

2. Automated Meta Data Generation

Objective

This project aims to develop an automated metadata generation system. This system will enhance document discoverability, classification, and subsequent analysis by producing scalable, consistent, and semantically rich metadata.

Key Areas of Focus

The development of this system will involve:

- Automating metadata generation: The system must automatically generate metadata for various unstructured document types.
- Content extraction: Capabilities for extracting text content from diverse formats like PDF, DOCX, and TXT, incorporating optical character recognition (OCR) where necessary, are essential.
- Semantic content identification: The system needs to identify and leverage the most meaningful sections of a document to inform metadata generation.
- Structured metadata creation: The system will be responsible for generating structured metadata outputs.
- User interface development: An intuitive web interface for document upload and metadata viewing will be a core component.
- System deployment: The entire system will be deployed to ensure public accessibility and ease of use.