



ORGANIZING AND RETRIEVING OF SCIENTIFIC ARTICLES USING ONTOLOGY

P. Maheswari¹, Dr. A. Jaya²

¹Research scholar, B.S. Abdur Rahman University,

²Professor and Head of the Department of Computer Application ,
B.S. Abdur Rahman University

Email: pmaheswari@live.in¹, jayavenkat2007@gmail.com²

Abstract

Nowadays, Researchers are emerging in various domains. They are in need of lot of research articles for reference on time to time. Browsing through web and categorizing the research articles are tedious task, time consumption, Document process. Also the researchers may come irrelevant materials to provide them with required information. The research focused on retrieving and organizing articles from web. This research works facilitate the researchers to browse the articles in a systematic way. Efficient organisation and retrieval of those articles is required. Using ontology, different domains are mapped with various similar concepts. Topic Index and Decision Tree Classification are used to classify the research articles to the scholar and retrieve the relevant article according to the query given by the scholar. The proposed framework supports the domain expert through the whole process of knowledge discovery from data integration to exploration. This research work supports the researchers in order to provide organizing and semantic search.

Index Terms: Decision Tree Classification, Domains, Ontology, Topic Index.

I. INTRODUCTION

With the growing of digital paradigm the use of web resources is increased among researchers. Web resources transformed the traditional libraries to hybrid one. Nowadays lots of lots researches are going enormously, researchers are in need of scientific articles for reference. Most

of the articles are stored in electronic form in a database. Scientific articles may be structured or semi structured. There have been a great deal of studies on modelling and implementation of semi structured data in recent research.

New research scholars may start doing their research in their respective domain. They may not have a finite idea about their domain, whereas in the web repository lots and lots of articles are available which may miss lead them or may come to a conclusion. This research work aims to help the research scholar to classify their research articles from the repository using the query. This research resolves the searching time since it is a domain specific it is time consuming to gather the needed information to the scholar.

II. LITERATURE SURVEY

In previous work on Information retrieval [1] the book discusses that Information retrieval consists of software programs that facilitate the user to find the information the user needs. IR system minimize the overhead of the user locating needed information. [2] This paper uses the technique a new hypertext resource discovery system called a Focused Crawler. The goal of a focused crawler is to selectively seek out pages that are relevant to a pre-defined set of topics. The topics are specified not using keywords, but using exemplary documents. focused crawler analyzes its crawl boundary to find the links that are likely to be most relevant to crawl, and avoid irrelevant regions of the web. [3] This paper SIEU (Semantic Information Extraction in University Domain) uses ontology as a knowledge base for the information retrieval

process. The advantages of the university domain it is not just a mere keyword search. The query is analyzed both syntactically and semantically. [4] This paper shows how filtering algorithm, tokenization, stemming, association rule mining techniques are used. Only the abstract were used for scanning. On the basis of the keyword extraction rules set and they applied further analysis steps of architecture; they found the relevant medicines for the disease based on the semantics. [5] This paper retrieve the documents according to the user's preference by using the knowledge model seekers as per the interest of the user.[6] This paper describes the methodology and software tools to build a database on the careers and productivity of academics. The software provides data crawling and data mining techniques used to transform web page based information and CV information into a relational database. It is important to differentiate between voluntary and forced mobility, as only the former is associated with higher research performance.

III. PROPOSED APPROACH

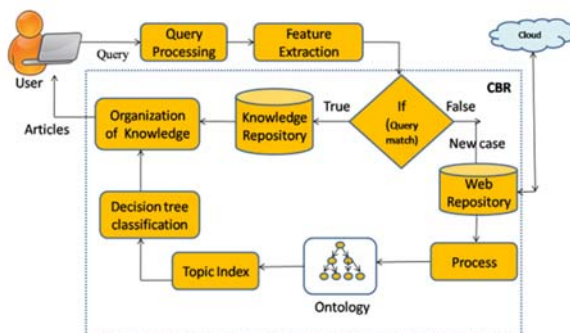


Fig.1 Conceptual diagram

The above figure 1 gives the conceptual diagram. This research work comprises of the following phases:

1. Query Processing
2. Feature Extraction
3. Web repository
4. Cloud
5. Process
6. Ontology
7. Topic Index
8. Decision Tree Classification
9. Knowledge repository
10. Organization of knowledge

A. Query processing

The first phase of the research work is Query processing. Query is received from the user is process into an internal form.

B. Feature Extraction

Feature Extraction is the second phase of the research work. It reduces the original data into a new data. Removal of stop words, Like is, at they, etc., and Tokenization are performed in the second phase of research. If the query given by the user matches then the control goes to the Knowledge Repository else the control searches for the new case through Web Repository.

C. Web Repository

Web Repository is the third phase of the research work. Web repository received the web pages from the crawler and from the cloud. If the query doesn't matches then the control searches the articles from the web repository.

C. Cloud

The fourth phase of the research work is cloud. Cloud is used to store and accessing the article over the Internet. The control searches the articles which are not available in the web repository .The cloud is used to make the proposed framework dynamically.

D. Process

The fifth phase of the research work is Process. Process is used to transform the data into a clear format and also filter in order to remove the noise data and irrelevant data and produce the clear format. Example: Title, Author name, Publication date etc.

E. Ontology

Ontology is the sixth phase of the research work. Ontology is used the define classes and sub classes. It separates the domain knowledge from the operational knowledge. Analyze the domain knowledge and reuse it. The proposed frame work supports the domain expert through the whole process of knowledge discovery from data integration to exploration. Example: A biologist and a programmer may use the same query" Virus" with different search context but the search system would result by re ranking result incorporating the user –interest. If the user specified the domain then the result would appropriate to the query search by the user.

F. Topic Index

Topic Index is the seventh phase of the research work. Users can group similar content by its topics and also reduce the dimensionality of text to the most important features. Topic has an ontological type. According to the topics it is hierarchy. Assign the class label to each article from a set of topic.

G. Decision Tree classification

The eighth phase of the framework is decision tree classification. Decision tree builds classification models in form of tree structure. It breaks down a dataset into smaller and smaller subset. The result is a tree with decision nodes and leaf nodes. The decision tree can handle both numerical and categorical data.

H. Knowledge repository

Knowledge Repository is the ninth phase of the research work. The repository is designed to get knowledge about the articles. If the query matches then the control goes to the knowledge

I. Organization of Knowledge

Categories are used to organized documents thematically into broad areas like Data Mining, ontology whereas title are described in specific topics like "A knowledge based schema matching system for ontology mapping". Articles that belong to two or more categories are unusual. A category can be assigned based not only on the document content, but also on its styles.

IV.EXPERIMENTS



Figure 2 Domain specifications into mind map according to the user

Figure 2 depicts the previous research work using Mind map domain specification. Mind mapping is a free hand drawing therefore classifying the research articles is tedious work

whereas the proposed research framework using ontology retrieves 83% of research articles compared to mind map.

V.CONCLUSION AND FUTURE ENHANCEMENT

This research framework provides to browse the scientific articles in a systematic way. It retrieves the research articles according to the domain expert. New research scholar can have the finite idea about their domain .it is also time consumption is less. The future works involves integrating retrieval algorithms also increase the relevance of the research articles.

VI.REFERENCES

- [1] R.E. Moore, Gerald Kowalski, "Information Retrieval Systems Theory and Implementation", Interval analysis (Englewood Cliffs, NJ: Prentice-Hall, 1966). (ISBN: 978-0-7923-9926-1 (Print) 978-0-585- 32090-8 (Online) Volume 1 1997).
- [2] Soumen Chakrabarti, Martin van den Berg and Byron Dom, "Focused crawling: a new approach to topic-specific Web resource discovery", Elsevier Science B.V, 1999.
- [3] Swathi Rajasurya , Tamizhamudhu Muralidharan , Sandhiya Devi,Dr.S.Swamynathan, "Semantic Information Retrieval Using Ontology In University Domain", arXiv preprint aXiv:1207.5745(2012).
- [4] R.Amirtharathna, "Information Retrieval:Identifying Medicines to Diseases Using Filtering Algorithm", International Journal Of Research In Advance Technology In Engineering (IJRATE),Volume 1, Special Issue, October 2013.
- [5] Ms leena More, "Analytical study of Information retrieval techniques and modified model of search engine",International Journal on computer science and engineering(IJCSE) vol 7No 6 June2015 ISSN :0975-3397
- [6] Aldo Geunaa,Rodrigo Kataishi,Manuel Toselli, Eduardo Guzmán,Cornelia Lawson,Ana Fernandez-Zubieta,Beatriz Barros, "SISOB DATA EXTRACTION AND CODIFICATION: A TOOL TO ANALYZE SCIENTIFIC CAREERS" ELSEVIER, The New Data Frontier ,Volume 44, Issue 9, November 2015, Pages 1645–1658.