Contribution ID: **138** Type: **Oral Presentation**

Towards Generative Archival Intelligence: Piloting Agentic AI Workflows for Generative Feature Discovery and Dynamic Knowledge Graph Construction from Historic Document Collections

Friday, 20 March 2026 09:22 (22 minutes)

This study investigates how LLMs can act as generative reasoning engines to interpret and restructure complex archival document collections. Building upon earlier work with the President's Personal File (PPF 9: Gifts) from the Franklin D. Roosevelt Presidential Library, the research explores how an LLM can infer relationships, sequences, and contextual features from textual and descriptive inputs derived from scanned historical pages. The project seeks to understand whether such models can perform higher-order interpretation, recognizing the tacit organization, document groupings, and semantic cues embedded in archival materials. These archival materials are composed of thousands of sequentially arranged pages, often grouped into file folders, and contain a mixture of printed and handwritten textual elements that signal relationships not captured in existing cataloging systems.

The experimental framework employs a multi-stage agentic workflow in which models analyze scanned documents to infer boundaries, sequences, and associations among individual pages, and the LLM serves as both interpreter and orchestrator. The AI workflow process is guided by a baseline knowledge graph that encodes known entities, relationships, and attributes provided by the archival team. With this knowledge, the agentic AI workflow process autonomously analyzes the extracted text and descriptive metadata to hypothesize how documents are bound or sequenced within each folder. As it reasons through patterns in dates, names, annotations, and file markings, the model proposes new candidate features adding to the novel entities, relationships, or attributes that enrich the evolving graph structure. These features are iteratively proposed, validated, and refined through human review, forming a continuous cycle of automated discovery and curatorial oversight.

Over multiple cycles, the knowledge graph becomes a living, self-updating representation of the archival corpus, capturing both its explicit and inferred organizational logic. By analyzing how the system interprets sequentiality, detects tacit groupings, and generates candidate features for approval, the research seeks to evaluate generative AI's capacity for understanding archival logic beyond explicit text recognition. A prototype conversational interface demonstrates how the evolving graphs can be queried through text and image inputs, offering a multimodal means of scholarly engagement with historical data.

This study contributes to the fields of digital humanities and computational archival science by proposing a new generative approach to archival interpretation, one that couples autonomous feature discovery with human interpretive judgment. The resulting framework offers a pathway toward dynamic, explainable, and continuously learning archival knowledge systems that reimagine how context and meaning can be reconstructed from complex historical document collections.

Primary authors: GNANASEKARAN, Rajesh Kumar (the University of Maryland); Prof. MARCIANO, Richard (University of Maryland, College Park)

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

Session Classification: Artificial Intelligence (AI) - IV

Track Classification: Track 10: Artificial Intelligence (AI)