

# Document-oriented Management System: Structuring Documents of an Academic Program for Accreditation

Mohammad Rashid Hussain, Mohammed Abdul Khaleel, Khalid Mohiuddin

**Abstract:** *This paper is first part of salami slicing publication. It presents the initial part of documents and evidence (DOVE) management system of an academic program which is offered in a university environment. The second part will present the entire system for managing program documents and supports its accreditation process. This study describes a development process of documents structuring with the goal of converting manual documents to automated system. The significance of the proposed system is to structure documents required for the program accreditation. It is a very complex process to organize all the required documents into a repository and responding users' query. It is more complex when structuring scan documents and warehousing them for future access. For developing the DOVE database, different levels of the program hierarchy has to be considered. It has been structured adopting both rational and hierarchical model. Accreditation requirements is focused in designing the system. For the initial part, DOVE is structured for the first three stages of the proposed system. It structures approximately 6900 documents from the multi-levels of the program constituency. The result shows a perfect naming convention to each document using indexing techniques.*

**Index Terms:** *Document-oriented database; documents structuring; document management system; program accreditation; academic institution*

## I. INTRODUCTION

Over the decades, most of the academic institutions have been struggling to structure the Documents and Evidence (DOVE) created by the academic practices. It is very difficult to organize the huge volumes of DOVE or documents for the future references [1]. Several document management systems have been developed for organizational activities; some of them are organization's specific [3]. Most of these systems are being developed for commercial organizations and less focus has given for educational institutions and their specific requirements. These systems usually classifies on their data model and being categorized as document-oriented, key-value, wide-column, and graph-based [2]. Generally, in academic institutions

document-oriented management systems are being placed to fulfill most of their institutional and academic practices requirements. Document-oriented databases or documents stores are needed to be user friendly and have potential to response random queries from the users [3]. Documents stores indeed have to be flexible data model, easy to maintain, and allows access to the required random documents [5, 6] in user-friendly interface.

This study addresses a real time requirement of structuring documents and evidence of an academic program and to meet the requirements of the program accreditation. It develops DOVE centric and document-oriented database management system i.e. "DOVE Management System (DEVMS)". This system collects documents of academic practices from the multiple levels of program hierarchy [9, 12]. The documents structure and organize based on nesting and hierarchical model [12]. Each document is structured with a naming convention (i.e. SRY1S1, SR1S2, and more) and stores at different levels of program hierarchy. The structuring of each document represents a structural variety to ensure the hierarchy of documents and basic structure for document retrieval [8]. It provides a strong database model that manage documents efficiently and develops a strong documents database i.e. DOVE Big Data.

Presently, we have been designing the system's framework considering the features such as user friendly, efficient access, secure documents store, and a powerful documents warehouse. This will be an efficient document-oriented database management system for managing program's documents. More focus has given to meet the accreditation requirements. The rest of the paper is organized as section two describes the methods adopted for the system development.

## II. LITERATURE REVIEW

This section presents a brief summary of document-oriented management systems in the literature. The proposed system is document centric and being developed considering the needs of the academic program. Initially, the study covers thirteen [1-13] (listed in 'references' section) other studies, exists in the literature. We will be very specific in the second part.

**Revised Manuscript Received on May 10 ,2019**

**Dr. Mohammad Rashid Hussain**, College Of Computer Science, King Khalid University, Abha, KSA

**Dr. Mohammed Abdul Khaleel**, College Of Computer Science, King Khalid University, Abha, KSA

**Dr. Khalid Mohiuddin**, College Of Computer Science, King Khalid University, Abha, KSA



### III. PROPOSED METHODOLOGY

In this section, we describe the preliminaries of documents structuring to facilitate the development of an automated system, say, “DOVE Management System (DEVMS)”. The system develops to automate the documents that are generated during the program delivery in a university environment. More specifically, the system is designed to facilitate the program’s accreditation process.

We present the methodology that is adopted in developing document-oriented system for the program’s document management [9]. It is explained in six stages according to the hierarchy of documents’ generation [7].

First stage: we collect the required documents requesting from the university data center, at the end of each semester. According to the university IT policies, we (the team involved) can’t retrieve directly all the required documents from the data center. The required documents of three different systems are collected and stores. The first one is the students’ registration system which generates documents that consist of students’ enrollment information and related reports. The second one is learning management systems (LMS), which is a very huge and produces bulk of documents during the semesters. And the third one is examination processing system (EPS) and generates documents in huge number.

Second stage: here we get documents from the stage one that is grouped into three systems and at the same level in the hierarchy. Here, the documents of five years only are extracted from the stage one. For the program’s accreditation, maximum of five years documents are required. In this case, the academic years from 2013-14 to 2017-18 are considered. At this stage, the total number of documents is approximately 710 (10 from SRS+300 from LMS + 400 from EPS). These documents integrates with stage three and grouped into stage four.

Stage three: this stage is very crucial and core part of the DEVMS. Here, the documents and evidence of academic activities of the program delivery are created during each semester of the academic year. The process repeats for five years and documents are stored for the future access. Stage three includes documents of three different parts of the program. Following practices have been followed during each semester.

Course file, it consists of course assessments, students’ samples, course specification, course report, and etc. We have selected 40 courses of the curriculum for the accreditation purpose. One file for each course, it includes minimum ten documents. Here the total number of documents from the course files are (40 courses \* 10 documents \* 10 semesters =4000).

Students’ outcomes measurement, which includes course measurement and performance analysis of each course and supporting evidence. The program skills described into 10 learning outcomes. For each learning outcomes, three courses are allocated, three supporting evidence for each outcome and this practice for ten semesters. Here the total number of documents is approximately 2900

Here approximately the total number of documents from three stages is 6900.

The remaining three stages (i.e. 4-6), we will describe in the second part of the studies.

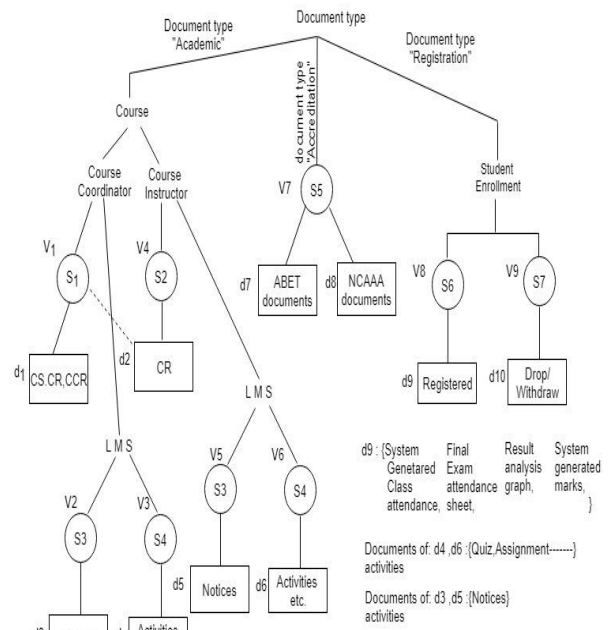


Figure 1 Collection in document oriented database by capturing hidden rule in the form of decision tree, to create, precise and concise DOVE approach

Documents of course files:  $d_{cf} = U_{d \in S(D)}$

$S(D_{CF}) = \{d_1, d_2, d_4, d_9\}$  where Set of documents

$S(D) = \{d_1, d_2, \dots, d_{10}\}$

Documents of Accreditation:  $d_{ACC} = S(D_{ACC}) =$

$U_{d \in S(D)} : \{d_7, d_8\}$

$d_7 : S(D_{ABET}) \quad d_8 : S(D_{NCAA})$

Documents of students Enrolled:

$d_{SE} = S(D_{SE}) = U_{d \in S(D)} : \{d_9, d_{10}\}$  Where  $d_9 : S(D_{Reg})$ , and

$d_{10} : S(D_{drop/Withdrawal})$

Our proposed approach have divided in three stages, first stage, the system have been developed for automated documents management. More specifically, the system is designed to facilitate the required documents for the program’s accreditation. In next stages, a set of experimental tests made to demonstrate the effectiveness and efficiency of our approach. In this paper we propose a technique, called documents and evidence (DOVE) Management System (DEVMS), to explain our approach within the collection of documents oriented database. In next stages of our work, the algorithms have introduced a combination of hierarchical and k-means clustering approach based on cosine similarity assessment technique to group the similar documents effectively. For document summarizations are Preprocessing, feature selection, context representation, content selection, context ordering, and sentence realization are considered. For better understanding of this strategies, consider a decision tree built in the domain of document management. Some of the important documents (Course Specification (CS), Course Report (CR), Cumulative Course Report (CCR), etc.) have been considered as leaf node in the tree, and internal nodes are associated with a document attribute that express value based condition that have been explained in our next stages work.



#### IV. DEVMS FRAMEWORK

In the process of system development, DOVE controller who has supposed to be heading the system development process. He has a complete understanding of documents required in the process of program accreditation, which is shown in figure 1. After developing the system the authorized users will access the system for uploading and downloading their generated documents. Figure 1 shows document resources such as university data [7] center and all the stages where documents are being generated.

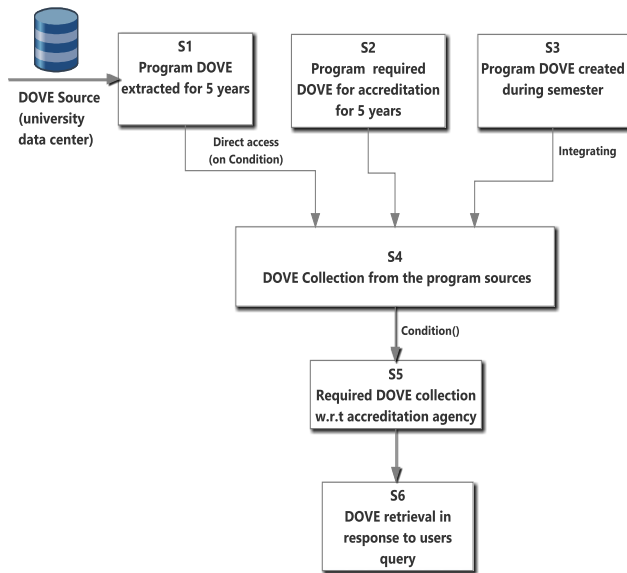


Figure 2 DOVE generates in program delivery

Defining DOVE The frame work describes structuring of documents and the process of converting manual document management to automated system [11]. The process describes for the first three stages and the remaining will be in the second part. The following variables and equations apply in structuring documents for the first three stages. The structuring of documents considers only for the previous five years, considering the requirements of program accreditation. But the same applies for any document generated during program delivery for any academic year.

$n =$  represents number of years that is “ $n=1$  to 5”

$N_{nt}$ :  $nt =$  Number of notices for stage 1 ;  $N_{act}$  :  $act =$  Number of activities for stage 2

$N_{crs}$ :  $crs =$  Number of courses (i.e. 40) in each semester for stage 3

$A_Y =$  Academic years (2013-14, 2014-15, 2015-16, 2016-17, 2017-18)

**Stage 1:** Student registration system (SRS) abbreviated “SRS

$$A_Y: S_{RS} = SR [((Y_n)_{\forall n}(S_1)); ((Y_n)_{\forall n}(S_2))] \dots(1)$$

**Stage 2:** Learning Management System (LMS): Two events have been considered “notices” and “activities” Learning Management Notices (LMN), Learning Management Activities (LMA)

$$A_Y: L_{MN} = LMN [(((Y_n)_{\forall n}(S_1))_{5 \times 1} ((N_{nt})_{\forall nt})_{1 \times nt}); (((Y_n)_{\forall n}(S_2))_{5 \times 1} ((N_{nt})_{\forall nt})_{1 \times nt})]_{5 \times nt} \dots(2)$$

$$A_Y: L_{MA} = LMA [(((Y_n)_{\forall n}(S_1))_{5 \times 1} ((N_{act})_{\forall act})_{1 \times act}); (((Y_n)_{\forall n}(S_2))_{5 \times 1} ((N_{act})_{\forall act})_{1 \times act})]_{5 \times act} \dots(3)$$

Equation (2) and (3) describe DOVE for two types of documents, for notices and activities

**Stage 3:** Examination Processing System (ExPS), generates documents for result for forty courses in each semester.

$$A_Y: E_{XPS} = EP [(((Y_n)_{\forall n}(S_1))_{5 \times 1} ((N_{crs})_{\forall crs})_{1 \times crs}); (((Y_n)_{\forall n}(S_2))_{5 \times 1} ((N_{crs})_{\forall crs})_{1 \times crs})]_{5 \times crs} \dots(4)$$

Equation (4) gives result of 40 courses for each semester of and academic year

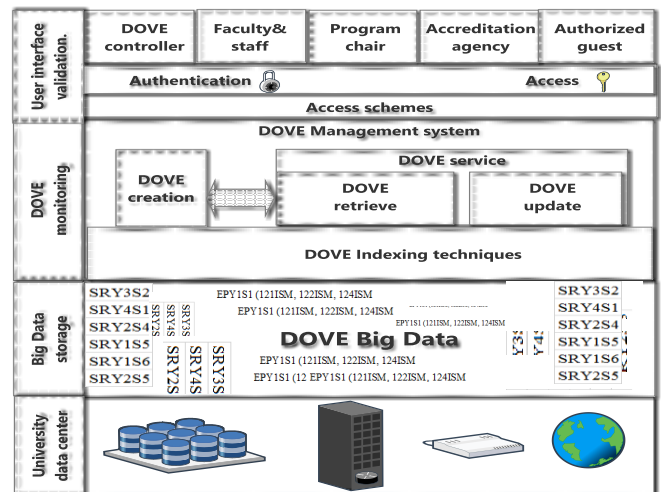


Figure 3 DEVMS multilayered architecture

Figure 2 describes the multilayered architecture for document-oriented database management system or DEVEM. The architecture is segmented into multiple layers and describes from the bottom, document source from the university data center, DOVE considered as Big Data, indexing technique (s), document management system, access schemes, and user interface on the top.

#### V. RESULT ANALYSIS

Path optimality algorithm

$A = 0$

for each ;  
MAKE – SET (v)  
for each (u,v) ordered by weight (u,v) increasing;

if FIND – SET (u) FIND – SET (v);

UNION (u,v)

Return A

Here u and v are the set of documents. The above path optimality algorithm has been compared with Bachelor and Wilkin’s algorithm to develop document management system to achieve our proposed approach.

Example:



Items	A	B	C	D
Variables (X <sub>1</sub> , X <sub>2</sub> )	(5, 3)	(-1, 1)	(1, -2)	(-3, -2)

	A	B	C	D
A	0	6.32	6.40	9.43
B	6.32	0	3.61	3.61
C	6.40	3.61	0	4
D	9.43	3.61	4	0

D (First Cluster Centre)	A		Clusters
Maximum A(9.43)	...		I {A}
Shift to B(3.61)	6.32	3.61 < 6.32 (no re-assignment is required)	II {B, C, D}
Another cluster C(4)	6.40	4 < 6.40 (no re-assignment is required)	

We proposed a mathematical models and algorithms required for developing the system. We designed an interface that facilitates users to upload or download the documents as they required. In this way, through examples, two clusters have created to do the required operation.

VI. CONCLUSION

The presented work is a part of large study and follows salami slicing principle to present the work. The study aims to develop a document-oriented database management system (DEVMS) for storing and retrieving documents. The first part presents structuring documents and evidence (DOVE) of the practices of an academic program which is offered in the university. The documents are generated during the program delivery over the years. The documents are more specifically structured based on nesting, relational and hierarchical model to manage program documents. The documents are structured for the first three stages which covers approximately 6900 documents of the program constituency.

REFERENCES

1. Tanmay Basu <sup>†</sup>, C.A. Murthy, "A similarity assessment technique for effective grouping of documents" Information Sciences 311 (2015) 149–162
2. Rodrigo Costa dos Santos<sup>a,\*</sup>, Maria Augusta Soares Machado<sup>c</sup>" FIRDOR - Fuzzy information retrieval for document recommendation" ScienceDirect Procedia Computer Science 139 (2018) 56–63
3. Mohamed L. Chouder <sup>a</sup>, Stefano Rizzi <sup>b,\*</sup>, Rachid Chalal <sup>a</sup>" EXODuS: Exploratory OLAP over Document Stores" Information Systems 79 (2019) 44–57
4. Larbi Guezoulia,<sup>\*</sup>, Hassane Essafi <sup>b,c</sup>" CAS-based information retrieval in semi-structured documents: CASISS model" Volume 3, Issue 2, December 2016, Pages 155-162
5. Jin Zhanga, Guannan Liub, Ming Renc,<sup>\*</sup>" Finding a representative subset from large-scale documents" Journal of Informetrics 10 (2016) 762–775

6. Ramzi Chaieba,<sup>e</sup>, Karim Kalti <sup>b,e</sup>, Muhammad Muzzamil Luqman<sup>c</sup>, Mickaël Coustaty<sup>c</sup>, Jean-Marc Ogier <sup>c</sup>, Najoua Essoukri Ben Amara<sup>"</sup> Fuzzy generalized median graphs computation: Application to content-based document retrieval" Pattern Recognition 72 (2017) 266–284
7. Umesh D. Dixita and M. S. Shirdhonkar<sup>"</sup> Face-based Document Image Retrieval System" Procedia Computer Science 132 (2018) 659–668
8. Kjetil Norva<sup>g</sup>" Granularity reduction in temporal document databases" Information Systems 31 (2006) 134–147
9. Ziwon Hyunga, Joon-Sang Park <sup>b</sup>, Kyogu Leea,<sup>"</sup> Utilizing context-relevant keywords extracted from a large collection of user-generated documents for music discovery" Information Processing and Management 53 (2017) 1185–1200
10. Diego Arroyuelo<sup>\*,1,a</sup>, Mauricio Oyarzún<sup>b</sup>, Senén González<sup>c</sup>, Victor Sepulveda<sup>"</sup> Hybrid compression of inverted lists for reordered document collections<sup>☆</sup>" Information Processing and Management 54 (2018) 1308–1324
11. Christopher Southan<sup>"</sup> Retrieving GPCR data from public databases" Current Opinion in Pharmacology 2016, 30:38–43
12. Pushpalatha K<sup>\*</sup>, Ananthanarayana V S<sup>"</sup> A tree based representation for effective pattern discovery from multimedia documents" Pattern Recognition Letters 93 (2017) 143–153
13. Enrico Gallinucci <sup>a,b</sup>, Matteo Golfarelli <sup>a,b</sup>, Stefano Rizzi <sup>a,b,\*</sup>" Schema profiling of document-oriented databases" Information Systems 75 (2018) 13–25

AUTHORS PROFILE



Dr. Mohammad Rashid Hussain is Assistant Professor, at College Of Computer Science, King Khalid University, Abha, Saudi Arabia, and holding PhD degree in Information Technology. His research interest includes educational development & review, educational data mining, cloud intelligence, mobile cloud computing and

optimizations.



Dr. Mohammed Abdul Khaleel received B.Sc. degree in Electronics from Osmania University, India and M.C.A degree from Osmania University, India. After that worked in Global Suhaimi Company Dammam Saudi Arabia as Senior Software Developer .Since 2008 serving as Lecturer at College of Computer Science, King Khalid University, Abha, kingdom of Saudi Arabia. Awarded Phd. in Computer science From Sambalpur University, India. His research interest includes Data Mining, Software Engineering.



Dr. Khalid Mohiuddin holds a PhD in computing and is recognized as researcher and an excellent teaching practitioner at the faculty of Computer Science at King Khalid University, Saudi Arabia. His interdisciplinary research involves cloud computing, mobile cloud, cloud intelligence, academic research, educational evaluation and development in higher education.

