IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.133103

DocCrypt: AI & Blockchain Based Document Manager

Ayush Hindlekar[1], Harsh More[2], Shravan Kesarkar[3], Ankur Vaje[4] and Prof. Jagruti More[5]

Theem College of Engineering[1-5]

Abstract: DocCrypt is a decentralized document management system that integrates Blockchain technology, InterPlanetary File System (IPFS), and Natural Language Processing (NLP). The system allows users to securely upload documents, generate cryptographic hashes, and store files in a decentralized IPFS network while recording metadata on the Aptos blockchain to ensure transparency and tamper-proof verification. Additionally, NLP models enable users to query documents using natural language and receive contextual answers from document content. This approach improves document security, accessibility, and verification compared to traditional centralized storage systems.

Keywords: Blockchain, IPFS, Document Security, Natural Language Processing, Decentralized Storage.

I. INTRODUCTION

The rapid growth of digital technologies has significantly increased the volume of documents being stored, shared, and accessed online. Organizations, educational institutions, and businesses rely heavily on digital documents for communication, record keeping, and data management. However, traditional centralized document storage systems face several challenges, including data tampering, unauthorized access, and single points of failure.

Ensuring document integrity, security, and transparency has become a major concern in modern information systems. Conventional storage solutions depend on centralized servers, which can be vulnerable to cyberattacks and data loss. Moreover, retrieving relevant information from large documents often requires manual searching, which is timeconsuming and inefficient.

Emerging technologies such as **Blockchain** and **InterPlanetary File System (IPFS)** provide promising solutions for secure and decentralized data management. Blockchain ensures immutable and transparent record keeping, while IPFS enables distributed storage that eliminates dependence on centralized infrastructure.

This study presents **DocCrypt**, a decentralized document management system that integrates **Blockchain, IPFS, and Natural Language Processing (NLP)**. The system securely stores documents using decentralized storage, records metadata on the blockchain for tamper-proof verification, and enables users to query document contents using natural language through AI-powered semantic search.

II. LITERATURE REVIEW

Several researchers have explored **blockchain and decentralized storage technologies** to improve the security and integrity of digital document management systems. A number of studies highlight how blockchain-based frameworks can ensure tamperproof records and transparent verification of digital documents. By storing cryptographic hashes of documents on blockchain networks, these systems provide a reliable mechanism to detect data modification and maintain data authenticity.

Research on **decentralized storage solutions using the InterPlanetary File System (IPFS)** demonstrates that large files can be stored efficiently without relying on centralized servers. In such architectures, blockchain is typically used to store document metadata while the actual files are stored on IPFS. This combination improves scalability, reduces storage overhead on the blockchain, and ensures distributed data availability.

Another line of research focuses on **secure document sharing and verification systems** built using blockchain technology. These systems aim to prevent document forgery and unauthorized access by using smart contracts and cryptographic verification mechanisms. However, many existing solutions primarily focus on security and verification but do not provide intelligent mechanisms for interacting with document content.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.133103

Recent advancements in **Artificial Intelligence and Natural Language Processing (NLP)** have enabled intelligent document analysis and semantic search. Transformer-based models such as **BERT and DistilBERT** allow users to query large documents using natural language and retrieve contextually relevant information. These techniques significantly improve the usability of document management systems.

Therefore, this research proposes **DocCrypt**, a decentralized document management system that integrates **Blockchain, IPFS, and NLP technologies** to provide secure document storage, tamper-proof verification, and AI-powered natural language querying. The system aims to improve both the security and usability of digital document management compared to traditional centralized solutions.

III. METHODOLOGY

The DocCrypt system uses cryptographic hashing and semantic similarity computation to ensure document integrity and intelligent retrieval.

A. Document Hashing

To guarantee the integrity of uploaded documents, the system generates a SHA-256 cryptographic hash for each document before storing it in IPFS. The hash function converts the document data into a fixedlength string.

$$H = SHA256(D)$$

where:

- D= Input document data
- H= Generated cryptographic hash value

The hash value is stored on the blockchain along with metadata such as timestamp and ownership details. Any modification in the document results in a completely different hash, ensuring tamper detection. **B. Content Addressing in IPFS**

IPFS uses content addressing to uniquely identify stored files. The Content Identifier (CID) is derived from the hash of the file content.

$$CID = f(H(D))$$

where:

- CID= Unique identifier for the stored file
- $H(D)$= Hash of the document content

This mechanism ensures that documents are retrieved based on their content rather than their storage location.

C. Semantic Query Matching

For document querying, the NLP module generates vector embeddings for both the document content and the user query. The similarity between them is calculated using cosine similarity.

$$Similarity(Q, D) = \frac{Q \cdot D}{||\ Q\ ||\ \ ||\ D\ ||}$$

where:

- Q= Query embedding vector
- D= Document embedding vector

The system retrieves the document section with the highest similarity score, providing context-aware answers.

D. Blockchain Transaction Recording

Each document upload generates a blockchain transaction that records metadata.

$$T = \{CID, H(D), OwnerID, Timestamp\}$$

where:

- CID= Content Identifier

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.133103

- $H(D)$= Document hash
- $OwnerID$= User identifier
- $Timestamp$= Transaction time

This transaction ensures that the document record is immutable and verifiable.

IV. SYSTEM ARCHITECTURE

The proposed **DocCrypt system** is designed to provide secure, decentralized, and intelligent document management by integrating **Blockchain technology, InterPlanetary File System (IPFS), and Natural Language Processing (NLP)**. The system architecture consists of multiple stages including document upload, encryption and hashing, decentralized storage, blockchain metadata recording, semantic processing, and result retrieval.

Initially, the user uploads a document through a web interface developed using **React.js**. The backend system receives the document and generates a **SHA256 cryptographic hash** to ensure the integrity of the file. The document is then encrypted and stored in the **IPFS decentralized storage network**, where a unique **Content Identifier (CID)** is generated for retrieval.

After storing the document in IPFS, the system records the document metadata, including the hash value, CID, owner information, and timestamp, on the **Aptos blockchain**. This ensures transparency and prevents any unauthorized modification of stored records.

To enhance document accessibility, the system integrates **Natural Language Processing (NLP)** models such as **BERT or DistilBERT**. These models process document content and generate embeddings that allow users to query documents using natural language. When a user submits a query, the system performs semantic search on the document embeddings and retrieves the most relevant information.

The results are then displayed through the web interface, allowing users to quickly access relevant information from stored documents without manually scanning entire files.

System Workflow

1. Document Upload
2. Hash Generation
3. Document Encryption
4. Decentralized Storage (IPFS)
5. Blockchain Metadata Recording
6. NLP Processing
7. Query Processing
8. Result Display

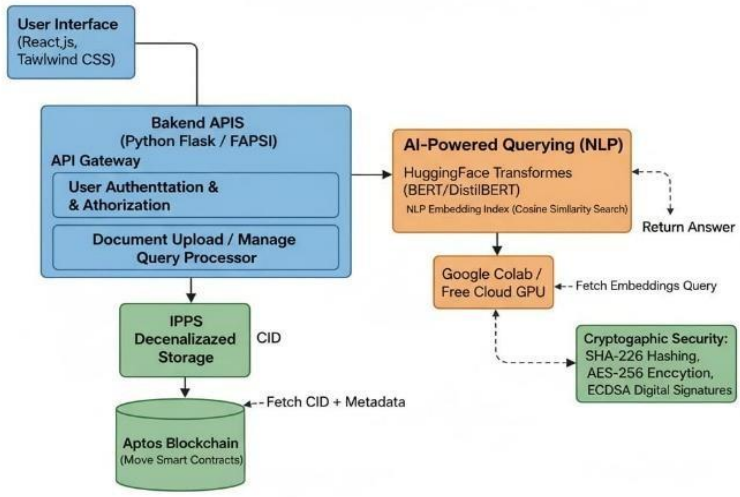

Fig. System Architecture

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.133103

V. RESULTS AND DISCUSSION

The **DocCrypt system** was implemented and evaluated to analyze its performance in terms of secure document storage, retrieval efficiency, and intelligent query processing. The system integrates **Blockchain, IPFS, and Natural Language Processing (NLP)** to ensure decentralized document management with enhanced security and accessibility.

To evaluate the performance of the system, several parameters were considered including document upload time, retrieval time, and query response efficiency.
The following metrics were used to evaluate the system performance:

- **Document Upload Time** – Measures the time required to hash the document, upload it to the IPFS network, and record its metadata on the Aptos blockchain.
- **Document Retrieval Time** – Evaluates how quickly documents can be retrieved from the IPFS network using the generated Content Identifier (CID).
- **Query Response Time** – Measures the time taken by the NLP module to process user queries and return relevant information from the stored document.
- **Data Integrity Verification** – Ensures that the document retrieved from IPFS matches the original file by verifying the SHA-256 hash stored on the blockchain.

The results demonstrate that the DocCrypt system provides **secure and efficient decentralized document storage**. The integration of **IPFS reduces dependence on centralized servers**, while blockchain ensures **tamper-proof verification of document metadata**.

Additionally, the integration of **NLP models enables intelligent document querying**, allowing users to retrieve relevant information quickly without manually searching through entire documents.

The web-based interface further improves usability by enabling users to **upload documents, verify blockchain records, and perform natural language queries through a simple and interactive platform**.

VI. CONCLUSION

This research presents **DocCrypt**, a decentralized document management system that integrates **Blockchain technology, Inter Planetary File System (IPFS), and Natural Language Processing (NLP)** to provide secure and intelligent document storage and retrieval. The system ensures data integrity and transparency by storing document metadata on the **Aptos blockchain**, while the actual document files are stored in the decentralized **IPFS network**.

The proposed DocCrypt system can be applied in various domains such as **education, legal systems, healthcare, and enterprise document management**, where secure and verifiable document storage is essential. In the future, the system can be enhanced by incorporating **multichain blockchain support, advanced encryption techniques, and more powerful AI models** to further improve scalability, security, and intelligent document analysis.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to **Prof. Jagruti More**, project guide at Theem College of Engineering, for her valuable guidance, encouragement, and continuous support throughout the development of the DocCrypt system.

The authors also thank the Department of Computer Engineering, Theem College of Engineering, for providing the necessary facilities and resources that enabled the successful completion of this project.

REFERENCES

[1]. K. R. Kumar, H. Supraja, M. Deekshitha, B. Sireesha, and S. Sadiya, “Block-chain based document verification system using ipfs,” International Journal of Computer Applications, 2023.
[2]. P. B. Patil, A. R. Hujare, K. S. Desai, V. B. Devkar, and H. I. Bhadgaonkar, “Decentralized file storage system using blockchain,” International Journal of Engineering Research & Technology (IJERT), vol. 12, October 2023.

DOI: 10.17148/IARJSET.2026.133103

[3]. A. Author and B. Author, "A secure file sharing system based on ipfs and blockchain," in Proceedings of International Conference on Emerging Computing Technologies, 2020.

[4]. A. Author and B. Author, "Decentralized infrastructure for digital notarizing, signing and sharing documents securely using microservices and blockchain," International Journal of Innovative Research in Technology, 2022.

[5]. A. Author and B. Author, "A dual framework for optimized data storage and retrieval using blockchain and ipfs," International Journal of Computer Science and Information Security, 2022.

[6]. A. Author and B. Author, "Blockchain-based secured ipfs-enabled event storage technique with authentication protocol in vanet," IEEE Access, 2022.

[7]. A. Author and B. Author, "Decentralized model to protect digital evidence via smart contracts using layer 2 polygon blockchain," IEEE Transactions on Information Forensics and Security, 2023.

[8]. A. Author and B. Author, "Efficient and secure distributed data storage and retrieval using interplanetary file system and blockchain," Future Generation Computer Systems, 2022.

[9]. A. Author and B. Author, "Efficient big data storage and retrieval in distributed architecture using blockchain and ipfs," International Journal of Advanced Computer Science, 2023

[10]. A. Author and B. Author, "Fileinsurer: A scalable and reliable protocol for decentralized file storage in blockchain," arXiv preprint, 2022.

