

Design and Development of a Publication Warehouse

M. S. Ahmed*, J. Sadique¹, M. A. Zulquarnain¹, M. K. Newaz and M. M. Islam²

Institute of Computer Science, Atomic Energy Research Establishment, Savar, Dhaka, Bangladesh

¹*Bangladesh Atomic Energy Commission, Agargaon, Dhaka, Bangladesh*

²*Health Physics and Radioactive Waste Management Unit, Institute of Nuclear Science & Technology*

Atomic Energy Research Establishment, Savar, Dhaka, Bangladesh

Abstract

It is a dire need of time to have a centralized database of published scientific/technical papers of the scientists of Bangladesh Atomic Energy Commission. This article presents a data warehouse which is developed to store research papers of the scientists. The developed publication warehouse is hosted in an internet server and remains open for all. Researchers around the globe will be benefited by reading these research papers.

Keywords: abstract, publication, database, data warehouse

1. Introduction

A digital library is usually a research platform containing full-text publications including journals, conference proceedings, etc., as well as an interlinked set of connections among authors, works, institutions and specialized communities. A digital library can be explored by browsing, searching or advanced searching [1]. An enterprise data warehouse serves as a system of record and enables an organization to create a single and integrated version of the truth of the enterprise data. The concept of the single source of truth is considered to be the most accurate [2]. Scientists of Bangladesh Atomic Energy Commission (BAEC) publish research papers in various international/ national journals and conference proceedings. BAEC wanting to have a web based system which would archive all these papers and thus initiated a conceptual project entitled “procedure for submitting title & abstract of published scientific/technical paper for publishing on the BAEC website (BAEC/SOP-004)”. A web based system is thus expected to be developed so that by following an official procedure any author from BAEC can submit their scientific/ technical paper(s) to publish in the online site. This focuses on developing a publication warehouse which is useful to search published papers.

2. Materials and Methods

2.1 Database Formulation

The process of developing a software system usually requires careful planning in order to complete all necessary activities. Requirement specification is an essential part of any software development. Data warehouse methodologies can be broadly classified based on the design approach: top-down, bottom-up, hybrid or federated [3]. We however followed the top-down approach by identifying the detailed business processes and prioritizing them. However detail requirement specification was not there to start with. Thus, business rules are assumed for the development of the publication warehouse.

2.2 Business Rules

A journal can have one or more paper, a paper is published in one and only one journal, an Institute can have one or more author, an author is posted to one and only one Institute, an author can write one or more paper, a paper can have one or more author, a paper must have one and only one corresponding author, a paper is published under one and only one Institute, an Institute can have zero or more paper.

From the business rules it is apparent that entities or tables required to develop the database are: author, journal, paper, institute and publication. Each entity usually has several characteristics or attributes. Also a good general practice is to make attribute atomic. A data dictionary is expected to contain all the attribute names and characteristics for each table in the system [4].

2.3 Data Dictionary

A detailed description of all tables for the proposed data warehouse is provided in Table 1.

For the Author table, instead of making Author name more atomic (like first name, middle name and last name), a single attribute (Author Name) is considered for ease of searching name. Pap Authors attribute in Paper table is however used to record all authors name as written in the original paper. For the Publication table, although publication ID (PubID) is not the primary key, unique auto increment numbers are used for ease of editing. Most of the attributes of all the tables are considered obligatory. Only the month the paper published and volume number in Paper table are made optional.

2.4 Entity Relationship Diagram

The business rules imply that the Journal table has one-to-many relationship with the paper table. The Institute table has one-to-many relationship with the author table. The Institute table also has one-to-many relationship with the paper table. But the author table has many-to-many relationship with the paper table. And thus to avoid this many-to-many relationship, the publication table is

*Corresponding author: shakil97@yahoo.com

considered. There is also a one-to-one relationship between a paper and its corresponding author. Fig. 1 below is the Entity Relationship Diagram (ERD) that reflects the conceptual schema of the system. MySQL Workbench [5] is used to draw the ER diagram.

Table 1: Data dictionary of publication warehouse

Table name	Attribute name	Contents	Type	Required	PK or FK
Author	AuthorID	Author ID	varchar(5)	Yes	PK
	AuthorName	Author Name	varchar(40)	Yes	
	AuthorEmail	Author Email	varchar(40)	Yes	
	InstShort	Institute Code	varchar(10)	Yes	FK
Journal	JournalID	Journal ID	varchar(5)	Yes	PK
	JournalName	Journal Name	varchar(200)	Yes	
	JournalPublisher	Journal Publisher Name	varchar(100)	Yes	
	JournalCountry	Journal Country Name	varchar(15)	Yes	
Institute	InstShort	Institute Code	varchar(10)	Yes	PK
	InstLong	Institute Name	varchar(60)	Yes	
	InstPlace	Institute Place	varchar(12)	Yes	
Paper	PapEntryNo	Paper Entry No	varchar(6)	Yes	PK
	PapEntryDate	Paper Entry Date	Date	Yes	
	PapTitle	Paper Title	varchar(220)	Yes	
	PapAuthors	Paper Authors Names	varchar(220)	Yes	
	AuthorID	Author ID	varchar(5)	Yes	FK
	JournalID	Journal ID	varchar(5)	Yes	FK
	PapMonth	Paper Publication Month	varchar(3)	No	
	PapYear	Paper Publication year	year(4)	Yes	
	PapVol	Paper Volume No	varchar(5)	No	
	PapPage	Paper Pages	varchar(10)	Yes	
	PapAbstract	Paper Abstract	Text	Yes	
	PapDiscipline	Paper Discipline	char(1)	Yes	
Publication	InstShort	Institute Code	varchar(10)	Yes	FK
	PubID	Publication ID	int(11)	Yes	
	PapEntryNo	Paper Entry No	varchar(6)	Yes	PK/FK
	AuthorID	Author ID	varchar(5)	Yes	PK/FK

PK= Primary Key; FK= Foreign Key; Varchar = Variable character length data; Char = Fixed character length data; Text = Variable-length character strings; Int = Integer value only

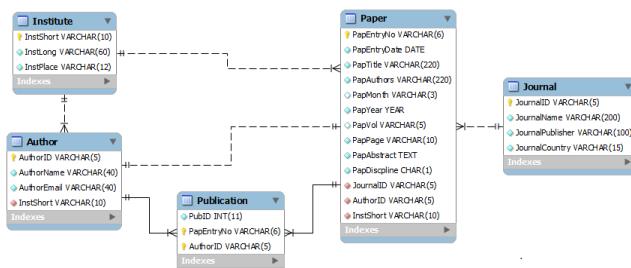


Fig. 1: Entity relationship diagram

2.5 Development and Deployment

In order to ensure project success, it is necessary to be

pragmatic when choosing the software to be used. A free tool, WAMP, was used to develop the web based application. WAMP is an application server used to run web application locally. Apache web server, MySQL and PHP are integral part of this windows based tool [6]. The backend database is developed from the ER mentioned above. The database was created using MySQL server which is bundled with the WAMP distribution. Structured Query Language (SQL) is used to develop business logics and create various reports from backend database. The key programming language used to realize the implementation is PHP programming language. *PHP* is a widely-used open source general-purpose scripting language that is especially suited to web development [7].

The Web application has been implemented to have two major modules: User module and administration module. A number of search options are there to come up with the user module. These are title search, author search, institute search and discipline search. Any user can freely perform these search functionalities. Detail of a paper can be seen by clicking the title in a list of publication. Add, edit and delete options are provided only for administrative purpose. The application has a login page though for the administrator.

The development of the project has initially been carried out on a Netbook computer, which runs on windows XP operating system. The software was developed as local host and later deployed online to a hosting service provider. The web application is hosted on an internet server, so it can be accessed through any standard or mobile web browser via the internet. The main user page of the publication Warehouse is shown in Fig. 2 below. The fact that web applications are not necessarily platform dependent is a big advantage over stand alone applications.



Fig. 2: Main page of publication warehouse

3. Results and Discussion

The purpose of conducting this research project is to verify that it is possible to make a repository of published research papers of the scientists of BAEC so that anyone interested can easily read a paper online. Four types of search options are provided for user: title search, author search, institute search and discipline search. For title search, user needs to enter a keyword in the title. The keyword can be one or more words. The system usually consider a keyword as a

string of character, then compare it with all the titles in the warehouse and then come up with a list of matched papers. User than can click on the desired title and see the detail of the paper. For author search, one needs to enter name of an author or a part of it. It returns a list of author/co-author containing the name. Then user can click on the name to see the list of his/her publication. For institute search, a user can select any institute from a drop-down menu. After submitting a request, the system comes up with a list of publication which are published under the institute.

The system also has option to search on discipline. Papers are divided into two disciplines: Physical Science/Engineering and Bio-Science. After submitting a query for a particular year, the system returns a list of publication published under that discipline. Usually there are cross-references in any list although abstract of a paper can be seen by clicking the title of the paper in the list. In addition to these search options, admin can see the current list of author, journal, paper, etc. after logging in the administrator page. Admin has also privilege to add, edit and delete from any list. Initial testing found that all the functionalities discussed are properly working.

4. Conclusion

The publication warehouse is on its beta phase and deployed on an internet server (www.baecbd.org/dbpublication/publication.php). The warehouse is developed in a way so that any user easily can search and find a required research paper and read the abstract. A link is also provided in the home page of BAEC website (www.baec.gov.bd) so that researches from home and abroad will be benefited.

References

1. The ACM Digital Library, Available: <http://dl.acm.org>, Accessed: 27/10/2015.
2. K. Krishnan, Data Warehousing in the Age of Big Data, Morgan Kaufmann (2013).
3. A. Alazmi, Data Warehousing Implementations: A Review, International Journal on Data Mining and Intelligent Info. Techno. Appl., **4**(1), 9-25 (2014).
4. C. Coronel, S. Morris and P. Rob, Database Systems: Design, Implementation and Management, 10th Ed., Course Technology, USA (2012).
5. MySQL Workbench, <http://www.mysql.com/products/workbench/>, Accessed: 10.10.2015
6. Wampserver, A Windows Web Development Environment. <http://www.wampserver.com/en/>, Accessed: 22.09.2016
7. PHP:Hypertext Preprocessor, <http://php.net>.