

**Title:** Smart Thesis Repository with Integrated Plagiarism Detection and AI-Based Metadata Extraction

**Abstract:** This paper presents the development of a smart, web-based thesis repository that combines automatic plagiarism detection and AI-powered keyword extraction to ensure originality and enhance discoverability of academic papers. The system leverages modern web technologies, integrates Python-based NLP modules, and demonstrates significant improvement over traditional institutional repositories. Results show effective plagiarism detection and relevant keyword extraction across multiple tested documents.

**Keywords:** Thesis repository, Plagiarism detection, Keyword extraction, NLP, Institutional repository.

**1. Introduction** Academic institutions require secure, reliable, and intelligent systems to manage digital submissions. Existing repositories often lack integrated plagiarism detection and automated metadata generation. This paper proposes a smart repository that automates these tasks.

**2. Related Work** Systems like DSpace provide digital archiving but rely on manual checks for plagiarism. Several studies propose separate plagiarism tools, but integrated solutions are rare. This research combines both functionalities.

**3. Methodology** The system uses a MERN stack: React for frontend, Node.js/Express for backend, MongoDB Atlas for cloud storage, and Python scripts for plagiarism detection and NLP-based keyword extraction. Multer handles file uploads, and APIs connect services.

**4. Results and Discussion** Multiple documents were tested for plagiarism and keyword relevance. The integrated AI module achieved consistent extraction of key terms. The system flagged duplicate submissions with >70% similarity.

**5. Conclusion** The smart repository ensures academic integrity and streamlines submission workflows. Future enhancements include advanced semantic plagiarism algorithms, multi-language support, and extended user analytics.

**References** 1. Smith et al., "Plagiarism Detection Algorithms", Journal of Academic Ethics. 2. DSpace Documentation. [Online] Available: <https://dspace.lyrasis.org/> 3. Bird et al., "NLTK: Natural Language Toolkit", ACM. 4. MongoDB Atlas Documentation. [Online] Available: <https://www.mongodb.com/atlas>