

Knowledge Discovery in Databases (KDD): A Comparative Evaluation of Scientific Databases

Divya Mishra and Ravindra Kumar

Guest Faculty, Department of Library and Information Science
Babasaheb Bhimrao Ambedkar University, Lucknow, Utter Pradesh, India
E-mail : divya.mishra67@gmail.com, kumarravindra324@gmail.com

(Received 30 June 2017; Revised 12 July 2017; Accepted 10 August 2017; Available online 16 August 2017)

Abstract - In this information explosion age, a large number of commercial and free online database provided by publishers of information resources is available on web, Libraries of every kind offering various services regarding use of online resources and services to fulfill the information requirements of a large group of users. The present study comparatively analyze the selected databases which aims to serves a scientific community. The Library Science and information personnel all over the world are focusing more and more on development of better, user friendly and affordable discovery solutions to fulfill the requirements of patrons.

Keywords: Knowledge discovery, WOS, PubMed, Scifinder

I. INTRODUCTION

Discovery tools are the latest state of work tools to start a discovery project, make necessary modifications in between the execution and also to arrive at a definite conclusion of any research work. In the Ocean of information that the internet and web have created, no appropriate information can be retrieved without the help of some discovery tools. Due to this reason every sphere of knowledge generation has some very definitive discovery tools. A range of techniques are exist to help out in extorting blueprints that when inferred offer important and previously unknown, insight into the stored data. The raw data is considered as a collection of elements from which little knowledge can be gleaned and Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data. The phase of KDD and Data mining can seize on several forms, the choice discoverer on the needed discovery results. KDD is a multi stair course of action that facilitates the alteration of data into constructive information. The unifying goal of the KDD process is to extract knowledge form data in the context of large databases

A. Work of KDD System

1. Developing an Understanding of Application Domain, Relevant Prior Knowledge and the aim of the end patrons;
2. Creating a target data set by selecting data set, subset and its variables on which discovery is to be performed
3. Cleaning and processing of data;
4. Projection and reduction of data to represent the useful and specific knowledge related to the goal of task;

5. Choosing data mining task of KDD process like classification and clustering etc
6. Searching for pattern of interest in a particular form;
7. Interpreting mined pattern.
8. Consolidating discovered knowledge.

B. Databases Discovery Services

Discovery services are designed to provide a simple Google-like search box, which enables library users to search all of the library's resources with a single query. Some Initiatives of Database Discovery Services are as follows.

1. NISO Open Discovery Initiative (ODI) by IEEE
2. EBSCO Discovery Service (EDS) by EBSCO
3. ExLibris Primo by IEEE
4. ProQuest Summon by IEEE
5. OCLC WorldCat Discovery Services by Library of congress
6. Google API by google

II. PURPOSE OF THE STUDY

We thought of making a preliminary study to know how it can function and what kind of resources and services provide by them. The study evaluates the knowledge discovery and refining options in databases and services offered by scientific databases to find the relevant information required by them and also the awareness rate among users regarding them.

III. SCOPE OF THE STUDY

The Study is restricted to only the scientific databases namely Pubmed, Scifinder, Scopus and Web of Science (WOS) subscribed by special libraries.

S.No.	Database Name	Publisher
1	PubMed	United States National Library of Medicine (NLM)
2	Scifinder	American Chemical Society
3	Scopus	Elsevier
4	Web of Science	Thomson Reuters

IV. MATERIAL AND METHODS

Data related with study are collected through accessing the selected scientific databases in libraries which subscribed it. Thus, collected and collated the data in order to find out information with an aim to the following:

- a. Discovery Fields Provided by Databases
- b. Discovery Refining options in Databases

- c. Reference Management Tools provided by Databases
- d. Citation Indicators in Databases
- e. Other Services Provided by Databases

V. RESULT AND DISCUSSION

To realize the objective of the study, data was collected accessing the selected scientific databases on various aspects of knowledge discovery and refining options in databases and services offered by scientific databases.

TABLE I DISCOVERY FIELDS PROVIDED BY DATABASES

S.No.	Discovery Fields	PubMed	Sci Finder	Scopus	WOS
1.	Author	√	√	√	√
2.	Group Authors	√	√	√	√
3.	Last Author	√	√	√	√
4.	Editors	√	√	√	√
5.	Source Title	√	√	√	√
6.	Book Series Title	√	√	√	√
7.	Conference Title	√	√	√	√
8.	Publication Year	√	√	√	√
9.	Organization	√	√	√	√
10.	Funding Agencies	√	√	√	√
11.	Languages	√	√	√	√
12.	Countries	√	√	√	√
13.	ISBN/ISSN	√	√	√	√
14.	Journal Name	√	√	√	√

In the present age numerous type of publications are available and patrons search it according to their memory and requirement. The above table represent the all the search field provided by databases.

TABLE II DISCOVERY REFINING OPTIONS IN DATABASES

S.No.	Discovery refining options	PubMed	Sci Finder	Scopus	WOS
Article Type					
1.	Review	√	√	√	√
2.	Note	√	√	√	√
3.	Meeting Abstract	√	√	√	√
4.	Proceeding Paper	√	√	√	√
5.	Letters	√	√	√	√
6.	Editorial Material	√	√	√	√
7.	Book Chapter	√	√	√	√
Text Availability					
8.	Abstract	√	√	√	√
9.	Free Full text	√	√	√	√
10.	Full text	√	√	√	√
Publication Date					
11.	-	√	√	√	√
Organization Name					
12.	-	√	√	√	√
Country Name					
13.	-	√	√	√	√

Table 2 shows the discovery refining options offered by databases to arrange searched information according to their need.

TABLE III REFERENCE MANAGEMENT TOOLS PROVIDED BY DATABASES

S.No	Reference Management Tools	PubMed	Sci Finder	Scopus	WOS
1.	End Note	×	×	×	√
2.	Sci planner	×	√	×	×
3.	Mandely	√	×	×	×
4.	Zotero	√	×	√	×

The above table reveals that all the selected scientific databases had reference management tools.

TABLE IV CITATION INDICATORS IN DATABASES

S.No	Citation Indicators	PubMed	Sci Finder	Scopus	WOS
1.	A Index	×	×	×	×
2.	G Index	×	×	×	×
3.	H Index	√	×	√	√
4.	I10 Index	√	×	√	√
5.	Z Factor	√	×	√	√
6.	Formula Index	×	√	×	×
7.	Subject Index	×	√	×	×
8.	Substance Index	×	√	×	×
9.	Patent Index	×	√	×	×
10.	Article Level Matrics	√	√	√	√

In the regard of measuring the productivity of scientific community evaluation indicators are used which are based on citations and journal impact factor. The above table

represent that all the selected databases provide evaluation indexes like h, i10, z factor etc.

TABLE V OTHER SERVICES PROVIDED BY DATABASES

S.No	PubMed	Sci Finder	Scopus	WOS
1.	Single Citation Match	Get Substances	H index	H index
2.	Batch Citation Match	Get Reaction	Impact factor of journal	Impact factor of journal
3.	-	Get Related Citation	Keep Me Posted Alert	Reference Management Tools
4.	-	Get references	-	-
5.	-	Markush Search	-	-

The above table shows the additional services provided by scientific databases. Table clearly indicates that all the selected databases provide additional services related to the subject coverage.

V. CONCLUSION

Knowledge Discovery in Databases is the process of searching for hidden knowledge in the immense quantity of data that we are technically able of generating and storing. Data, in its raw form, is simply a collection of elements, from which little knowledge can be gleaned. With the development of knowledge discovery methods the significance of the data is considerably enhanced. The Library Science and information personnel all over the world are focusing more and more on development of better, user friendly and affordable discovery solutions to fulfill the requirements of patrons.

REFERENCES

- [1] U. Fayyad, and E. Simoudis, "Knowledge Discovery and Data Mining Tutorial MA". *Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1997.
- [2] U. Fayyad, S. Piatetsky- and P. Smyth, From "Data Mining to Knowledge Discovery: An overview." *Advances in Knowledge discovery and Data Mining*, pp. 1-36, 1996.
- [3] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth., "The KKD Process for Extracting Useful Knowledge from Volumes of Data." *Communications of the ACM*, Vol. 39, No.11, pp. 27-34, 1996.
- [4] D.J. Hand, *Discrimination and Classification*. Chichester, U.K.: John Wiley and Sons, 1981.
- [5] E. Simoudis, Reality Check for Data Mining. *IEEE Expert*. pp.26-33, 1996.
- [6] J. Way, and E.A Smith, "The evolution of Synthetic Radar Systems and Their Progression to the EOS SAR". *IEEE Trans. Geoscience and Remote Sensing*. Vol. 29, No.6, pp. 962-985, 1991.
- [7] <https://www.ncbi.nlm.nih.gov/pubmed/>
- [8] <http://www.cas.org/>
- [9] <https://www.scopus.com/>
- [10] <http://online-wosa.gov.in>