Sreenidhi Institute of Science and Technology

Department of Computer Science and Engineering

Abstract seminar on

AI-Powered Repository Analysis and Visualization System using RAG and LLMs

Batch No: G1

MANOJ 22311A05HU

SAHIL 22311A05JB

MANI RAJA 2231105Z3

Internal Guide:

Dr. Preethi Jeevan

project Coordinator:

Padmini .K

Contents

Abstract

- This is an AI-powered automation system that analyzes GitHub repositories,
- extracts their structure, and generates natural-language explanations using LLMs and Retrieval-Augmented Generation (RAG).
- It helps developers quickly understand large codebases through intelligent summaries and mind maps.

Introduction

- Developers often struggle to understand large or poorly documented repositories.
- Manual exploration consumes time and effort.
- This automates this by combining code parsing, embeddings, and LLMs for smarter analysis.

Literature Overview

- Existing works focus on static code analysis, documentation generation, and AI-assisted summarization.
- However, they lack automated workflow integration and contextual retrieval.
- This bridges this gap using a RAG-based architecture integrated with workflow automation.

Existing System

- Manual code exploration and documentation.
- Time-consuming and error-prone.
- Lacks AI-based summarization or visualization features.
- No automation or intelligent code insight retrieval.

Proposed System

- Automates repository analysis and visualization.
- Uses LLMs with RAG for accurate context retrieval.
- Generates mind maps for repository structures.
- Streamlines developer onboarding and understanding.

System Architecture

- 1. Input: GitHub repository link.
- 2. Clone & parse repository using GitPython.
- 3. Chunk files and create embeddings.
- 4. Store in vector database (FAISS/Chroma).
- 5. Retrieve relevant code context via RAG.
- 6. Summarize using LLM API and visualize as a mind map.

Hardware & Software Requirements

- Hardware: 8GB RAM, 20GB Storage, quad-core CPU
- Software: Python 3.10+, MySQL, GitPython, LangChain, LLM API OR (LOCAL LLM OLLAMA), FAISS, n8n

- Local : A Power full GPU (8 to 32 gb vram)

Results & Output

- Successfully analyzed multiple GitHub repositories.
- Generated summaries and file structure mind maps.
- Provided contextual insights via RAG pipeline.
- Improved developer understanding and productivity.

Future Scope

- Add interactive visualization dashboard.
- Integrate with GitHub Actions for continuous updates.
- Expand to multiple LLMs for code optimization and review.
- Include chat-based interaction with repositories.

Conclusion

- This demonstrates how AI and workflow automation can revolutionize repository analysis.
- It enhances code comprehension, reduces manual effort, and promotes intelligent software engineering.

References

- OpenAI API Documentation
- LangChain Framework Docs
- GitPython Documentation
- Hugging Face Transformers Docs
- MySQL Reference Manual
- N8n manual