International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

# Document Management System, Open Source and Secure

**Shraddha More[1], Snehal Pagare[2], Shivani pawar[3], Prof. S.K. Gondhalekar[4]**

*[1,2,3,4]UG Students, Computer Engineering Dept, KKWCOE & R,Nashik , Maharashtra, India*
*[5]Assistant Professor, Computer Engineering Dept, KKWCOE & R,Nashik , Maharashtra, India*

---

**Abstract -** : *There are various documents that are generating on daily basis in fields like education, government sectors, medical etc. There are various applications available for storing and managing these generated documents. The data generated by these applications are large in amount. So it is necessary to build the system which will store the data on cloud and on standalone server. The proposed system will let the user to scan, upload or import documents to Document Management System. The system will facilitate compression without quality or data loss and provide authentication and authorization mechanism using OAuth2 and two-factor authentication. System will store the documents in PDF format at designated sever locations (cloud) and user can view documents within the system i.e. in Web Application. It will categorize files based upon Date etc. User can share files using links with password protection and limited time access. The system will use Tortoise CVS for version control of documents. It will track changes of each and every activity of logged in user. Our aim is to develop system for healthcare purpose to efficiently store and access documents, saving time and resources in open source environment where document security is important.*

***Key Words*: *Index Terms*—OAuth2, RSA algorithm, Tortoise CVS, Cloud, Open source.**

## 1. INTRODUCTION

The Open Source Document Management system (DMS) is a document management system for organizations for efficient management of documents in their organization. Up till now, there has not been a proper solution for managing documents that is open source. Documents are still transferred from one department to another by physical means (such as pen-drives, compact disc or are mailed via some web services (Email, file sharing solutions etc.). Even if the organization has a file-server and a Local Area Network setup, it still is not the most efficient way to manage the documents. Thus, there arises a need for a Document Management System for effective document management for various enterprises. Various software solutions like Domain, Share point (from Microsoft), Secure Doc came into existence to solve this need but they are not Open Source and Cost efficient. Thus, in our paper, we are designing and giving a layout of a n efficient Document Management System as solution that addresses the needs of the organizations in today's competitive world in a cost effective manner. The proposed Document management system will give user opportunity to retrieval documents and information fast and without wasting time. DMS have features like uploading document in any format, storing, searching and retrieving it in PDF documents, as well as advanced features such as Version and access control. DMS help user to have a space for storing and sharing documents with other users in system using link password protection. The proposed DMS will provide Security by using O-Auth2 Authentication and authorization framework. The documents are stored in compressed format without losing data thus saving the space for storage.

### 1.1 ARCHITECTURE

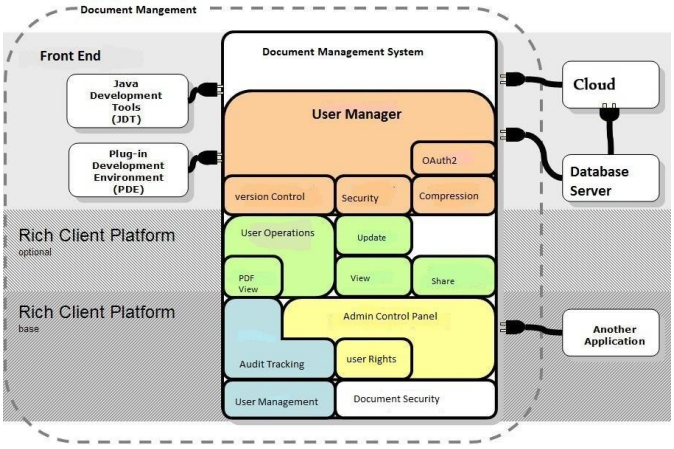


**fig .** System Architecture

*A. Java Development Toolkit*

The Java Development Kit (JDK) is an implementation on the Java Platform. The JDK has as its primary components as collection of java programming tools such as javac, java, applet etc.

*B. Plug-in Development Environment*

The Plug-in Development Environment (PDE) provides tools to create, develop, test, debug, build and deploy plug-ins, fragments, features, update sites and RCP products. PDE also provides comprehensive OSGi tooling, which makes it an ideal environment for component programming.

*C. Client Platform*

Client able to upload, view, download documents. Whereas client is able to access his own platform only.

*D. User Management*

- OAuth2 :

The O Auth version 2 protocols or framework allows third-party applications to grant limited access to an HTTP service, either on behalf of a resource owner or by allowing the third-party application to obtain access on its own behalf.

- Version Control :

Version control is the process by which different drafts and versions of a document or record are managed. Integrated Tortoise CVS will be used for version control.

- Document Security:

Stored files will encrypted and provide encryption in transit. Files to be encrypted with an RSA based 2048- bit key created through perfect forward secret key.

- Document Compression:

System will provide compression if size of documents exceeds than the limit of storage. System will achieve this without data as well as quality loss.

- User Operations:
  - View: System will view uploaded PDF documents within the system.
  - Update: System will update documents if required.
  - Share: System will Share file using links with pass- word protection and limited time access
- Admin Control Panel
  - Audit: Tracking changes of each and every activity of logged in user.
  - User rights: System will provide User rights to access documents.
- Cloud server Cloud is provided as physical or virtual infrastructure that per-forms application- and information- processing storage.
- Database Server A database server houses a database application that provides database services to client.

## 2. SPECIAL SYSTEM FEATURES

The core task of any Document Management System is the storage and retrieval of documents from one or various sources in a centralized manner. The DMS improves on this basic model by providing certain other features that are very important in today's world

*A. Encrypted Document Transfer*

This feature allows the user to send their documents over to the main server for storage purposes in the organization. However, before the file can be transferred it must be encrypted for security purposes [5]. This is one of the most crucial functions of the system. The Document(s) chosen for the transfer are themselves encrypted using RSA asymmetric encryption algorithm

We shall elaborate on this below.

### 2.1. Rivest-Shamir-Adleman(RSA) Algorithm:

The RSA cryptosystem is the most widely-used public key cryptography algorithm in the world. It can be used to encrypt a message without the need to exchange a secret key separately. The RSA algorithm can be used for both public key encryption and digital signatures. Its security is based on the difficulty of factoring large integers. Party A can send an encrypted message to party B without any prior exchange of secret keys. A just uses B's public key to encrypt the message and B decrypts it using the private key, which only he knows. RSA can also be used to sign a message, so A can sign a message using their private key and B can verify it using A's public key.

**Key Generation Algorithm:**

1. Generate two large random primes, p and q, of approximately equal size such that their product n = pq is of the required bit length, e.g. 1024 bits.
2. Compute n = pq and = (p1) (q1).
3. Choose an integer e, 1¡e¡1¡e¡, such that gcd (e,)=1gcd(e,)=1. 4. Compute the secret exponent dd, 1¡d¡1¡d¡, such that ed1moded1mod.
4. The public key is (n,e) (n,e) and the private key (d,p,q) (d,p,q). Keep all the values d, p, q and secret. [Sometimes the private key is written as (n,d) (n,d) because you need the value of n when using d. Other times we might write the key pair as ((N,e),d)((N,e),d).

n is known as the modulus.

e is known as the public exponent or encryption exponent or just the exponent.

d is known as the secret exponent or decryption exponent

**Encryption:** Sender A does the following:-

1. Obtains the recipient B's public key (n,e)(n,e).
2. Represents the plaintext message as a positive integer mm with 1¡m¡n1¡m¡n
3. Computes the ciphertext c= mem odn.
4. Sends the ciphertext c to B.

**Decryption:** Recipient B does the following:-

1. Uses his private key (n,d)(n,d) to compute m=cd mod nm= cd mod n.
2. Extracts the plaintext from the message representative mm.

*B. Document (file) Compression*

This feature shall compress the size of the encrypted file which resides over on the server. The encrypted file has a size that is larger than the plaintext file. Thus compression shall allow us to optimize the size of the encrypted document.

*C. File Versioning*

This feature of the system keeps the track of the versions of a particular document over time. User viewing particular document may want to revert to a previous version of the same document. For this the server will keeps multiple copies of the same document on it.

Fig shows the directory structure which provides the file versioning capability. Every Document has a hidden folder (in the same parent directory) of the same name.

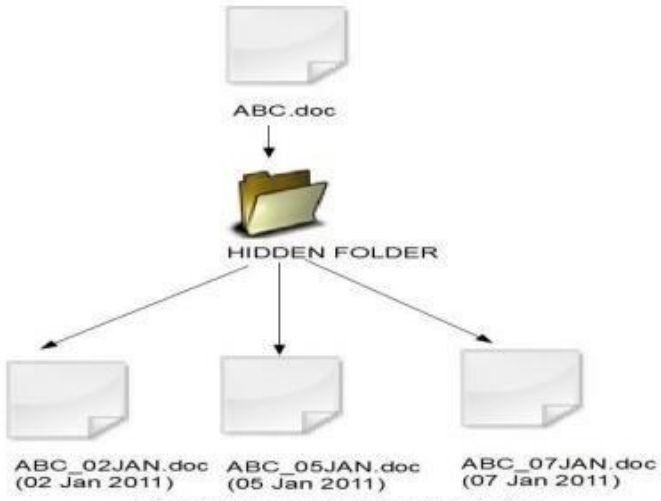

Fig A.5 Directory structure for File Versioning

*D.* Searching Technique

Every good Document Management System must provide a search feature that enables its users to quickly find a particular file kept in the system. So a general approach would be to keep all the file information (file name, location, user, etc in a table inside the backend Postgress SQL, whenever a user will search for a file, its entry would be checked in that table. And the file would be retrieved for the user, depending on relevance to the user that asked for it.

## 3. GENERAL SYSTEM REQUIREMENTS

1) *Database Requirements:* Postgress SQL
2) *Software Requirements:*
   1. AngularJS , Jquery, Spring Boot Framework
   2. Tortoise CVS version control
   3. Spring Boot Framework
   4. Operating System: Windows or Linux
3) *Hardware Requirements:* 1.Database Server
   2. Cloud

## ADVANTAGES

*A. Reduced Storage Space*

Commercial property costs are increasing and so is the expense to store paper documents. A software-based document management system (DMS) that can reduce the need for file cabinets, boxes and storage bins is a valuable asset to any enterprise, freeing up precious office space.

*B. Enhanced Security*

Document security is vital to many enterprises. DMS provides better control over sensitive documents. Access to documents can be controlled at the folder level for different groups or individuals. Also, a DMS leaves an audit trail of who has viewed a document, when it was accessed, and how it may have been modified. Managed documents are highly traceable and can be tagged to allow for automated alerts. User posts a fresh version of a particular file, the system not only replaces the old file but also places a copy of the file in the hidden folder associated with that particular file. The files inside the Hidden Folder maintain the versions of the file with the date of posting concatenated with the actual name of the file.

*C. Easier Retrieval*

Searching for and retrieving documents can be very time consuming. A document management solution, specific for your enterprise, can be a powerful, time-saving tool. Depend- ing on the solution implemented, a DMS can retrieve files by a word or phrase in a document full text search, or DMS can apply index categories to a document or folder. Easier integration with business applications facilitates access to critical information. DMS also gives the opportunity to access documents remotely. As long as

there is internet connection, documents can be accessed from virtually anywhere.

*D. Better Collaboration*

With DMS, information sharing and collaboration can be a lot easier. Documents captured from different sources can be accessed from multiple locations. Electronic imaging makes sharing documents over a network via email or the Internet possible.DMS provides greater visibility to business processes and can allow for better work-flow monitoring. Authorized access by external users can be allowed and monitored.DMS also offers Version Control which allows for older versions of the same document to be recovered if needed. This feature is important if several parties work on the same document and there are changes made to a document that were not authorized or meant to happen.

*E. Better Backup and Disaster Recovery*

Any document management solution should include a data backup and disaster recovery plan. With a DMS, documents are highly traceable and can be tracked within a range of criteria. Also, because imaged documents are centrally stored they cannot be lost or misplaced after viewing. Documents are less likely to be misfiled, and if so they are easier to locate with cross-indexing. Documents can be digitized and archived at their point of entry into the system.

## 4. CONCLUSION AND FUTURE WORK

The proposed system i.e DMS ensures that documents is safe, accurate, and securely accessible. It will works as an open source System that stores and manage documents of users. System will provide version control, authentication mechanism, security mechanism, compression functionality to users. In future we can develop same type of System for Android and IOS users.

## REFERENCES

1. V. L. Orlov, E. A. Kurako Electronic document management systems and distributed large-scale systems, published in 16 November 2017 on IEEE Explorer..
2. Sengol mary J, Usha S WEB BASED DOCUMENT MANAGEMENT SYSTEMS IN LIFE SCIENCE ORGANIZATION, published in 19 April 2016 on IEEE Explorer.
3. Tongjin Lee, Jun Iio Document Management System Based on ISAD(G), ,published in 10 December 2015 on IEEE Explorer.
4. Yeong-Tae Song, Sungchul HonJ, Jinie Pak Empowering Patients using Cloud Based Personal Health Record System, published in 06 August 2015 on IEEE Explorer.
5. Hediyeh Baban,Salimah Mokhtar An Enhanced Document Management System for SME, published in 20 August 2012 on IEEE Explorer.
6. Hediyeh Baban, Salimah Mokhtar Online Document Management Sys- tem for Academic Institutes, published in 20 January 2011 on IEEE Explorer..
7. Mohamed Amin Sakkar, Bruno Defude, and Jorge Tellez EDocument Provenance in the Cloud: Constraints and Challenges, published in 10 october 2010 on IEEE Explorer.
8. Kyrene G. Alberto, Camille M. Abella, Ma. Gracia Corazon E. Sicat, Jasmin D. Niguidula , Jonathan M. Caballero Compiling Remote Files: Redefining Electronic Document Management System Infrastructure (CReED), published in 15 January 2010 on IEEE Explorer.
9. Hu Ye, Hu Tao Secure documents management in supply chain system, published in 22 April 2003 on IEEE Explorer.
10. Changtao Qu, Thomas Engel, Christoph Meinel Implementation Of A Document Management System Based On WEBDAV Protocol, pub- lished in 06 August 2002 on IEEE Explorer.