

# **(REMOTE) Leveraging Knowledge Graph-Enhanced RAG and LLMs for Historical Archival Analysis: A Case Study of State of Maryland's Legacy of Slavery Dataset Collections**

*Friday, 21 March 2025 09:20 (20 minutes)*

Integrating Artificial Intelligence in digital humanities has created unprecedented opportunities for analyzing historical archives. Building upon established work with Learning-as-a-Service solutions for Maryland State Archives' Legacy of Slavery collections, specifically the Domestic Traffic Advertisements dataset, this research proposes an innovative approach combining Knowledge Graph-based Retrieval Augmented Generation (KG-RAG) with Large Language Models to analyze three strategically chosen collections: Certificates of Freedom, Domestic Traffic Advertisements, and Manumissions. These collections were selected for their intricate historical relationships - Certificates of Freedom documents provide legal proof of an enslaved individual's free status, often referencing prior Manumission records issued, while Domestic Traffic Advertisements offer crucial contextual information about slave trading patterns that frequently preceded manumission or freedom certification. The project introduces a novel architecture that enhances traditional RAG systems by incorporating knowledge graphs to capture complex relationships and temporal-spatial connections between these historically linked dataset collections. Unlike traditional RAG systems, this knowledge graph-enhanced approach enables natural language interactions for archive patrons and researchers, allowing them to explore complex historical narratives through intuitive conversations rather than traditional database queries. The system employs a three-layer approach: a knowledge graph layer mapping relationships between entities across collections using Neo4j, an RAG layer augmented with knowledge graph embeddings for contextual retrieval, and an LLM layer for natural language interaction and insight generation. This is an approach to transform how users can discover connections across collections, trace individual histories, and uncover previously hidden relationships in the archives. This natural language interface improves accessibility by eliminating the need for specialized database knowledge or understanding of archival organization systems. The research advances AI-enabled scientific workflows through specialized prompt engineering patterns for cross-collection analysis and custom embedding techniques for historical document representation. This approach improves the trustworthiness of AI responses by grounding them in verified historical relationships while enhancing the accuracy of cross-collection insights by leveraging the knowledge graph's ability to capture complex historical narratives and relationships. The system demonstrates how AI can democratize access to complex historical archives while maintaining the integrity and context of sensitive cultural materials. This research contributes to ISGC 2025 Track 10's focus on AI-enabled scientific workflows and novel approaches in scientific applications adopting machine learning techniques while advancing the state-of-the-art in knowledge graph-enhanced RAG systems for digital humanities research.

**Primary author:** GNANASEKARAN, Rajesh Kumar (the University of Maryland)

**Presenter:** GNANASEKARAN, Rajesh Kumar (the University of Maryland)

**Session Classification:** Artificial Intelligence (AI) - III

**Track Classification:** Track 10: Artificial Intelligence (AI)