International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 10, October 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

IMPLEMENTATION OF A ROBUST SEARCH ALGORITHM TO MINE DATA FROM ELITERATURE DATABASE

Dr.G.Charles Babu[1], Dr.K.Ratna Babu[2]

[1]*Professor, Department of CSE, Malla Reddy Engineering College (A), Hyderabad, Telangana*

[2]*Lecturer, Department of Computer Engineering Government Polytechnic, Addanki, Prakasam Dt.*

ABSTRACT

Mining data from databases as seen many upgrades since few decades considering the development of huge data repositories with the advent of internet revolution worldwide. Inherently, the importance of search algorithms to mine data has gained prominence over a period of time. Many programming languages have been used to retrieve relevant information from databases. In this paper, we present the eliterature database created in MySQL with all possible entries such as ISSN, Publisher name, publication type etc are incorporated. Authors from different geographical regions can also be searched from the database. The search algorithm code implemented in the work is used to search the database with varied options such as 'Abstract', 'keywords', 'affiliation', 'country', 'ISSN' etc. Each search option and the relevant code were written in php. Binary search algorithm has been implemented in the work to perform search routine. Apart from general search option, a robust search method which combines various search combinations called 'combination search' can be used to efficiently mine data.

Keywords: eliterature, data mining, text mining, binary search algorithm, combination search

I. INTRODUCTION

The rapid growth of the web in the last decade makes it the largest publicly accessible data source in the world. The amount of data/information on the web is huge and still growing. Data of all types exist such as structured tables, semi-structured web pages, unstructured texts, heterogeneous data, hyperlinked texts etc [1] [2].The World Wide Web has witnessed esteem for its capability of storing huge amount of data, wherein millions of such repositories are available online. Data mining is also called knowledge discovery in databases (KDD). It is commonly defined as the process of discovering useful patterns or knowledge from data sources, e.g., databases, texts, images, the Web, etc [3]. Data retrieval aims at retrieving all objects which satisfy the defined conditions and consists mainly of determining which documents of a database contain the keywords in the user query [4].

Databases are not just confined for depositing and data retrieving medium but can also be used for analyzing data from huge repositories. It also aids research to analyze the data with the help of data mining algorithms which delivers an role being played by database technology in the data mining process [5]. Much research has been devoted in the area of text-mining since few decades, where the main intention was to explore and train on considerable knowledge from huge data repositories. Internet has become the excellent mode to disseminate the

wealth of information related to the topic. Owing to this perspective, search algorithms that efficiently extract either exact or related data to the user have gained much prominence [6]. Though information is perceived from online sources, many programming languages have been used to retrieve the relevant data from huge databases. Here, we report eliterature database, created in MySQL and implemented binary search algorithm concepts to mine abstract related information including journals, year, keywords, abstract etc. The rationale behind the work is based on the huge volume of information in journals publishing manuscripts having fewer or limited search options. Hence, a specific, broad, intuitive search options are implemented in this work to represent robust initiatives of search algorithms.

II. TYPE STYLE AND FONTS

In order to develop a local database, abstract related data was extracted from PubMed database [7] and Google Scholar [8]. PubMed is a freeware of National Center for Biotechnology Information (NCBI) maintained at the National Library of Medicine (NLM). PubMed provides with accessibility for ease in searching certain topics using generic mechanisms, using MeSH terms, publisher's name, title, patterns, phrases, names of publications. Google Scholar was used to retrieve subject specific data and the extracted information was stored in eliterature database.

A. Database architecture in MySQL

```
-- Table structure for table `eliterature_table`
--

CREATE TABLE IF NOT EXISTS `eliterature_table` (
  `SNo` int(11) NOT NULL,
  `Journal_ID` varchar(50) NOT NULL,
  `Structure` longblob,
  `Journal_Name` text,
  `ISSN` int(11) NOT NULL,
  `Publisher` text,
  `Pub_type` text NOT NULL,
  `Article_title` text NOT NULL,
  `Authors` text NOT NULL,
  `Affiliation` text NOT NULL,
  `Country` text NOT NULL,
  `Volume` decimal(10,0) NOT NULL,
  `Issue` decimal(2,0) NOT NULL,
  `Page_nos` int(11) NOT NULL,
`Abstract` text NOT NULL,
  `Keywords` text NOT NULL,
  `Impact_factor` int(11) NOT NULL,
```

```
 `Year` text NOT NULL,
 PRIMARY KEY (`Journal_ID`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Dumping data for table `eliterature_table`
--
INSERT INTO `eliterature_table` (`SNo`, `Journal_ID`, `Structure`, `Journal_Name`, `ISSN`, `Publisher`,
`Pub_type`, `Article_title`, `Authors`, `Affiliation`, `Country`, `Volume`, `Issue`, `Page_nos`, `Abstract`,
`Keywords`, `Impact_factor`, `Year`) VALUES
```

B. Extracting data from online databases

In order to insert specific data in local database, PubMed, Google Scholar etc are searched for the presence of keyword, 'data mining'. However, it was observed that most relevant data was obtained from Google Scholar but not from PubMed. PubMed database has much of data related to biology and hence using keyword 'data mining' resulted in more biological papers which are not significant to this study.

Moreover, it should be noted that the PubMed database is very huge with lots of information wealth necessary to any scientist or researcher working in all fields of science. As the study has been restricted to data mining, biological aspects of mining are only considered for extraction from PubMed database.

In the next step, Google Scholar was searched for the presence of keyword, 'data mining' without any limitations on search query. Limitations such as year, date and relevance can be supplied to the query. This search resulted in many hits which are more than expected and far from PubMed database which suggests the fact that this database is very huge and score more than PubMed.

III. RESULTS AND DISCUSSION

Mining data from database requires defining search algorithm to mine data using the user supplied specified input of keywords. Literature data, in general is huge and enormous in size owing to the submission of research and review papers being published in various journals of scientific significance. It should be noted that the data is also diverse and more informative. Hence journals which publish such information from authors have been diversified from general to more specific based on the field of study. Hence, to mine information, it is deemed to be necessary to implement robust and user friendly search functions. Search scripts are written in java in most of the cases. As we can see from available online databases, php and java scripts are used.

Therefore, in this study, a robust search mechanism implementing javascript and php features are used to mine textual information from local database.

The eliterature_table was created in MySQL database with all possible entries and the image is given in Figures1 and 2. Entry items such as ISSN, Publisher name, publication type (research/review/case study) are incorporated. Authors from different geographical regions can also be searched from the database. For example, if a work on 'IP networking' has to be searched from authors representing a country, this can be made possible using the database to retrieve entries from specified country. When the results are displayed, the user will have an option to choose the number of hits and can download the data as a single zip file.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 10, October 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

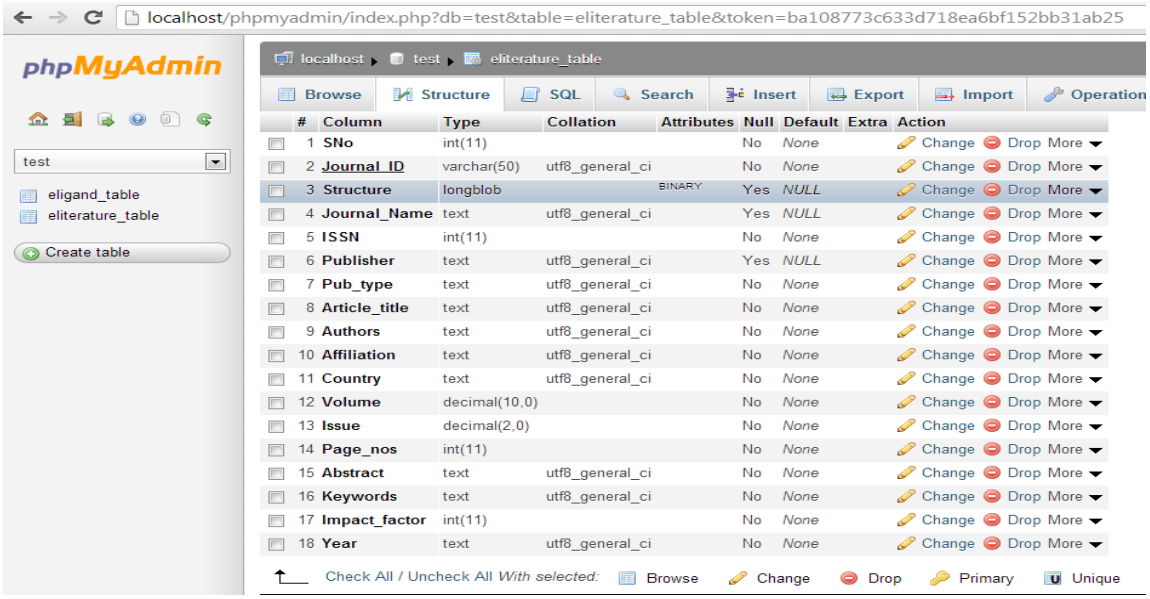

Figure 1. Database 'eliterature_table' structure created in MySQL

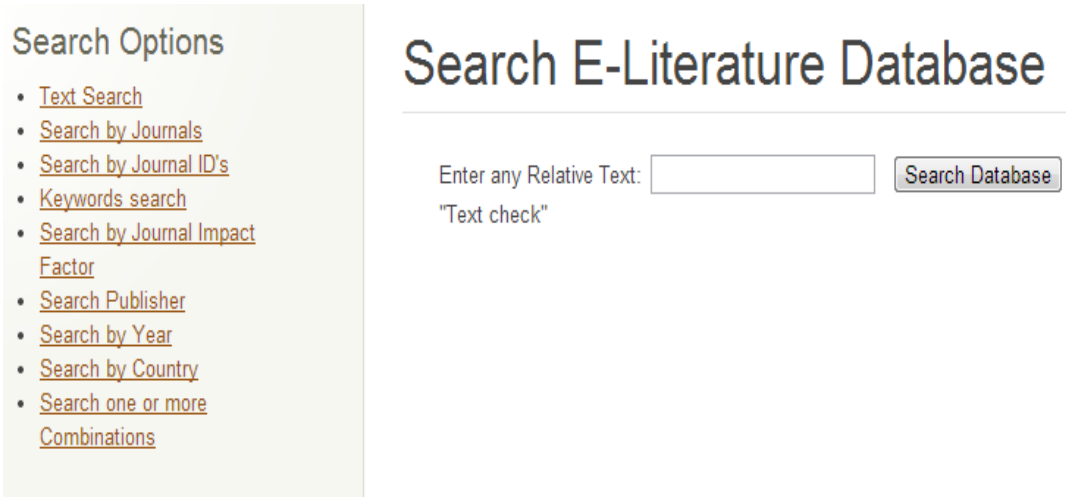

Figure 2. Database search page

An image of combination search as seen in html page is given below

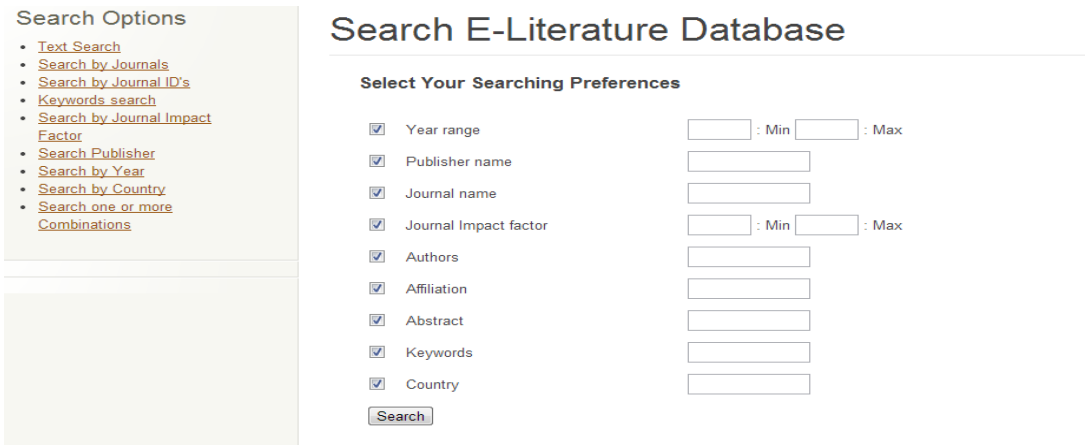

Figure 3.Combination of search terms in eliterature

A Survey on Security Challenges in Cloud Computing

Authors	Dr.K.Ratna Babu, Dr.G.Charles Babu
Affiliation	Lecturer, Department of Computer Engineering, Government Polytechnic, Addanki, Prakasam Dt, A.P.India. Professor, Department of C S E, Malla Reddy Engineering College(Autonomous), Hyderabad, Telangana, India.
Country	India
Article_title	A Survey on Security Challenges in Cloud Computing
Journal_Name	International Journal of Innovative Research in Science, Engineering and Technology
ISSN	2319-8753
Year	2016
Volume	5
Issue	5
Page Numbers	7022-7028
Abstract	The importance of the cloud is increasing exponentially and people start realising the reliability, scalability, and efficiency of the cloud computing. In recent generation since around 2008 the hype increasing like anything. Dramatically the cloud size increasing, and the problems also in the same fashion. Despite the potential advantages of cloud the organizers are slow down accepting the cloud services provided by service providers. In cloud the organizations store their data by handover it to service providers or third party who own the infrastructure. In this way the data is most vulnerable in cloud network it is obvious that there is lots of security issues need to be concern both for the normal client and business oriented client. Although the service providers guarantee the security in terms of technical aspects, but it is very difficult to achieve clients trust. In the present paper a review on security issues, how to challenge them and cloud manageability is presented.

Figure 4. Image showing abstract output

From the above, it is evidenced that combination search which is the most robust method implemented in this project can be used to mine data with much ease as the method is simple and user friendly. For example, an author can mine data based on specified years published on particular keywords originating from a specific country of origin can reveal the importance and observable data on the specified keyword

IV. CONCLUSION

Retrieving information from databases either online or standalone has seen implementation of improved search techniques using various programming languages. The eliterature database presented here is one of its kind where binary search algorithm was implemented to create a robust and efficient search option to retrieve data. When compared to PubMed or Google scholar, the database has more options to search and even an user can utilize combination search to narrow down the results. Further work is in progress to compare efficiency of algorithms implemented in eliterature, PubMed and Google Scholar.

REFERENCES

[1] Bing Liu. Web Data Mining Springer-Verlag Berlin Heidelberg (2007) pp.6 ISBN-10 3- 540-37881-2

[2] A. Arasu and H. Garcia-Molina. Extracting Structured Data from Web Pages. In Proc. of the ACM SIGMOD Intl. Conf. on Management of Data (SIGMOD'03), Pp.337–348, 2003

[3] Bing Liu. Web Data Mining Springer-Verlag Berlin Heidelberg (2007) pg.6 ISBN-10 3-540-37881-2

[4] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval.Addison–Wesley, 1999

[5] M. Agosti, G. Gradenigo and P. Marchetti. A hypertext Environment for interacting with large textual databases. Information processing and Management, 28(3): 371- 387, 1992

[6] Adele Howe and Danielle Dreilinger. Savvysearch: A metasearch engine that learns which search engines to query. AI Magazine 18(2): 19-25, 1997

[7] http://www.ncbi.nlm.nih.gov

[8] http://www.scholar.google.com