

Document Management System using Blockchain

Prof. Renuka Vaidya¹, Ms. Sanskriti Punde², Mr. Kartikey Yadav³,

Mr. Chakradhar Ghute⁴, Ms. Namrata Shinde⁵

Assistant Professor, Department of Information Technology¹

Students, Department of Information Technology^{2,3,4,5}

Sinhgad College of Engineering, Pune, India

Abstract: Under mission Digital India we are growing exponentially in the field of technology. As a developing country, India has contributed a lot to the field of technology. And still, in many government offices and corporations, the problem that we are still facing is the way documents are managed. After doing some research we found out that the recent steps taken by the CBSE board to use blockchain technology for storing 12th results were one of the revolutionary and impactful activities. We then started to explore more on this idea of managing documents. We wanted to tackle the biggest problem which is the security and storage of the documents and we thought that moving documents on the web can really help us in achieving our goal. The solution DOCMAN will try to solve most of the problems by making the whole process online which will include all the features that a process of the document takes from creating the document itself to the whole life cycle of the document starting from tracking to user access level to generating reports everything is covered and for the main security aspect we are integrating blockchain to make documents are more secure and reliable.

Keywords: Blockchain, Document, Storage, Security, Web Development, Software Engineering

I. INTRODUCTION

The whole thought process behind 'Docman' is to do something that could have an impact on the current document management system. The idea is to develop a web-based document management system that will help in managing, tracking, and securely storing and sharing documents. The system should properly have access levels and should give flexibility for each department to change the structure in which documents are sent or received. After doing our research we started to understand the whole problem and we wanted to do something but to go to the ground level we wanted to understand the key problems that are the biggest obstacle that can be on our way to solve the problem. We identified 4 major issues with the current system: The first and foremost issue that we noticed was that storage of documents was not an easy task, people would just pile these documents in some cabinet and this process goes on and on, but isn't there a way of like discarding the documents, which means that can't we just throw away these documents if they are years old, but we came to know that many files are needed or can be needed in the future. So, throwing them away is not an option so how can we have to think about how we can solve that problem The security of the documents - how can we be sure that the document is not tampered with, there are serious punishments for such an act but what if that happens, if we are talking about the real world there are cases when such incidents have occurred and tampering of documents is very much possible What about the cases when files are misplaced, the problem here is the tracking of files, we are not able to keep track of the files and that is possible because when there are such a large number of files it starts to become almost impossible to keep track of files and important documents especially when they are older The processing of the documents was taking a lot of time, when there are so many files to go through one by one and that process itself takes time, every document is properly checked and if everything goes well it is approved or rejected or forwarded to the bigger authority, it can be possible that the whole process will take time If you see the above problems that we have mentioned you will see a pattern and that is most of the problems are caused because of other problems, it's like a chain that goes on and on and leads to a lot of minor and major problem.

II. BACKGROUND AND LITERATURE REVIEW

Research was done on a lot of papers to find out if the solution was already present but none of the papers were doing the same exact thing as we had planned. The paper Blockchain-Based Automatic Tracking and Extracting Construction Document for Claim and Dispute Support that was being evaluated appears to have limitations in its access level and user rights management capabilities. While it offers basic tracking and monitoring features, it may not be suitable for managing complex document workflows within larger organizations or teams. The system's limited access controls could pose security risks, as unauthorized users may be able to access or modify sensitive documents. Furthermore, the lack of granular access controls could hinder collaboration, as team members may not have the necessary permissions to work on relevant documents. For organizations that require robust access controls and permissions management, the system may not be sufficient. It may be necessary to explore alternative software solutions that offer more advanced access level and user rights management capabilities. This is especially important for organizations with complex document workflows and collaboration requirements. By addressing these limitations, organizations can ensure that their document management system is secure, efficient, and capable of meeting their unique needs.

- The second paper has the following limitations: The Highly private blockchain-based management system for digital COVID-19 certificates has a centralized architecture, while potentially useful for certain applications, could limit its overall usability across a wider range of scenarios. By transitioning to a distributed system, organizations can ensure that their document management processes are more secure, efficient, and scalable to their unique needs. While transitioning to a distributed system can require additional investment and effort, the benefits of improved accessibility, scalability, and security can make it a worthwhile endeavor for organizations that require more robust document management capabilities.
- The third paper DOC-BLOCK: A Blockchain Based Authentication System for Digital Documents has focused on document verification, while document verification is an important aspect of document management, it is not the only consideration when it comes to managing complex document workflows. Document management systems that only focus on document verification may not be sufficient for organizations that require more granular control over document workflows. By exploring document management systems that offer more advanced workflow management capabilities, organizations can ensure that their document management processes are efficient, secure, and tailored to their specific needs.

The fourth research paper under review notes that the system in question is centralized only for the judicial embargoes process. This indicates that the system may have been designed specifically to handle this particular process, rather than being a more general-purpose document management system. The centralized nature of the system for the judicial embargoes process may have been chosen to provide greater control and security over sensitive documents but the potential limitations of a centralized system, such as single points of failure, and explore alternative solutions that may offer greater resilience and flexibility.

The fifth research paper under review highlights that the document management system described in the research paper focuses only on construction documents. While this may be suitable for some organizations in the construction industry, it may not be sufficient for managing other types of documents or workflows. To address these limitations, it may be necessary to explore more flexible and customizable document management systems that can be tailored to the specific needs of different organizations and industries.

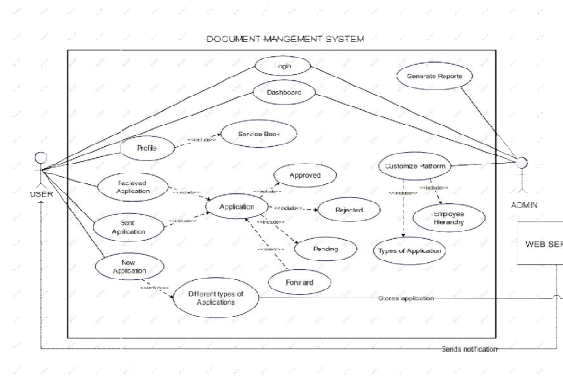
III. METHODOLOGY

The software is mainly composed of two very different technologies - blockchain and web development. Web scraping is the process of extracting data from a webpage, while web development involves the process of building a functional dynamic web application.

The book data that is to be displayed on Book Finder, is scraped from Amazon, which is known to host millions of books and was originally an online book store itself. The technologies involved are Python and Scrapy bundled with Pandas. The scraping module can be hosted on persistent computational service clouds like Digital Ocean or AWS. These scripts may run for about 2-4 weeks depending upon the size of the underlying data. The data obtained during this process is on the

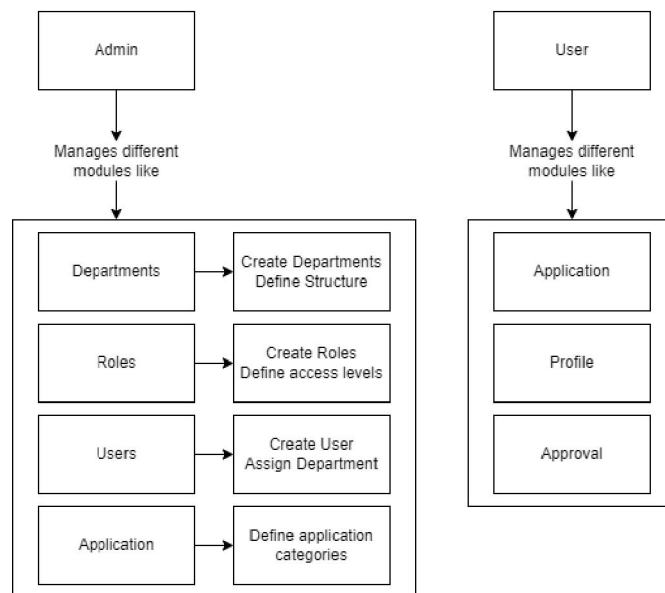
scale of 600,000-700,000 unique books. The scraping script involves recursively traversing through all desired pages on Amazon and finding required data through the use of *XPath*. Some of the *paths* used in the script are as follows:
 For web application development, the MERN stack is being used along with Tailwind CSS paired with MongoDB as a database. The user interface is being developed on Figma. The backend consists of different modules like User, Books, Scraper, Recommendation System, Sorting, Searching, Filtering, etc.
 The recommendation system will be used to populate a recommendations list for users where books that a user might like to read next are given. These books are devised from a rule-based algorithm that uses data about a user like which genres they prefer and which books they have already read. Based on such data, the rule-based algorithm will find books from similar genres, authors, and many such parameters.

IV. DIAGRAM SYSTEM ARCHITECTURE



V. SYSTEM ARCHITECTURE

Document management system using Blockchain



VI. IMPLEMENTATION

Let's start to understand the architecture of the project "DOCMAN". Firstly what we are doing is making a document management platform using blockchain for government institutes. We have thought that we are going to make two things: first admin and an interface for the normal employee or user.

Modules of the project:

ADMIN

So for the first part we are going to make admin interface

1. Department and subdepts for the initial on boarding we are going to let the admin add the name of the department and logo and if he wants can create sub depts
2. Roles: Then following that we let the admin put the roles inside of the department , after defining the roles which can be anything for the specific department eg SWE, SQA etc.
After doing that admin will add the schema for each role, the schema will be basically name , email, phone no or anything specific for the role itself.

Creating a schema for the role/ user will be just normal clicks and he will be done after that. If he wishes to add a specific field which is directly related to the department's role then we can provide an add attribute button and he will be able to define it in that text box.

Mostly we can stop the onboarding screen and then move onto the normal screen that will show him a column consists of the department (for SIH members reference - just like Calyxpod) and that column will contain different things like department , roles, access levels, users, files etc.. (but mostly limited to these things)

3. Users: Creating new user, admin can create new user by selecting the role and department and then the schema created in the roles section will be shown to enter the user details and admin can create the user after which the user will receive the mail for signing in to the system with maybe username/email and password
4. Access Levels: He can also define the access levels of the roles and then he can proceed into it by rearranging the roles, we can show a instruction which will say that top means the highest role and the bottom means the lowest roles
5. File: The next thing that the admin will do is that he can do thing in the file section - mostly from our discussion we think that admin can add categories of files and that will be used by our user We have thought that we are going to keep multiple admins and every admin will be having the same privileges

USER

1. Sign In: For the user part we have thought that every user will get a username or email and a password from the admin and then we are going to let the user change the password according to how he wishes.
2. Tutorial/ How to use system/ Demo: For every user we are going to show a tutorial section that will show what are things that are present on the interface.
3. Application Status: The interface will mostly contain things like approved application, pending application, rejected application, and if sent back by higher authority for review or any updates that will be shown under review category.
4. Application category: We can also make filters that will show the messages that are based on particular criteria which can be basically - circulars, leave applications or any other things that the admin specified in his file section.
5. Create application: The user can make applications which are a part of file creation or document creation. After the user clicks on the + icon he will be able to create documents. In the creation phase he will have options for:
 - Attaching documents basically uploading from the local file system.
 - Write text according to the requirement of the document.
 - Adding the category of the document itself (defined by the admin)
 - Receiver of the application/ to whom he wants to send the application.

Details of the implementation of certain features

Document movement: Unicast, Multicast, Broadcast

- Whom is the document redirected to (is to only one person or a broadcast message OR MULTICAST)

Document access level

1. We have thought that we are going to the read and write access to every role that is above the documents creator role
2. and everyone below it will have read only access only if the user sends
3. unless and until specifically defined by the document owner that he wants the people of the lower group as well to have the write access to the documents

Document Attachment File types

We are trying to support pdf, sheets and images which can be of file size 25 MB. Now that user has created the document itself that is he has uploaded all the attachments, did all the writing and everything, he can now send this to other people and that is done.

Receiver of the file (user)

On the receiving end we think that the person whom the file is directly to can approve, reject or make changes in document itself (if he has the access level above the sender) - this is where file annotations come into play when he makes changes inside of the documents attachments. The receiver is basically the person who is going to get the document, he can change the document because we are going to give the edit access to him (dependent on access level).

- Approval - In this we are just going to let the user approve the document by clicking on a button and then the user can add a signature directly to the document and approve them.
- Reject - We are just going to send it back to the sender letting him know that the file was rejected.
- Make changes - He can make changes in the document or the file attached by using our built in editor for pdfs but currently are not supporting editing of excel sheets or any file type.
- Forward - Can forward the application to others as mentioned in Document movement.
- Annotations - User can annotate the application
 - Add signature
 - Add textbox
 - Highlight text
- REVERT BACK FILE TO SENDER FOR
- CHANGES: If the receiver needs to have some more info he can send back the file for correction?

File changes/ Annotation

If he decides to make changes in the file itself then we are going to give an interface where he can

Add text boxes

- Edit the text (FFF - Far Future Feature) - not going to happen.
- Move the text boxes.
- Add a signature which he can upload directly from his computer.

And then finally send it back to the sender.

VII. CONCLUSION

This study has found that there are a lot of institutes that are not able to manage documents in an effective manner and secure manner so we are proposing a solution that will help them to solve this issue. With the help of this application, the institute or department will be able to manage applications such as leave applications, approvals, and circulars effectively and will be able to keep documents secure from various cyber security threats. DocMan is a secure, scalable, and easy-to-use web application which is customizable and can be used effectively to manage the department's application process digitally at high speeds and transparency. We are providing a tracking feature to help understand where the application is right now and what its status is.

REFERENCES

- [1]. V. Sudha; R. Kalaiselvi; D. Sathya (2022), Blockchain based student information management system, [7th Asia Conference on Power and Electrical Engineering (ACPEE)]
- [2]. Eu Wang Kim, Min Seo Park, Kyoungmin Kim, and Kyong Ju Kim (2022), Blockchain-Based Automatic Tracking and Extracting Construction Document for Claim and Dispute Support
- [3]. Iftekher Toufique Imam, Yamin Arafat, Kazi Saeed Alam and Shaikh Akib Shahriyar (2021), DOC- BLOCK: A Blockchain Based Authentication System for Digital Document
- [4]. Rosa Pericàs-Gornals¹, Macià Mut-Puigserver, M. Magdalena Payeras-Capellà(2022), Highly private blockchain-based management system for digital COVID-19 certificates.
- [5]. Document Management System using Blockchain, Shanmugaraja P; Susmitha K. S; Swadha S; Vijay R; Naveen G [n.d.]. Advanced Encryption Standard. https://en.wikipedia.org/wiki/Advanced_Encryption_Standard.
- [6]. Fernando, H., Hewavitharana, T., Perera, A.: Evaluation of electronic document management (EDM) systems for construction organizations, pp. 273–278 (2019)
- [7]. Opitz, F., Windisch, R., Scherer, R.J.: Integration of document-and model-based building information for project management support. *Procedia Eng.* 85, 403–411 (2014)
- [8]. Xie, L.: Evaluation of electronic document and record management programs in Canadian municipalities. The University of British Columbia (UBC) (2006)
- [9]. Bazlamit, S.M., Ahmad, H., Ayoush, M.: Document management systems in small and medium size construction companies in Jordan. In: *Proceedings of the 6th International Conference on Engineering, Project, and Production Management (EPPM 2015)*.
- [10]. Hasan, H.R., Salah, K.: Combating deepfake videos using blockchain and smart contracts. *IEEE Access* 7, 41596–41606 (2019)
- [11]. Khaled, S., Habib, M., Nishara, N., Ala, F.: Blockchain for AI: review and open research challenges. *IEEE Access* (2019)
- [12]. Chaer, A., Salah, K., Claudio, L., Ray, P.P., Sheltami, T.R.: Blockchain for 5G: opportunities and challenges. In: *Proceedings of IEEE GLOBECOM 2019, Waikoloa, USA* (2019)
- [13]. Heiskanen, A.: The technology of trust: how the Internet of Things and blockchain could usher in a new era of construction productivity. *Constr. Res. Innov.* 8(2), 66–70 (2017)
- [14]. Morabito, V.: *Business Innovation Through Blockchain. The B³ Perspective*. Springer, Cham (2017).
- [15]. <https://doi.org/10.1007/978-3-319-48478-5>
- [16]. Chuen, D., Deng, R.: *Handbook of Blockchain Digital Finance and Inclusion*, vol. 1. China Tech, Mobile Security, and Distributed Ledger, USA (2017)
- [17]. Gupta, M.: *BlockChain for Dummies*. Wiley, USA (2017)
- [18]. Brunnler, K.: A logic of blockchain updates, Scopus. In: *Conference Papper* (2017)
- [19]. Davidson, S., de Filippi, P., Potts, J.: Blockchains and the economic institutions of capitalism. *J. Inst. Econ.*, 1–2 (2018).
- [20]. Greenspan, G.: *Multichain Private BlockChain White Paper*.
- [21]. <https://www.multichain.com/download/MultiChain-White-Paper.pdf>
- [22]. Nakamoto, S.: *Bitcoin: A Peer-to-Peer Electronic Cash System*. <https://bitcoin.org/bitcoin.pdf>
- [23]. Haughwout, J.: Tracking medicine by transparent blockchain. *Pharm. Process.* 33(1), 24–26 (2018).
- [24]. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041910404&partnerID=40&md5=612062b1d6063ef645c89cff7c793d93>
- [25]. Dagher, G.G., Mohler, J., Milojkovic, M., Marella, P.B.: Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustain. Cities Soc.* (2018). <https://doi.org/10.1016/j.scs.2018.02.014>

- [26]. Pazaitis, A., De Filippi, P., Kostakis, V.: Blockchain and value systems in the sharing economy: the illustrative case of Backfeed. *Technol. Forecast. Soc. Chang.* 125, 105–115 (2017). <https://doi.org/10.1016/j.techfore.2017.05.025>