrieval Module

Fetches relevant research papers based on the user's topic.

8.2 Metadata Generation Module

Stores structured information such as title, authors, year, abstract, summary, and source URL.

8.3 Analysis Module

Generates concise summaries using AI models.

8.4 Cross-Paper Comparison Module

Analyzes three selected research papers and identifies similarities, differences, and key insights.

8.5 Workflow Module

Controls the execution pipeline using a directed graph model.

8.6 Review & Refinement Module

Enhances the academic quality of the generated report.

---

9. Implementation Details

- The user provides a research topic.
- The system retrieves academic papers.
- Metadata is extracted and stored in **`research_data.csv`.**
- Three papers are selected for AI-based comparison.
- A comparison report is generated in **`comparison_report.txt`**.
- The workflow execution is represented through a graph structure.

---

10. Output Results

The system produces the following outputs:

- **research_data.csv** — structured research metadata.
- **comparison_report.txt** — AI-generated cross-paper analysis report.

---

11. Conclusion

The proposed AI system significantly simplifies the literature review process by automating paper retrieval, analysis, comparison, and reporting. The system demonstrates the practical integration of modern AI models into academic research workflows, offering a scalable and efficient research assistance solution.

---

12. Future Scope

- Integration of citation analysis.
- Full-length research paper drafting.
- Web-based user interface.
- Advanced topic modeling and trend detection.