A Document Management System for Searching and Storing Documents

Ammar Poonawala[1], Harsh Pereira[1], Chetan Vatharkar[1], Advait Trivedi[1], Shraddha Khonde[2]

Student, Dept. of Computer Engineering, Modern Education Society's College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India[1]

Professor, Dept. of Computer Engineering, Modern Education Society's College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India[2]

ABSTRACT: Document Management System has generated great interest in the business world.A welldesigned document management system entails effective and efficient retrieval and processing of stored data.Information Retrieval is the branch of data-mining which focuses on the text-search in a collection of huge set of document-store. The information retrieval process begins when a user enters a search query.User query is matched in the database and documents are ranked accordingly.In many industries, document management systems and their applications have transformed the way people work. This is because mobile devices act as the linchpin for the other technology mega trends, delivering into peoples hands the critical information and insight they need in a given moment, wherever they are.A document basically contains a rich source of information. As the amount of documents required for large-scale and small-scale enterprise business is huge, so we need an efficient and accurate Document Management System.

KEYWORDS: Document Search, Cost Effective,Cost Cutting Initiatives, Document Retrieval, Information Retrieval, Spatial Search, Multi level search.

I. INTRODUCTION

A document management system(DMS) is a system (based on computer programs in the case of the management of digital documents) used to track, manage and store documents and reduce paper. Most are capable of keeping a record of the various versions created and modified by different users (history tracking).The term has some overlap with the concepts of content management systems. It is often viewed as a component of enterprise content management (ECM) systems and related to digital asset management, document imaging, workflow systems and records managementsystems.The earliest electronic document management (EDM) systems managed either proprietary file types, or a limited number of file formats. Many of these systems later became known as document imaging systems, because they focused on the capture, storage, indexing and retrieval of image file formats. EDM systems evolved to a point where systems could manage any type of file format that could be stored on the network. The applications grew to encompass electronic documents, collaboration tools, security, workflow, and auditing capabilities. These systems enabled an organization to capture faxes and forms, to save copies of the documents as images, and to store the image files in the repository for security and quick retrieval (retrieval made possible because the system handled the extraction of the text from the document in the process of capture, and the text-indexer function provided text-retrieval capabilities).While many EDM systems store documents in their native file format (Microsoft Word or Excel, PDF), some web-based document management systems are beginning to store content in the form of html. These policy management systems require content to be imported into the system. However, once content is imported, the software (ex. Corona Document Management System) acts like a search engine so users can find what they are looking for faster. The html format allows for better application of search capabilities such as full-text searching and stemming.

II. LITERATURE SURVEY

Presently, information retrieval can be accomplished simply and rapidly with the use of search engines. This allows the users to specify search criteria to obtain results. The earliest electronic document management system managed either proprietary files types, or a limited number of file formats. Earlier adding files and folders onto the computer was a bulky and frustrating process. Also many document management system don't provide the facility to search particular documents or provide limited search facility.

Currently the process of searching the documents is a long and lengthy process, as the documents are scattered all across the internet.Also the search functionality provided by many applications is very basic, and as such searching is very cumbersome, and the user is unable to get find relevant results.

III. PROPOSED SYSTEM

The System Architecture for this application gives a complete detailed view of the functionalities provided which will not only facilitate the user to store documents and search them using our advanced search functionality, but will also show how the general public can access view and search the files, that have public scope. Additionally we have also added a admin functionality wherein the admin can manage the entire system, all through a web interface. The architecture is given below.
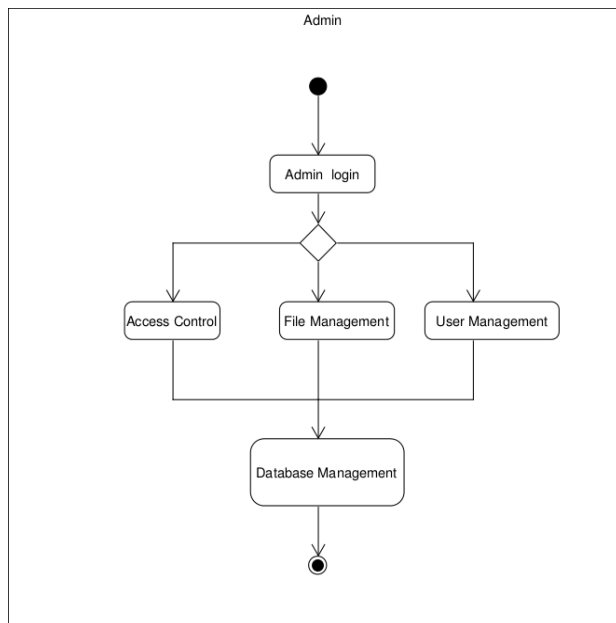

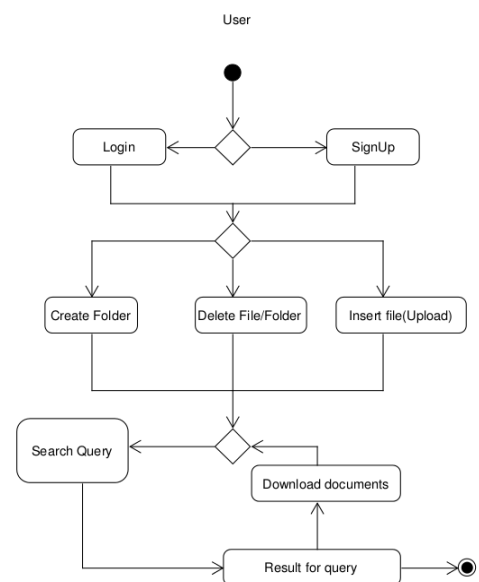
Fig .1. Admin Flow   Fig.2. User Flow

Here the included modules will not only provide functionalities but will also reduce the overall complexity.

1. User login
Here we provide a user signup page, and a login page. Also a additional login page for the admin is provided.

2.Manage your environment(i.e users' files and folders).

Here the user can upload documents and create his folder. Also a user can grant his file public access scope. When the user uploads a file the corresponding location is also noted.

3.Edit folder's and files
User can edit his files and folders metadata. And other information associated with the above.

4.User Edit
The user can edit his account details such as email address, password etc.

5.User Search
The user can search through all his files, based on the file content and metadata.
A multilevel search is provided where in the user can create his own query dynamically.

6.Public Search
The public (i.e) not registered users can search through the all the files based on language, location, and a multilevel search.

7.Admin Functionality
The admin can manage the users and files. The admin can also disable users, and change the metadata and location of the files.

IV. PURPOSE OF THE SYSTEM

The application is used to simplify the purpose of sharing files. It does by giving the user a functionality where in he can easily make a document either private or public.Also using our advanced search functionality the public can easily over the vast collection of documents, using the language, location and multi level search functionality.The system automatically detects the location and language of the document so as to not burden the user while he is uploading the documents.The location based search helps the public to find documents near them(As obviously the public is interested in relevant data example. If you were to search for documents related to news, you would obviously want news repots from nearby).

V. SIMULATION & RESULTS

We have implemented the above system as a web application, using ASP .NET for our backend, and Apache Solr for searching. The application is currently being run on a Microsoft IIS Server. We have also used the HTML5 Web API to help in retrieving the users location The language detection functionality is implemented using the LangDetect Language Detection, and Apache Tikka to help us parse our PDF files.
The main search functionality is shown in the diagram.

Fig.3. Advance Search

VI. CONCLUSION AND FUTURE WORK

In this paper we describe an efficient document management system where user can handle various documents by functionalities like upload,download,edit etc.using Apache Solr search we have designed multi-level and spatial search feature for searching documents. We have also admin functionality that controls our entire system. Our system can be used for searching purpose in many business enterprises. It is thus an efficient system for tracking user documents by spatial search.

REFERENCES

1. Hongpu Hu1, Guoqiang Sun,Panpan Qin, Hongyan Liu,Fan Yu, 'The State of the Art of Electronic Medical Document Management.', IEEE pp. 978-1-5090-5521-0/16/ 2016, 2016.
2. Sugao Sumiya and Takashi , Saito,'Development of a Multimedia Document Management System for Cooperative Work Environment', IEEE, 978-1-4799-6000, 2003.
3. Shyamalesh Khan,B V N Prasad,S Selvi,Usha Rani,'Document Management System: An Explicit Knowledge Management System', IEEE, 78-9-3805-4416-8,2015.
4. Srikant Krishnan, Nary Subramanian,'Evaluating Carbon-Reducing Impact of Document Management',IEEE, 978-1-4799-8884-6.,2015.
5. Rashidah, Mokhtar,Safura Adeela, Sukiman,Nor Fauziah, Abu Bakar,' An Integrated Document Management System for Managing Self Programme Accreditation Using Scrum Approach ', International Symposium on Technology Management and Emerging Technologies (ISTMET 2014), May 27 2014.