The Digital Library and the Archiving System for Educational Institutes

Article *in* Pakistan Journal of Library and Information Science · February 2019

DOI: 10.47657/2018201453

CITATIONS
30

READS
1,703

2 authors:

Atta Rahman
Imam Abdul Rahman bin Faisal University
213 PUBLICATIONS **2,494** CITATIONS
SEE PROFILE

Fahd Al-Haidari
Imam Abdul Rahman bin Faisal University
47 PUBLICATIONS **1,219** CITATIONS
SEE PROFILE

The Digital Library and the Archiving System for Educational Institutes

Atta ur Rahman
Imam Abdulrahman Bin Faisal University (IAU), Dammam, Kingdom of Saudi Arabia
Email: aaurrahman@iau.edu.sa

Fahd Abdulsalam Alhaidari
Imam Abdulrahman Bin Faisal University (IAU), Dammam, Kingdom of Saudi Arabia
Email: faalhaidari@iau.edu.sa

At present, there are several formats that exist through which data is distributed among online stakeholders. An example of this is the XML, which like other such formats is helpful for traditional inquiry methods and for forming the foundation of query languages such as SPARQL and SQL. Information about primary representation demands a broader assistance for the languages where every piece of data from any resource can substantiate the original queries for searching. Such models are useful for XML based retrieval since several cooperative XML search engines have been developed already. These search engines perform semantic investigation of XML files with data surrounded by the important fields. Therefore, XML files are used to store and index data intended for competent retrieval. In this research, an attempt is made to fill this gap of customized representation and retrieval with a focus on the educational domain. An institute's repository of books, e-books, journals, articles and research theses has been used to retrieve results. A system has been proposed and developed to store the contents of Institute's Databank as an object of the Digital Library. A structured method has been proposed to organize all the data and a system has been developed which extracts meaningful information from the Data Bank. The information repository is established, and the entire data is represented in terms of a unit called Digital Object in the Digital Library. The single unit is represented by recording some quantitative data about it referred to as 'Metadata'. The search is focused on extracting meaningful information from the repository by applying some filtration strategies to get relevant information, best matched with the query terms. At the end, a partitioning and parallelism focused architecture to archive the information for sharing, back-up and collaboration is also proposed. Comparison of the proposed

scheme with state of the art schemes is provided in terms of computational complexity and recall measurement.

Keywords: Digital Library, Information Retrieval, Archiving, IBM Domino, Databank

INTRODUCTION

The Digital Library (DL) becomes an essential part of knowledge in the field of systematic library systems. This is comparatively a novel concept, however there exist many loopholes that are yet to be fixed such as grouping, storing and retrieving relative information etc. Arms (2005) describes that DL is a software-based structure in which the focal point is web expertise. In web, electronic form of data can be accumulated and regained easily by providing rights to privilege the primary data. The primary set of data is focused on digital depiction just like in the form of editorials, published study material, theses, document, e-books and so on. According to Atta-ur-Rahman and Alhiyafi (2018), these types of data might be saved in different formats such as PDF, DOC/DOCX, XML, HTML and multimedia extensions etc.

Bartunov and Sigaev (2007) state that knowledge processing involves an important piece of fact about DL where the digital entities structure corresponds to the relevant data known as meta-data. As a result, these entities are converted into fruitful part of a DL as a block structure. Therefore, practical knowledge constructs a strong structure along with their sub categories e.g., wording contents, pictorial representation, auditory and videotape material. In addition every entity keeps an exclusive pattern identifier. These identifiers include extra knowledge about entities as user privileges and access rules. These rules may correspond to sharing, modification and updating. According to Canela et al. (2007), a major factor behind the DL system is efficiency of retrieval instead of storage material. There are several ways to archive and access the digital data including simple to complex structures. The fundamental structural design segregates entities formation through the innovative originator and saves them in a storage area in the appropriate form. According to De-Smet (2010), the DL implementation phase requires adequacy since it is an expensive and complex task and requires various resources, hence, it is advised to go for best practices. The well-known rules are comprised of a selection of contents and stakeholders, usable design strategies, open access, automated and responsive system, following standards and quality assurance. According to Hatano et al. (2002), DLs provide a multidisciplinary ground like digitization, information extraction, storage management and retrieval equipped with human computer

interaction (HCI) and Artificial Intelligence (AI) tools. Lee et al. (2005) emphasized on various advantages of DLs over the traditional ones, such as accessibility to the user, better searching and browsing, information sharing, information availability and versatility. According to Maria et al. (2004) and McCray and Gallagher (2001), the scope of DLs has been widened to exchanging information, social networking, vigorous exchanging information between scholars, researchers, developers and data analysts etc. Better abstracting, indexing services can be provided. All types of information-related data is stored in DL systems including text, audio, video, image and other types of media. Some famous DL projects are Europeana and GoogleBooks that provide such platforms.

According to Singh (2003), the internet analogy of public libraries i.e. highly reliable, high-quality community services have only recently begun to appear. This is because digital libraries are expensive to create and maintain. A serious obstacle to their creation is the provision of appropriate cataloguing information. Without creating a knowledge base, it is hard to offer the meaningful information retrieval. According to Tenopir (2003), the web-based environment has enabled full-text retrieval of documents as a standard feature. A digital library software, therefore, will not only offer the resulting data (the meta-data records) from a search, but also the full documents themselves from which search keys might also have been used for retrieval. This full-text retrieval feature of digital library software offers major benefits to physical libraries which are typically only indexed by their bibliographic data. According to Atta-ur-Rahman (2013), Shahzadi et al. (2011), Shahzadi et al. (2012), Atta-ur-Rahman (2016), Atta-ur-Rahman et al. (2018) and Faisal et al. (2018) intelligent information retrieval is an emerging area of research for more than two decades. In this type of information retrieval data mining, AI, soft-computing along with Natural Language Processing (NLP) have been investigated to achieve meaningful results from the complex data (semi-structured and/or unstructured).

The basic idea behind this research is to generate a collaborative DL for an educational institute. Primarily, the focus was an investigation of the students' study behavior with regards to library books and the factors related to large data entries about the books, as well as, the space problem. The DL structure with its progressive phase is the collaboration of the users and researchers' response with their personal understanding about the proposed system. The proposed system provides a means of efficient data storage and retrieval with ease. The second aspect of the proposed framework is the structural design with a collaborative intranet platform, creating an XML depository, and efficient searching in institute's

databank. An ETL (Extract, Transform and Load) strategy is proposed to construct the efficient archiving solution for institute's Data Bank. IBM Infosphere Data stage is used as a Data warehouse development tool for building up data archiving jobs. IBM Lotus Domino Environment is used to develop the proposed system. Lotus Dominos provides a custom designed data warehouse and data accessibility solution both locally (offline) and online. The metadata format, design and other structure can easily be customized (programmed) according to the type of data and implementation. Whereas, in Europeana, the client must follow their structure and data representation (metadata) standards etc. Moreover, it is purely an online solution which may not be affective for remote area institutes where internet connectivity is usually an issue. Moreover, being a custom solution, it provides more user-centric front-end design, privacy and security.

The rest of the paper is organized as follows: Section 2 contains proposed system model, Section 3 highlights implementational details, results and discussion while Section 4 concludes the paper.

PROPOSED SYSTEM MODEL

The Institute's Digital library solution has been proposed and implemented via XML as metadata storage and by building the institute data repository in a common platform after incorporating all the resources. XQuery language is used as the solution for searching the digital contents from the DL in a meaningful way. Here XQuery seems a better fit instead of SPARQL that is suitable for Ontological data stored in RDF, a case of semantic web, (Atta-ur-Rahman & Alhaidari, 2018). The application is developed into IBM Lotus domino 8.5 Environment being more suitable for an institute's requirements and provides an excellent platform for collaboration, workflow and communication. The generated DL objects can easily be imported/exported from/to the online resources for the sake of collaboration. Similarly, small repositories may be merged to the main data warehouse that can easily incorporate the diverse sources. ETL processes are designed for better archiving and interoperability in DL using IBM Infosphere Datastage Designer. The main components of the proposed framework are given below.

Digital Library System Architecture

The proposed DL system architecture is depicted in Fig-1. Initially, there are several resources that are added to the system. However, they can be divided into two categories, namely physical data sources and digital data sources. Physical data sources are comprised of sources in hard copies available in traditional library and

manually catalogued. Digital data sources are comprised of e-books, editorials, published articles, conference proceedings, students' theses or some online blogs etc. First, a standard digital format is devised that describes the type of information (metadata), an institute wants to store about a resource. Metadata fields are selected carefully and then data about the resources is transformed in XML format. Obviously, this process was more challenging for the physical resources compared to digital resources.

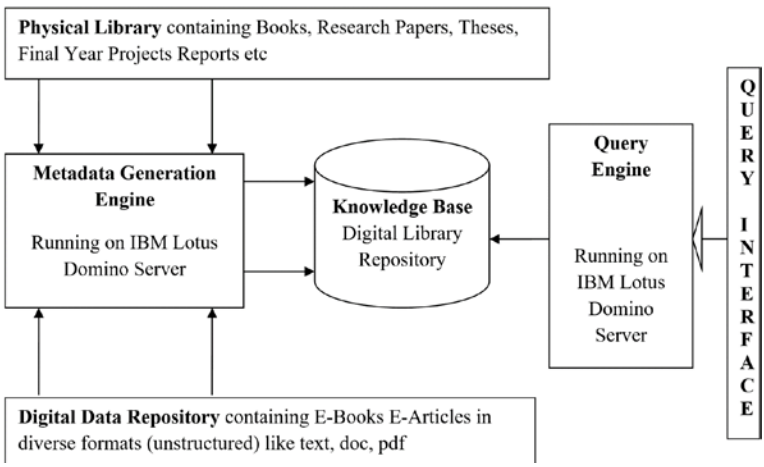

Figure 1. Proposed System Architecture

Meta-data generation. The resources are transformed into digital objects in XML with metadata fields given in Table 1.

Table 1

Metadata Fields in the Institute's Digital Library

Field Name	Purpose
Title	Title of a digital object e.g. Book title, Journal title, Article title, Thesis title, Report title etc.
ISBN/DOI	ISBN assigned to a digital object of type book and DOI in case of published article. In case of student projects and unpublished contents it may refer to Student's Registration number or Faculty ID of the author.
Author (s)	Author (s) of a digital object
Publisher	Publisher information of a digital object
Year of Publication	Year of publication of a digital object

Field Name	Purpose
Description	The description of a digital resource e.g. in case of thesis or article if may be 'Abstract' and in case of books it may be the 'Preface'
Contents	View of contents of a digital resource. It might be in terms of Table of Contents (ToC) etc.
Chapters Information	The names of chapters within a digital resource

Building metadata repository database. Entire metadata repository is converted into digital objects in XML form which is serving as the basis of proposed knowledge-base and has limitless capabilities and strengths for archiving and retrieval of DL. Figure 2 shows a sample objects of DL expressed in XML.

```
<InstituteRepository>
        <Book>
                <BookTitle>The Best Damn Windows Server 2003 Book Period</BookTitle>
                <ISBN>9781931836128</ISBN>
                <AuthorList>
                        <Author>Debra Littlejohn Shinder, Thomas Shinder</Author>
                </AuthorList>
                <Publisher>Syngress</Publisher>
                <PubYear>2004</PubYear>
                <Description>Not Applicable</Description>
        </Book>

        <Book>
                <BookTitle>The Craft of Information Visualization Readings and
                Reflections </BookTitle>
                <ISBN>9781558609150</ISBN>
                <AuthorList>
                        <Author>Benjamin B. Bederson, Ben Shneiderman</Author>
                </AuthorList>
                <Publisher>Morgan Kaufmann</Publisher>
                <PubYear>2003</PubYear>
                <Description>Interactive Technologies</Description>
        </Book>

        <Book>
                <BookTitle>The Designer's Guide to VHDL</BookTitle>
                <ISBN>9781558606746</ISBN>
                <AuthorList>
                        <Author>Peter J. Ashenden</Author>
                </AuthorList>
                <Publisher>Morgan Kaufmann</Publisher>
                <PubYear>2002</PubYear>
                <Description>Systems on Silicon</Description>
        </Book>

        <Book>
                <BookTitle>The Economics of Open Source Software
                Development</BookTitle>
                <ISBN>9780444527691</ISBN>
                <AuthorList>
                        <Author>Jürgen Bitzer / Philipp Schröder</Author>
                </AuthorList>
                <Publisher>Elsevier</Publisher>
                <PubYear>2006</PubYear>
                <Description>Not Applicable</Description>
        </Book>
</InstituteRepository>
```

Figure 2. XML view of DL objects

Application Development Tool

IBM Lotus Domino version 8.5 is the selected tool used for application development due to its collaborative environment and BaseX is used as query engine for XQuery to fetch the queried results from the institute's DL. IBM Infosphere Server along with IBM Infosphere Datastage designer are used as ETL implement tool for sake of executing ETL jobs.

Import Engine. This is a very important part of the proposed framework because the data sources are diverse. For instance, physical data source information is contained in MS Excel files, digital data sources are available in the form of spreadsheet, as well as comma separated format (CSV); moreover, the data from outside the institute. To deal with these diverse data sources, code is written in Lotus script that dynamically imports the data and stores it in the central repository.

User interface. Interface of the application given in Figure 3, is simple to use and provides links to all part of the system for various tasks like adding new source, search engine, administrative controls etc.

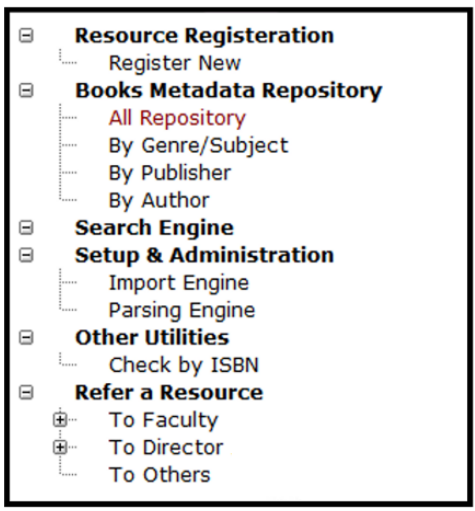

Figure 3. DL Navigation Bar

Registering a new resource. This task can be done through system default setting triggered by "Register New" tab and new source may be added after inserting the required fields.

Repository search. The desired search handle is implemented in BaseX using XQuery for fetching the desired results from the DL. It also provides mechanism to import new sources and database creation etc. The dialogue is shown in Figure 4.

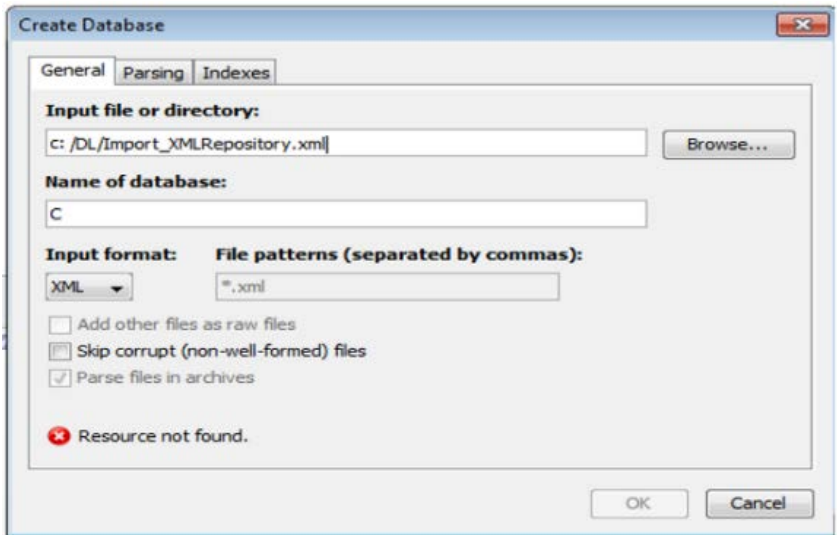

Figure 4. New XML Database in BaseX

For default setting, several choices are given in the parsing Tab (Figure 4) for environmental setup. Figure 5 illustrated all choices for parsing.

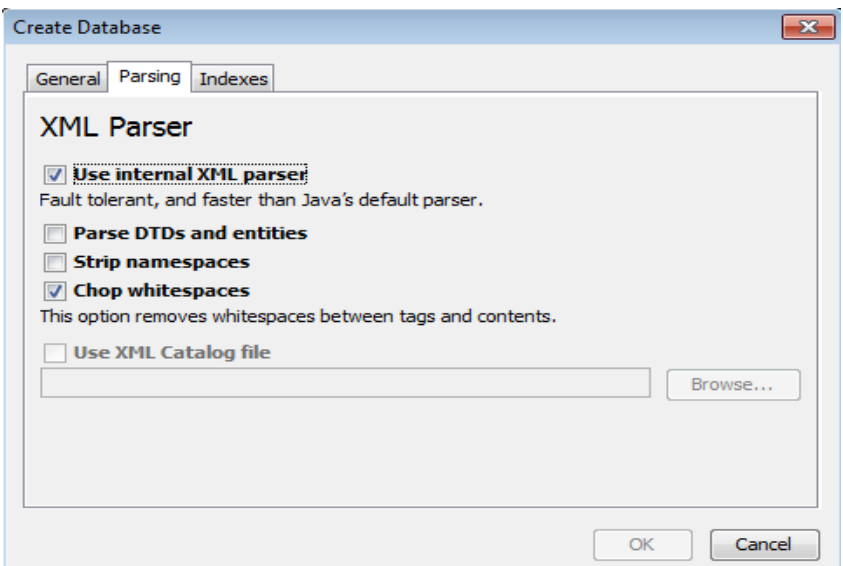

Figure 5. Parsing Tab of New Database Definition in BaseX

BaseX editor has many options related to indexing schemes and these can be applied on upcoming databases during import process. This is shown in Figure 6.

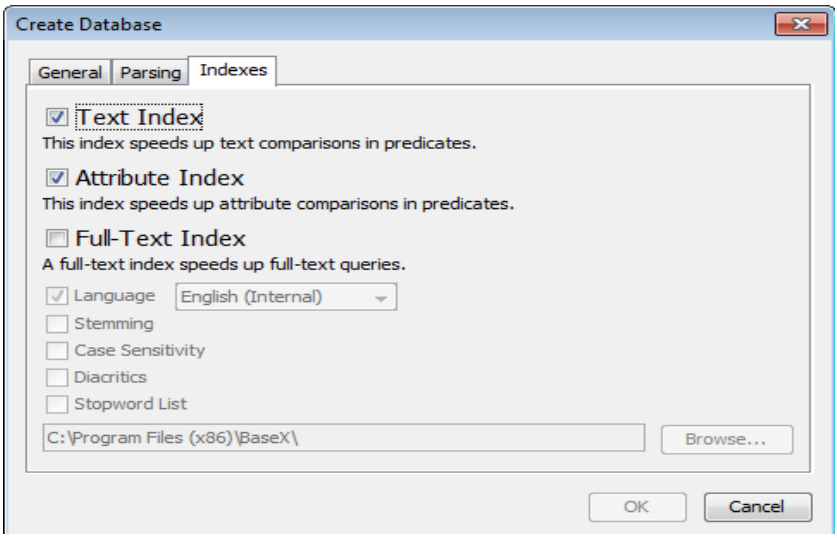

Figure 6. Indexes Tab of New Database Definition in BaseX

The main problem while analyzing the results from XQuery is that the result set is the node found in search query based on the information applied as filters. XQuery just traverse the XML tree and find the nodes in the hierarchy as directed by the query. The information is not sometimes meaningful. Therefore, it is needed to develop an approach to find meaningful information from the XML dataset.

In this paper, we developed an intelligent approach to find appropriate and meaningful information based on some rules developed after closely examining the XML file structure, which is discussed in the next section.

Information retrieval algorithm for the proposed digital library. From a critical analysis of the XML structure, it is observed that the structure of node and its children are an important aspect of object in terms of relevancy. It depends on the domain specific metadata and its tag-structure inside the XML file. Based on these facts, following algorithm has been proposed and implemented in the DL. This algorithm not only retrieves better results but exhibits better time efficiency as well.

Retrieval algorithm.

- *If a node has less than two children that means it has one child or it is a leaf node and it can be deduced that said node describes a specific attribute*

and does not carry useful information. Such node is nominated as Type-I node

- *If the node has two or more children having same name and structure, then it represents a specific dimension and hence does not contain meaningful information. Such node is nominated as Type-II node*

Nodes other than above types are meaningful. Based on the discussion, following rules are built.

If a specific node is in Type-I or in Type-II, the bottom up strategy should be followed that goes on to determine its all predecessor until the node satisfies the condition of being meaningful. Therefore, according to the node's children number, children's name and structure, we can identify the meaningful information.

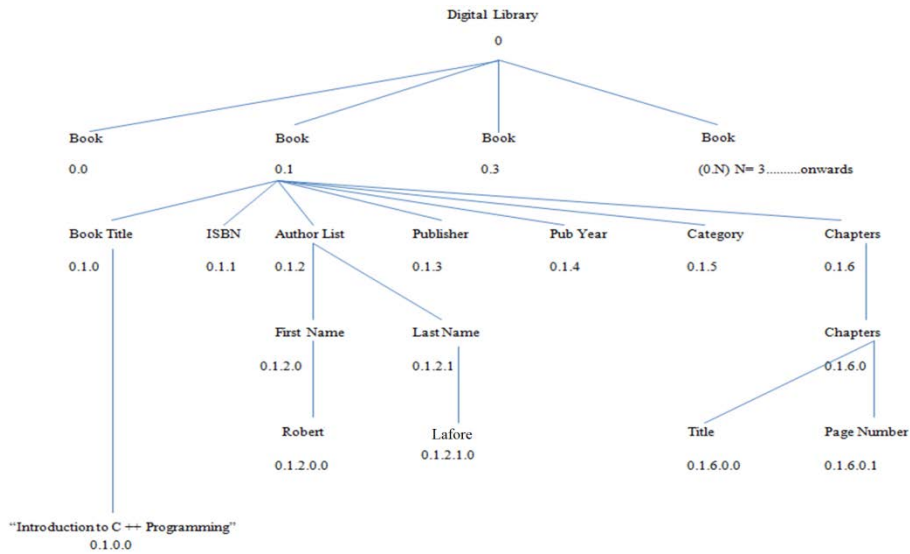

Figure 7. XML Tree of the Repository

For example, let us look at a sample XML tree in Figure 7, the node ID 0.0.1 and ID 0.0.1.0 belong to *type I*, the node ID 0.0.2 and ID 0.0.2.0.3 belong to *type II*, so they aren't enough meaningful nodes, the node ID 0.0 and ID 0.0.2.0 are meaningful nodes.

The model adopted to store the Digital Objects as XML data *T* as a rooted, labelled and unordered tree (like the one shown in Figure 7). Every internal node *v* in *T* has a node name, which in turn provides us with the information regarding a specific resource in the digital library and data value is contained in each leaf node.

A Dewey label is assigned to each node in the XML tree as a unique Identifier, to help locate the node position and help in parsing.

We built the mechanism in the direction of decision, the comparative location for a node with its associations linked these locations through dot "." which is the starting point through the root, to create the Dewey ID as of the node.

As an example, the node 0.1.0.0 corresponds to keyword *Book Title* and the node 0.1 corresponds to the node *Book*. Thus, Dewey ID performs several operations such as sequence detection, nodes associations and information about the root node etc. To make it further clear, following rules are considered.

1. The start tag of a node U appears before that of node V in an XML document if and only if Dewey(U) is smaller than Dewey(V) by comparing each component in order. This is to ensure the sequencing that will help at the time of parsing.
2. A node U is an ancestor of node V if and only if Dewey(U) is a prefix of Dewey(V). This will help finding the ancestors in the tree.
3. A node U is a sibling of node V if and only if Dewey(U) differs from Dewey(V) only in the last component.

To support XML keywords searches, reversed indexes are created. They reflect the hierarchical information suitable for traversing and retrieval. Reversed indexes design is illustrated in Table 2 while every keyword used as an item in the metadata.

Table 2

Inverted Indexes Table

Keyword	Dewey ID
…….	…….
Book Title	0.1.0
Introduction to C Programming	0.1.0.0
ISBN	0.1.1
Author List	0.1.2
FirstName	0.1.2.0
LastName	0.1.2.1

In the proposed mechanism, for meaningful information retrieval the decision criteria are the number of children a particular node possesses. This is a significant information which affects the meaningful information being fetched against a query, therefore we need to capture the number of children the node

possesses in our model and in the Dewey Indexes, keep a track of node name and the number of children it possesses. To get the name of the node and child number according to the DeweyID of the node, Dewey Indexes are created. Its structure is shown in Table 3.

Table 3

Dewey Indexes Table

Node Name, Childcount	Corresponding Dewey ID
Book Title, 1	0.1.0
Introduction to C ++ Programming, 0	0.1.0.0
Author, 2	0.1.1
ISBN, 1	0.1.2
First Name, 1	0.1.2.0
Last Name, 1	0.1.2.1

The main data indexes creation follows the above tables, so it is indexed and sorted data based on the key column. Inverted and Dewey indexes are stored in Lotus Notes Domino 8.5. Domino have a general-purpose embedded database engine capable of providing a variety of data management services. As the XML document structure is proposed, the data indexes are generated and stored. A novel algorithm is proposed that processes keyword search on XML data. The algorithm is based on the approach by Candela et al. (2007) to compute potential nodes which satisfy the query; it enriches its meaningful return information as does XSeek system, but the algorithm is fast and yields greater accuracy and precision in terms of returning meaningful data nodes. A detailed algorithm is described in Figure 8.

Now there is a detailed description regarding the implementation of the algorithm. At first, all the potential nodes are selected from the DL XML Tree, afterwards every potential node which a single existence of a query keyword have at least utilizes the function *getNearest_Meaningful_Node(v)* to determine whether it is a significant result node or not. Besides this, if the result is significant then improvements can be added into the set M, or else search another significant node, loop calculating function to each node in Potential Node list, till the last node, eventually M is returned.

Example

Let's query Q1 (C++ Programming) and pursue the algorithm, the Potential node is the leaf node ID 0.1.0.0, having no child (Type-I), consequently go to its root node *title*, having single child (type-I), further detection for root node, the node article has seven children with no similar names, thus the significant result set is node *article* (ID 0.1.0) which is a tree rooted at article.

Comparison

In XKSearch, the query result is the leaf node ID 0.1.0.0, in XSeek, the result is the tree rooted node article. For the result, we can observe that this result is more informative than XKSearch, however, it has same result as XSeek.

Archiving of data. An archiving solution is proposed and implemented for an Institute's DL by IBM Infosphere Datastage Designer and this belongs to IBM Infosphere Server family used as an ETL in data warehouse domain. It possesses different types of services that can be configured according to the application's needs. For example, there are two ways to improve the speed by parallelism, namely pipelining and partitioning. These are application dependent approaches and their effectiveness may vary over different applications. Both have their own pros and cons depending on the type of ETL services. To make the proposed DL framework more robust, a mixed parallelism architecture is configured, which combines the advantages of both approaches and results in a quicker processing environment. That is one of the main advantages of using IBM Domino over the open source solutions provided by *Europeana* and *GoogleBooks* etc.

RETRIEVAL ALGORITHM

```
Main Method
Begin:
        Q={k1,k2,…,kn};//k1,k2…kn query keywords
        //Select the potential candidate nodes from the XML tree in which there is at least a
        //single entry of query keywords
        S=Select_Potential_Nodes(T, Q);
        M={}; //query result set nodes
        for(each v v∈S)
        {
                v' =getNearest_Meaningful_Node(v);
                //Adding the meaningful results to the result
                M=v' ∪M;       //Union operation
        }
        return M;
End

// Method to find the meaningful nodes

DeweyID getNearest_Meaningful_Node(DeweyID v)
{
            //Get the children count for the node v
        childCount=get_child_Count(v);
        if(childCount < 2)
        {
                //It is a Type-I node we need to get its parent node
                vparent=getparent(v);
                return getNearest_Meaningful_Node(vparent)
        }
        else
        {
                //U is the set of names of all the children nodes of v
                U=getchildName(v);
                if( ui =  uj ( FOR ALL ui , uj  belongs to U ))
                        // if v's all children node has the same name
                {
                        vparent=getparent(v);
                        return getNearest_Meaningful_Node(vparent)
                }
                else
                        return v;
        }
}
```

Figure 8. Proposed Retrieval Algorithm

RESULTS AND DISCUSSION

This section demonstrates the Graphical User Interface (GUI) of the proposed system to highlight different aspects, and further evaluates it in terms of performance. Further the performance of proposed algorithm is compared with others in terms of recall and time. This is obtained from the results of the queries performed on the knowledge base in contrast to the legacy systems. The architecture of the ETL jobs which are designed for data archiving are also presented.

Import Interface

We design a structure for importing process about meta-data which obtains all types of data through Excel sheets, CSV format and other data files to the institute's DL. Script is written in Lotus for import mechanism. Figure 9 shows the dialogue window to import data from a diverse resource to the institute's DL.

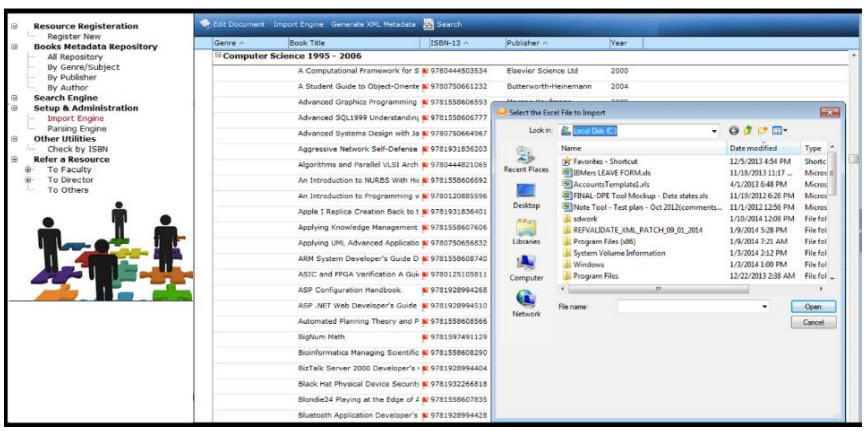

Figure 9. Import Dialog to attach the metadata files

Registration of Digital Object

According to the addition process of a new source, simply click on "Register New", it will open a new menu in which the registration form will open. For the new source, fill all the fields related to the meta-data as shown in Figure 10. After filling up all the form with required fields, click the store button and the resource will be added to the repository. Table 3 contains an example of a book as resource being added to the repository with its attributes, fields and corresponding XML tags.

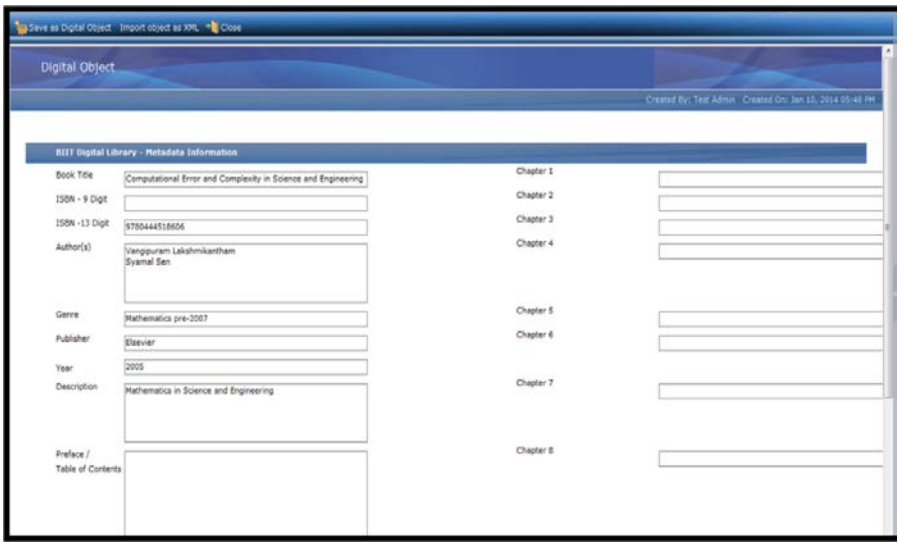

Figure 10. Registering a New DL Resource

Table 4

Mapping of Book Attributes in Digital Library

Registration Field	Corresponding Book Attribute	Corresponding XML Tag
Book	Name of the Book	<BookTitle></BookTitle>
Authors	Author Information	<Author></Author>
Publisher	Publisher	<Publisher></Publisher>
Year	Year of Publication	<PubYear></PubYear>
Description	Table of Contents	<Description></Description>
Book Path	Location	<BookPath></BookPath>
ISBN	ISBN Number	<ISBN></ISBN>
Chapter 1	Title for Chapter 1	<Chapter1></Chapter1>
Chapter 10	Title for Chapter 10	<Chapter10></Chapter10>

XML Transformation

Each attribute of the source being added will be transformed to XML tags and then will be added to the repository. In doing so, it can be searched, traversed and summarized by the query engine during retrieval.

Knowledge Base Creation

Figure 11 shows the process of creating the whole knowledge base in terms of a flowchart.

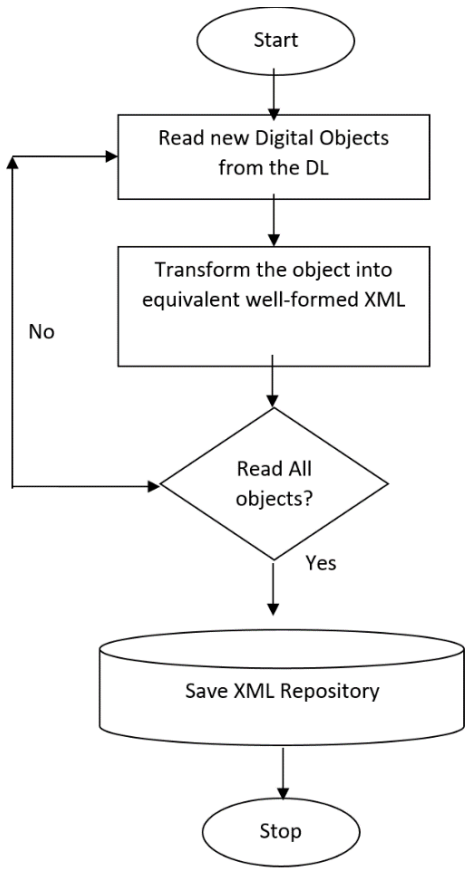

Figure 11. XML Knowledgebase Creation Flowchart

Graphical User Interface

In this section, the GUI of the proposed Digital Library is demonstrated. It is comprised of the navigational links for using the system, registration and reporting screens of the repository. The system is following the client server model and this application is deployed on IBM Lotus Domino Server.

Logging into the digital library. The Digital Library is deployed on Lotus Domino 8.5 which is server client environment. So the main Digital Library will be based on Lotus Domino Server 8.5 whereas the clients work on Lotus Notes Client

8.5. Access to the library is subject to a pre-registration process. It is shown in Figure 12.

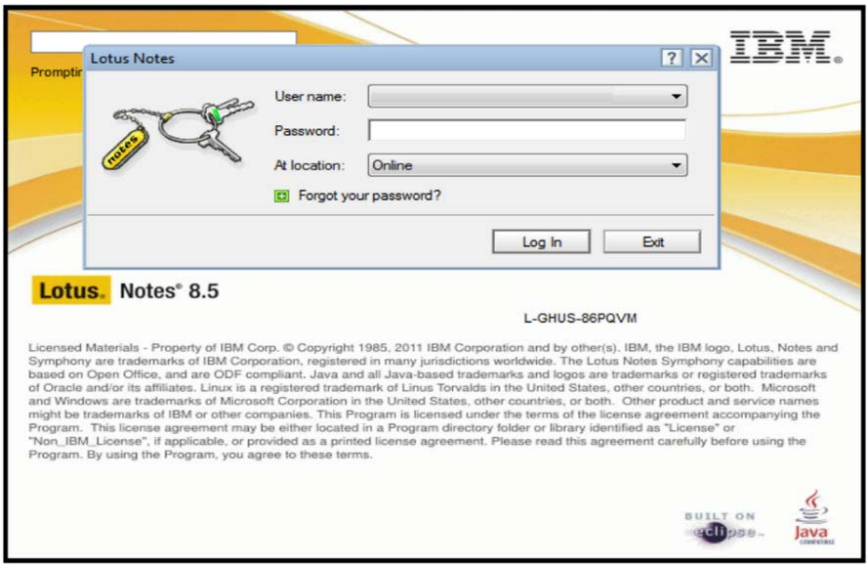

Figure 12: Digital Library Login

Upon successful login, the user will be directed to system area based on the rights allocated to him/her.

Viewing the registered digital objects. The digital objects registered in the digital library can easily be viewed since they are categorized with respect to their attributes. This is an easy way to visualize them. For example, categorization of the digital objects with respect to their genre and publisher is shown in Figure 13 and Figure 14 respectively.

Genre	Book Title	ISBN-13	Publisher	Year
Agricultural and Biological Sciences pre-2007				
	Advances in Plant GlycosidesCher	9780444501806	Elsevier	1999
	Advances in Plant PathologyVolum	9780120337118	Academic Press	1995
	Agriculture's Ethical Horizon	9780123705112	Academic Press	2006
	Air-Breathing Fishes	9780122948602	Academic Press	1997
	Algal Ecology	9780126684506	Academic Press	1996
	Amniote Origins	9780126764604	Academic Press	1997
	Ancient Marine Reptiles	9780121552107	Academic Press	1997
	Animal Cognition in Nature	9780120770304	Academic Press	1998
	Antarcticasoilsweathering process	9780444427847	Not Available	1987
	Applications of Physiological Ecolo	9780124359550	Academic Press	1996
	Aquatic Ecosystems Interactivity c	9780122563713	Academic Press	2003
	Avian Molecular Evolution and Sys	9780124983151	Academic Press	1997
	Bauxites	9780444408884	Not Available	1972
	Beef Cattle Feeding and Nutrition	9780125520522	Academic Press	1995
	Behavioral Ecology of Tropical Bir	9780126755558	Academic Press	2001
	Benguela Predicting a Large Marin	9780444527592	Elsevier	2006
	Bioavailability of Nutrients for Anir	9780120562503	Academic Press	1995
	Biodiversity of Fungi	9780125095518	Academic Press	2004
	Biogeochemistry of Trace Element	9780444519979	Elsevier	2005
	Biology and Culture of Channel C	9780444505767	Elsevier	2004
	Biology of Brassica Coenospecies	9780444502780	Elsevier	1999
	Biology of the Hard Clam	9780444819086	Elsevier	2001
	Biology of the Lobster	9780122475702	Academic Press	1995
	Biotechnology and Safety Assessr	9780126887211	Academic Press	2002

Figure 13. Viewing Digital Objects by Genre

Gulf Professional Publishing			
	Advanced Reservoir Engineering	9780750677332	2005
	Advances in Engineering Fluid Mechanics Multiphase Reactor ar	9780884154976	1996
	Advances in Environmental Control Technology Health and Tox	9780884153863	1997
	Applied Process Design for Chemical and Petrochemical Plants	9780884150251	1995
	Applied Process Design for Chemical and Petrochemical Plants	9780884151012	1997
	Applied Process Design for Chemical and Petrochemical Plants	9780884156512	2001
	Blowout and Well Control Handbook	9780750677080	2004
	Chemical Process Equipment Selection and Design	9780750675109	2005
	Chemical Process Safety Learning from Case Histories	9780750677493	2004
	Chemistry of Petrochemical Processes	9780884153153	2001
	Compressors Selection and Sizing	9780750675451	2005
	Computational Rheology for Pipeline and Annular Flow Non-Ne	9780884153207	2001
	Computer Control and Human Error	9780884152699	1995
	Containment Systems A Design Guide	9780750676120	2003
	Data Reconciliation and Gross Error Detection An Intelligent Us	9780884152552	1999
	Deep Challenge Our Quest for Energy Beneath the Sea	9780884152194	1998
	Developments in Offshore Engineering Wave Phenomena and (	9780884153801	1998

Figure 14. Viewing Digital Objects by Publisher

Evaluation of the Proposed DL

In this section, the results of the information retrieval using the proposed algorithm are presented. We have also compared the results with the well-known schemes in the literature.

Results comparison. The experiential results are conducted on Intel Core i5 system with 2GB RAM in Windows 7 Operating System. The execution of the application is performed in IBM Lotus Domino Client Server 2 tier structural design. Reversed and Dewey directories creation and storage processes performed in IBM Lotus Domino database. During the experimentation, the original data about the institute's books and written material in the form of thesis, articles etc. was added to DL. Nearly 5000 records saved in the DL system and experiment conducted for the evaluation of the proposed algorithm's efficiency through its performance measure along the XKSearch and XSeek (Liu et.al., 2007) using native XQuery. Four types of query sets (Table 5) selected for the algorithm, to check the running time and recall in contrast to above mentioned systems. The recall operation will be performed according to the mentioned formula:

$$recall = \frac{|Rel \cap Ret|}{|Rel|}$$

Wherever Re*l* indicates that appropriate nodes set which returned by the proposed DL system, Re*t* indicates system results (desired search results) and recall measures the percentage of the desired nodes that are output. Figure 15 shows the processing time of three approaches. It can be observed that the proposed algorithm outperforms XSeek algorithm while comparable to XKSearch algorithm in terms of execution time. Figure 16 shows that the proposed algorithm exhibits a higher recall rate compared to both algorithms on the average.

Table 5

Query Set

#	Query Statement
Q1	Title Introduction to C++ Programming
Q2	Robert
Q3	Author Lafore
Q4	Book Author Robert

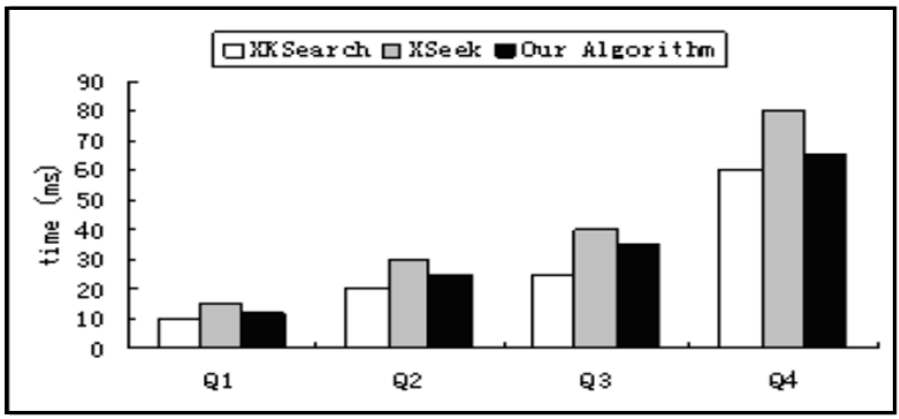

Figure 15. Performance Comparison in terms of Processing Time

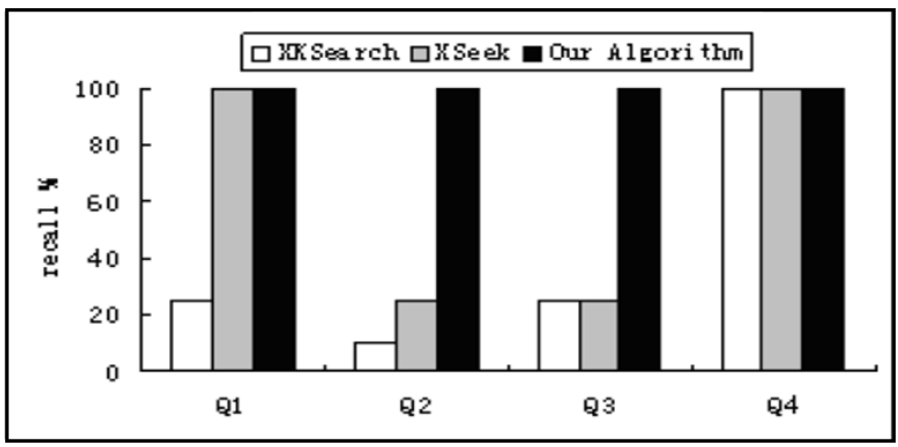

Figure 16: Performance Comparison in terms of Recall Measurement

CONCLUSIONS

In this paper, a digital library system for educational institutes has been proposed, designed and evaluated while focusing on the needs of academic domains. IBM Lotus Domino Client-Server architecture has been utilized to develop the system for internet as well as intranet environment, for the sake of exploiting the capabilities for defining the workflows and live collaboration among the users. XML is used as a transformation platform to import all types of an institute's data resources (books, technical reports, research articles and theses et.) into common standard structural format. The archiving and interoperability are achieved using the ETL processes. A customized retrieval algorithm has been proposed. BaseX is used as a query engine to execute XQuery over the XML knowledgebase and for

performance evaluation of the proposed retrieval algorithm compared to well-known algorithms. The performance results show that the algorithm is promising in terms of execution time and recall factor. While overall framework is ideal for an institute's digital library. In future, the concept can be extended to the next level and aspects of data mining and information extraction may also be investigated.

REFERENCES

Arms, W.Y. (1995). Key Concepts in the Architecture of the Digital Library. *D-Lib Magazine, 13* (17), Reston, Virginia.

Atta-ur-Rahman. (2013). Teacher Assessment and Profiling using Fuzzy Rule based System and Apriori Algorithm. *International Journal of Computer Applications*, *65*(5), 22-28.

Atta-ur-Rahman. (2016). Knowledge Representation: A Semantic Network Approach. *Handbook of Research on Computational Intelligence Applications in Bioinformatics* (1st ed.), Chapter: 4, IGI Global.

Atta-ur-Rahman & Alhaidari, F.A. (2018). Querying RDF Data. *Journal of Theoretical and Applied Information Technology, 26*(22), 7599-7614.

Atta-ur-Rahman & Alhiyafi, J. (2018). Health Level Seven Generic Web Interface. *Journal of Computational and Theoretical Nanoscience. 15*(4), 1261–1274.

Bartunov, O. & Sigaev, T. (2007). Full-text Search in PostgreSQL: A Gentle Introduction. *Moscow University Press*, 1, Moscow, Russia.

Candela, L., Castelli, D., Pagano, P., Thanos, C., Ioannidis, Y., Koutrika, G., Ross, S., Schek, S., & Schuldt, H. (2007). Setting the Foundation of Digital Libraries. *D-Lib Magazine, 13*(3/4).

De-Smet, E. (2010). Some ISIS-software history and technical background to the new FOSS integrated library system 'ABCD'. *Liber quarterly: the journal of European Research Libraries / European Research Libraries, 19*(3/4), 324-335.

Faisal, H.M., Ahmed, M., Asghar, S., & Atta-ur-Rahman. (2017). Intelligent Quranic story builder, *International Journal of Hybrid Intelligent Systems*, preprint.

Hatano, K., Hiroko, K., Yoshikawa, M. & Uemura, S. (2002). Information Retrieval System for XML Documents. *DEXA '02 Proceedings of the 13th International Conference on Database and Expert Systems Applications*. 758-767. LNCS. ISBN: 3-540-44126-3.

Lee, G.T., Dahlan, N., Ramayah, T., Karia, N. & Asaari, M.H.A.H. (2005). Impact of Interface Characteristics on Digital Libraries Usage. *Malaysian online Journal of Instructional Technology, 2*(1), 1-7.

Liu, Z., Walker, J., & Chen, Y. (2007). XSeek: A Semantic XML Search Engine using Keywords. *In Proceedings of the 33rd International Conference on Very Large Data Bases* (VLDB '07). 1330-1333, ISBN: 978-1-59593-649-3.

Maria, M.I., Azevedo, L., Amorim, P., & Ziviani, N. (2004). A Universal Model for XML Information Retrieval. *INEX'04 Proceedings of the Third international conference on Initiative for the Evaluation of XML Retrieval*, 311-321. LNCS. ISBN:3-540-26166-4 978.

McCray, A.T. & Gallagher, M.E. (2001). Principles for Digital Library Development, *Communications of the ACM*, *44*(5), May, New York, USA.

Shahzadi, N., Atta-ur-rahman, & Shaheen, A. (2011). Semantic Network based Semantic Search of Religious Repository. *International Journal of Computer Applications*, *36*(9), 1-5.

Shahzadi, N., Atta-ur-rahman & Sawar, M.J. (2012). Semantic Network based Classifier of Holy Quran. *International Journal of Computer Applications, 39*(5), 43-47.

Singh, S. (2003). Digital Library: Definition to Implementation, *Lecture Delivered at Ranganathan Research Circle*, Delhi.

Tenopir, C. (2003). Use and Users of Electronic Library Resources: An overview and Analysis of Recent Research Studies. *Council on Library and Information Resources*. Washington, D.C. ERIC Number: ED499383