Abstract

The exponential growth of digital knowledge has led to vast amounts of multimodal data, including research papers, news articles, images, and audio. Traditional digital libraries, which rely primarily on keyword-based search, struggle to capture semantic meaning and the complex relationships between people, organizations, events, and concepts.

This project, Cognitive Digital Library with a Multimodal Knowledge Graph (CDL-MMKG), addresses these challenges by integrating MongoDB's scalable NoSQL architecture with natural language processing (NLP), vector search, and graph-style modeling. The system ingests large-scale datasets (e.g., Wikipedia, arXiv abstracts, and news archives), extracts entities and events using NLP, and generates embeddings that are indexed with MongoDB Atlas Vector Search for semantic retrieval. Extracted relations are stored as a multimodal knowledge graph, enabling advanced queries that combine keyword, semantic, temporal, and relational filters.

A React-based dashboard provides intuitive exploration through hybrid search and graph visualizations. Expected outcomes include a scalable backend, semantic retrieval capabilities, and structured knowledge graph exploration. The project lays the foundation for next-generation digital libraries, supporting intelligent research assistance, knowledge discovery, and AI-driven question answering.

Introduction

Digital libraries have transformed the way knowledge is stored and accessed, evolving from traditional catalog systems to large-scale digital repositories that host books, research articles, news archives, images, and audio-visual content. However, most existing systems rely on keyword-based search mechanisms, which often fail to capture the true semantic meaning of user queries and cannot effectively handle multimodal data. For instance, while a user may search for "AI conferences in Europe after 2020," conventional systems struggle to connect relevant documents, event records, figures, and participants in a meaningful way.

The increasing diversity and scale of data demand a more intelligent and scalable approach to knowledge management. This project, Cognitive Digital Library with a Multimodal Knowledge Graph (CDL-MMKG), proposes a solution that leverages MongoDB's flexible NoSQL database alongside modern natural language processing (NLP), embeddings, and vector search. By extracting entities, events, and relationships from heterogeneous sources and linking them in a knowledge graph, the system enables semantic and relational queries that go beyond keyword matching.

With the addition of vector search for semantic similarity, graph-style relations for knowledge connectivity, and a React-based dashboard for visualization, CDL-MMKG transforms a static collection of documents into a dynamic, intelligent knowledge platform. This approach not only enhances information retrieval but also supports advanced applications such as research assistance, digital archiving, and AI-driven question answering.

Problem Statement

Traditional digital libraries face significant challenges:
1. Lack of semantic understanding: keyword-only search cannot capture deep contextual meanings.
2. Limited multimodal support: metadata is often text-only; images, figures, and audio lack semantic linkage.
3. Scalability issues: relational systems struggle with large unstructured datasets.
4. Poor knowledge connectivity: entities, events, and documents are not explicitly linked.

This project aims to overcome these challenges by building a scalable MongoDB-based system with vector search, NLP pipelines, and knowledge graph modeling.

Objectives

The key objectives are:
1. Ingest and normalize text, images, and audio metadata into MongoDB.
2. Apply NLP and embedding models to extract entities, events, and relations.
3. Store embeddings and enable semantic search using MongoDB Atlas Vector Search.
4. Construct a multimodal knowledge graph linking people, organizations, topics, events, and media.
5. Develop a user-facing dashboard to support hybrid semantic + keyword queries and graph exploration.
6. Demonstrate scalability, retrieval quality, and real-time semantic querying.

Methodology

The system will be developed in the following phases:
1. **Data Ingestion**: Wikipedia, arXiv abstracts, and curated news archives as initial corpora.
2. **NLP Pipeline**: Named Entity Recognition (NER), event extraction, entity resolution, and linking.
3. **Embedding**: Sentence-transformers for text, optional CLIP embeddings for images.
4. **Storage & Indexing**: MongoDB with Atlas Vector Search, GridFS for media, and

sharding for scale.
5. **Query API**: FastAPI service for semantic retrieval, hybrid queries, and relation traversal.
6. **Frontend**: React-based dashboard with search bar, filters, timeline, and interactive graph visualization.
7. **Evaluation:** Retrieval quality (Precision@10), entity linking accuracy, ingestion throughput, and query latency.

Expected Outcomes

The project is expected to deliver:
- A functional multimodal digital library backend powered by MongoDB.
- Vector search–enabled semantic retrieval over documents, entities, and events.
- A knowledge graph that connects multimodal entities with explainable provenance.
- A React-based dashboard for hybrid search and graph exploration.
- Performance evaluation report (throughput, query latency, retrieval quality).
- Foundation for future extensions: cross-modal search, temporal reasoning, and RAG-based Q&A.

Tools and Technologies

1. **Database** – MongoDB Atlas (Vector Search, GridFS, Sharding)
2. **Backend** – Python (FastAPI), Celery for jobs
3. **NLP/ML** – Hugging Face Transformers, Sentence-Transformers, spaCy
4. **Frontend** – React + Tailwind CSS + D3/vis.js
5. **Deployment** – Docker Compose for local, MongoDB Atlas for cloud
6. **Version Control** – Git/GitHub
7. **Optional** – Terraform/Atlas CLI for cluster automation