

Available online at <http://www.mecspress.net/ijeme>

An Efficient Approach for Web Mining using Semantic Web

Md. Motiur Rahman ^a, Ferdusee Akter ^a

^a *Chittagong Veterinary and Animal Sciences University, Chittagong, Bangladesh*

Received: 19 March 2018; Accepted: 08 June 2018; Published: 08 September 2018

Abstract

The volume of data on the Web is increasing rapidly. The rapidly increased data in Web have brought an urgent need to develop a method to organize that data. At the same time, the level of user expectation of getting precise data is increased highly. Hence, it is tough to satisfy the user satisfaction through the existing system. In this paper, we proposed a model to organize the large volume of data over the Web and retrieve the more relevant data to the user. As an implementation of the proposed model, we built two demo search engine (one for RDF based semantic searching and another for existing searching). We use two different sets of data for testing. For every set of data, the RDF based searching returns more precise data than existing searching. The efficiency of the proposed model is better than the existing searching strategies. In the proposed model, we considered both traditional web and RDF based ontology library to organize the data effectively.

Index Terms: RDF, Semantic Web, Web Ontology, Web Mining.

© 2018 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

Semantic web technologies were primarily designed before the rise of the Web 2.0 technologies for organizing large volumes socially contributed content. The World Wide Web is mostly accessible to humans, whereas machines have a very sketchy knowledge of its content [1]. The main vision of the semantic web is that the content available on the web or offered by the web should be understandable by computers and can exploit the standard meaning. The large volume of linked data available on the web has made a clear advert that in the near future, a web representation may hold the human readable and machine understandable information[1- 2].

Semantic web technology will operate in two major ways: it adds structure to user data or organizes the user data using web mining and connects the existing storage of data that describe the Web 2.0 landscape [3]. Web

* Corresponding author.

E-mail address: motiur@cvasu.ac.bd, ferdusee2008@gmail.com

mining is the process of applying data mining techniques to automatically discover and extract the meaning from web contents and features. Although web mining is the branch of data mining, it is not similar to data mining [4]. To organize the unorganized data and to structure the unstructured data web mining is an important strategy. Web mining acts on different ways. Web content mining is the process of extracting the web contents and retrieves the useful information. Here web content adverts to the web pages that are designed to satisfy the user query. It includes text, audio, video, image and so on [4-5]. Web structure mining (WSM) is the process of forming relation among web pages as nodes and hyperlinks as edges. WSM brings all web pages under a structured shape. Usage mining is an application of data mining that finds the useful usage pattern from web content [6]. Hence it can understand the needs of web based applications. Text Mining is another part of data mining that deploy data mining technique to extract information from a collection of web documents and Web Ontology Language (OWL) is used to mine the meaning of the text in the web. OWL is a new language that introduces for representing ontologies in the semantic web. OWL includes many features to represent language. To represent data about category of objects and the interrelation among data, OWL was initially designed [7]. In the field of semantic web, OWL is designed to play a tremendous role. To automatically process information like as intelligent agents and to access information OWL plays a key role. Specifically, OWL is hoped to bring the huge volume of data into a structured format and to provide structured vocabularies that exhibit the interrelation among different terms and hence the intelligent agent can interpret their meaning. The OWL can represent objects of the classes and the property values of each object. Equivalence statements about classes and properties of objects, disjoint statements on classes, equality- inequality can be made between objects [2, 8]. Our case study to bring data into structured format is shown in Fig. 1.

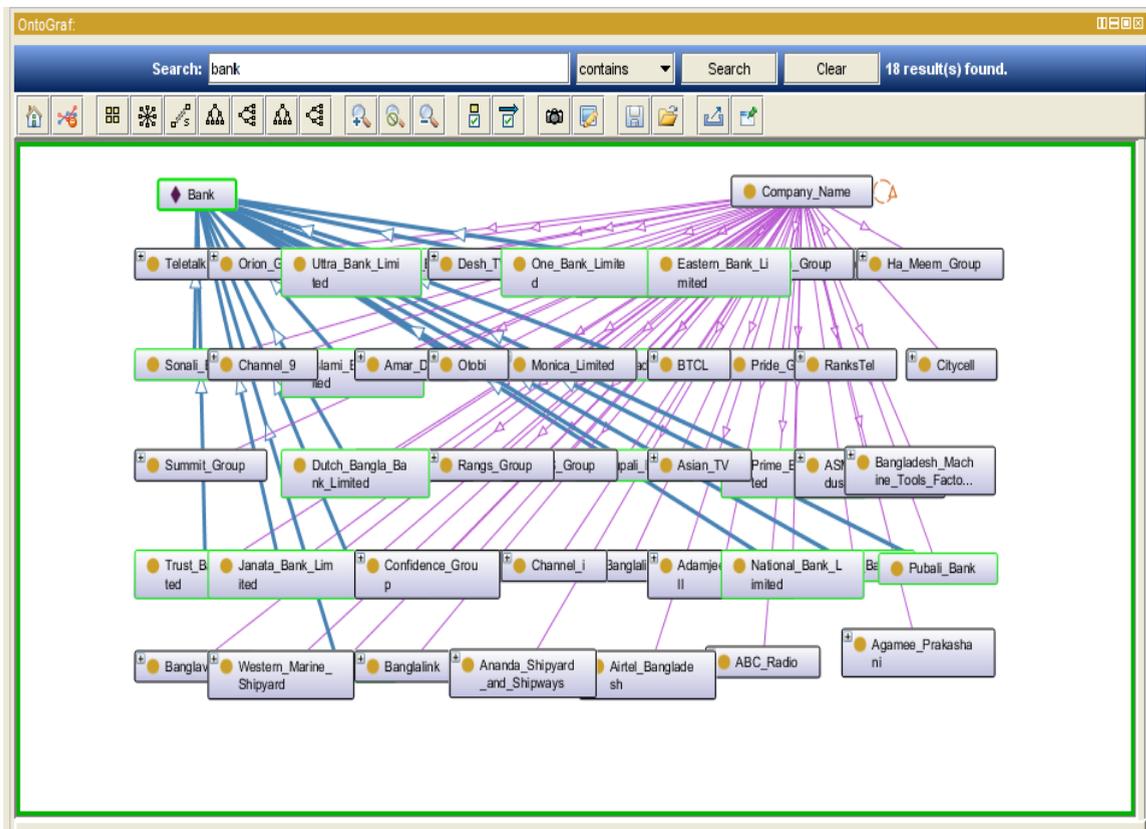


Fig.1. Ontology Graph of Banks under Study

Here we take some Banks of Bangladesh and divide them into classes and subclasses. To organize these classes and objects in a subclass or hierarchy OWL takes the help of Resource Description Framework (RDF). RDF is a language that has been designed for providing a mechanism to describe the web contents and interrelation between these web contents. RDF has been designed to organize the unorganized web contents, structure the unstructured Web contents [9]. RDF offers several useful features for web mining. A key feature of RDF is that it uses International Resource Identifiers (IRIs) – a generalization of Uniform Resource Locator (URL) to advert the resources. RDF directly allow to reference non- local resources which facilitates the information to integrate [9-10]. The underlying data structure of RDF is labelled as directed graph and it is syntactically form triple [13]. Triple is an edge that consists of three components named subject, predicate and object. The triple represents an edge labelled with predicate joining two nodes (subject and object). Triple describes a binary relationship between subject and object using predicate [16-18].

2. Related Work

Semantic Web is getting more popular because of its some interesting feature of data organization. It is a mesh of data that can process data by computer more efficiently than human. However, it is not so easy to implement the idea of Semantic Web. In order to bring the era of Semantic Web into light several research work have been done around the world. Some of the notable work have been presented here.

Kaur N. et al. [3] reviewed five ontology development editors like Apollo1.0, SWOOP 2.3Beta4, Protégé5.0, Graffoo 1.0 and Neon 2.5.2 and compared with their updated versions. They also performed Comparison of two main data models ontology and RDBMS. They also considered classification of ontology languages from those reported in the Literature, with a special attention accorded to the interoperability between them.

Galib S M. et al. [7] focused on data organization and a way to get more pr écised data. To accomplish this task they proposed a method and used a simulation tool to represent how data will organize in Semantic Web. Their contribution also include the integration of K-Means clustering algorithm to get more relevant data. However, they did not provide any proof in favor of proposed model.

Rettinger A et al. [8] studied the applicability of machine learning in Semantic Web and performed some experiments to present their compatibility in Semantic Web. They covered similarity and distance-based methods, kernel machines, multivariate prediction models, relational graphical models and first-order probabilistic learning approaches for mining the Semantic Web statistically.

Singh R et al. [15] proposed semantic web mining for an ERP application based on educational domain. The proposed system helps to find suitable semantic data related to students, faculties and courses for the clients. They also made comparison between traditional query and ontology based query. Their proposed ontology based system performed better than traditional query.

Bhatia C S et al. [16] introduced Grammatical Rule Extraction Technique to extract Semantics through the process of ontology learning. They exploded a new semantic structure and that improve the performance of system in the sense of relevancy.

Meirong T et al. [17] work focused on proving agent-based framework for mining semantic web contents employing clustering techniques. Clustering will help provide user with query relevant cluster of web contents, which will better satisfy user requirement.

3. Methodology

The nature of the data in the existing web is not well organized and unstructured. The returned result against a query is not satisfactory. Most of the time, it returns a huge amount of data as results and most of them are irrelevant. The non-technical people who are not familiar with the existing web technology face a lot of problems during the searching. To make web technology more user-friendly and reduce the limitations of web technology, we propose a model that is shown in Fig. 2. To make web technology more user-friendly and reduce the limitations of web technology, we propose a model that is shown in Fig. 2. This proposed model

operates as follows:

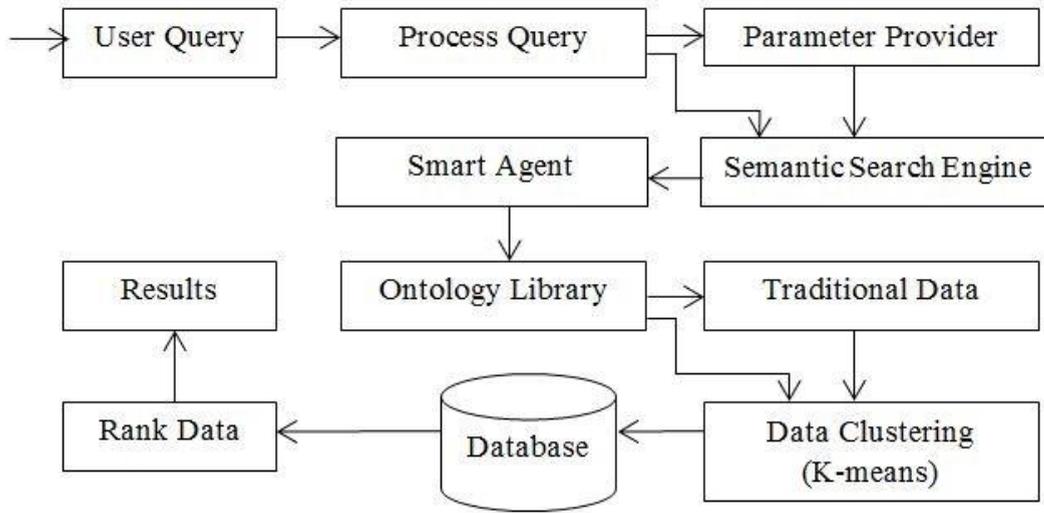


Fig.2. Proposed Model

The user enters his query which is processed in the query processor. The query processor checks whether the parameter is sufficient. If the parameter is enough to search, then the query is passed to the Semantic Search Engine. Otherwise, it is passed to the parameter provider which suggests relevant queries. The proposed model is based on an agent and semantic search engine invokes the smart agent to process the user query. The smart agent searches the result in the Ontology Library that contains RDF based data and also in the traditional database. Then the returned result is sent for clustering in order to obtain more precise results. The clustered result is passed to the database for temporarily storing, then ranking algorithm to rank the results and finally passes the results to the user.

4. Result and Discussion

For the support to our Proposed Model, we have created two demos Search Engine, one of Keyword based and other for RDF based searching. To create Keyword based and RDF based searching, we use PHP, HTML, MYSQL and PHP, RDF, HTML. We used two Databases with the same volume of data, and applied both searching strategies to it. The snapshot from keyword based searching is as shown below in Fig. 3.

As we see from the Fig. 3 that the keyword based searching has returned 8 results against our searching parameter. From the outcome, we can see that most of them are irrelevant. That is, from all 50 data in the database, 8 of them are returned by search engine. Similarly, when we commenced the search with same query parameter in RDF based search, we see only 3 results are returned as shown in Fig. 4. The volume of data in database for RDF search is same as in the keyword based search engine. From the figure we can also observe that the relevancy of data is quite efficient compared to keyword based search. The comparison of above two snapshots gives a strong support to our proposal. We further tabulated the - number of observed data from above results for statistical calculation. In Table 1, we have maintained the number of returned results against the total number of data in the respective databases of each search engine.

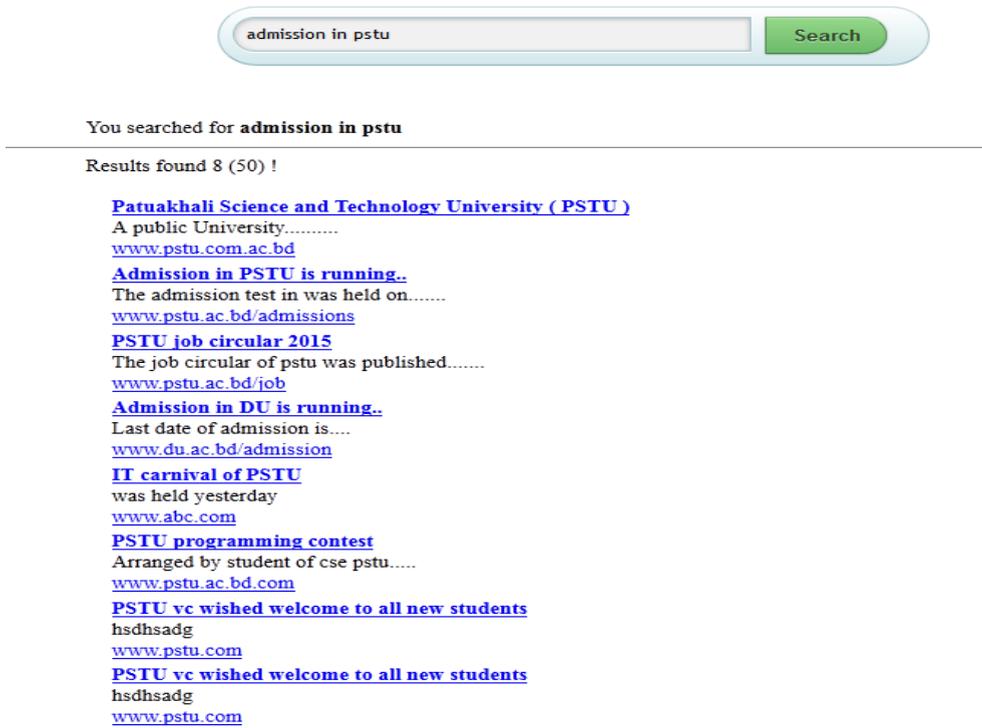


Fig.3. Snapshot of Keyword Based Searching

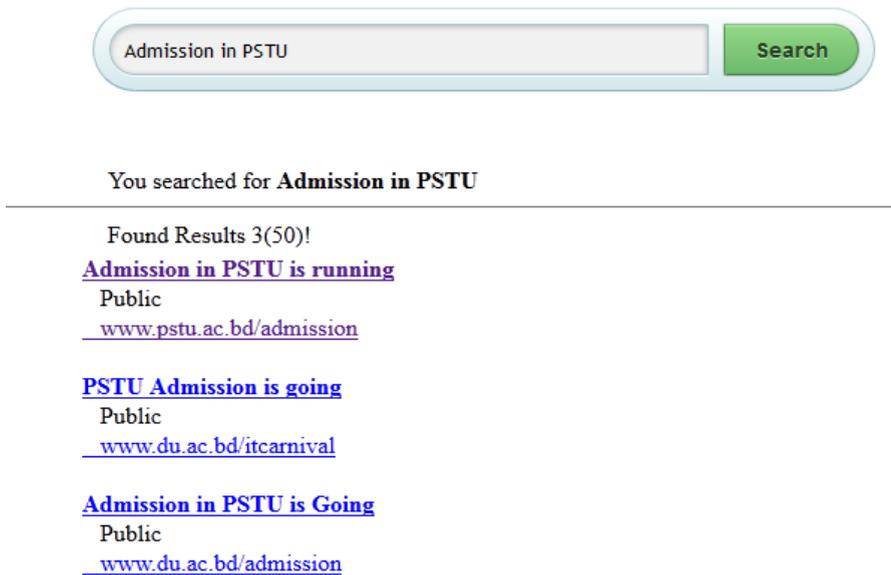


Fig.4. Snapshot of RDF base d Searching

From the table, we can see that, 6 results are returned in keyword based search and 1 in RDF based search for same total number of 20 data in the database. For 50 data in the database, the number of results returned by keyword based search engine is 8 and by RDF based search engine is 3. Similarly, for 100 data in the database, keyword based search engine returns 18 results whereas, RDF based search engine returns 7 results. Here the relevancy between outputs from two search engine is in favor of RDF based search engine.

Table 1. Results from Keyword based and RDF based Search

No. of Data in Database	Returned Results	
	Keyword Based	RDF Based
10	5	1
20	6	1
30	12	2
40	9	4
50	8	3
60	29	6
70	15	8
80	45	12
90	32	15
100	18	7

Table 2. Results from Keyword based and RDF based Search

No. of Data in Database	Returned Results	
	Keyword Based	RDF Based
10	3	2
20	5	2
30	14	5
40	12	6
50	6	3
60	21	4
70	19	8
80	27	13
90	29	15
100	12	10

The comparison of returned results from keyword based search engine and RDF based search engine is graphically shown in Fig. 5. Here, the number of data in the database against results returned is plotted. The blue bar represents the results from keyword based search while the red bar represents the results from RDF based search.

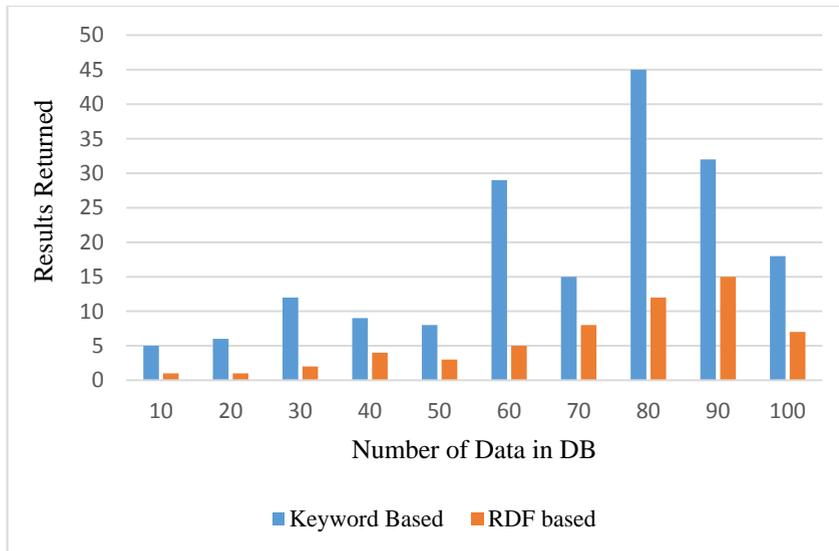


Fig.5. Comparison between Keyword based and RDF based Search

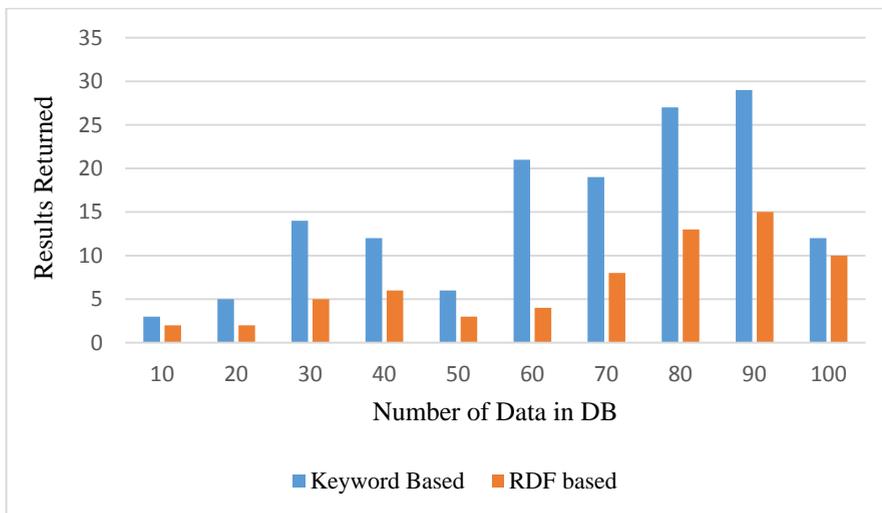


Fig.6. Comparison between Keyword based and RDF based Search

We also perform analysis on another group of data and the observed results are shown in Table 2. From Table 2, we see that for every set of data the RDF based searching perform better than keyword based searching. The comparison of returned results from a keyword based search engine and RDF based search engine on a different set of data is graphically shown in Fig. 6. Here, the number of data in the database against results returned is plotted. The blue bar represents the results from keyword based search while the red bar represents the results from RDF based search. From Fig. 5 and Fig. 6, we see that the RDF based searching that supports the proposed model returns more relevant result with comparison the existing keyword based searching. And the proposed model will eliminate the limitations of existing search strategies.

5. Conclusion and Future Scope

In keyword based searching, most of the returned results are irrelevant and user unsatisfactory. The existing search engine returns huge websites against user query that causes information overloading [12]. The reason behind is that it even returns those data as a result that slightly match with keywords. Hence, most of the returned results are not that a user actually wants. Therefore, search engines cannot differentiate among different meanings of the words, and ultimately a huge number of websites are returned to the user, some of which may be relevant, nevertheless most of them are irrelevant [2]. To overcome these obstacles, we proposed a model based on RDF strategies and we built two demo searching engine (RDF based and Keyword based) to verify the proposed model.

RDF based semantic web could minimize these limitations of existing keyword based searching because the semantic web maintain a relationship among data. The comparison of RDF based searching and keyword based searching are shown in Fig. 5 and Fig. 6. Here we use two different sets of data and for each set of data the RDF based searching returns more relevant data. Most of the returned results in keyword based searching shown in Fig. 3 are not relevant. In contrast, the returned results in RDF based searching shown in Fig. 4 are relevant to the user query. Hence the efficiency of RDF based semantic web is more than the existing keyword based searching. Moreover, the efficiency of existing keyword based searching can be increased by using the proposed model.

References

- [1] Gopalchari M V, Sammulal P. "A Survey on Semantic Web and Knowledge Processing" *International Journal of Innovative Research in Computer and Communication Engineering*, 2013, Vol. 1, Issue 2.
- [2] Aghaei S, Nematbakhsh M A, Farsani H K. "Evolution of the World Wide Web: From Web 1.0 to Web 4.0" *International Journal of Web & Semantic Technology (IJWesT)*, 2012, Vol.3, No.1.
- [3] Kaur N, Aggarwal H, "Evaluation of Information Retrieval Based Ontology Development Editors for Semantic Web" *I.J. Modern Education and Computer Science*, 2017, 7, 63-73.
- [4] Kabir S, Ripon S, Rahman M. and Rahman T. "Knowledge-Based Data Mining Using Semantic Web," *International Conference on Applied Computing, Computer Science, and Computer Engineering*, 2013.
- [5] Calbimonte J P, Jeung H, Corcho O, Aberer K. "Enabling Query Technologies for the Semantic Sensor Web" *International Journal on Semantic Web and Information Systems*, 2012, 8(1), 43-63.
- [6] Hepp M. "Semantic Web and Semantic Web Services: Father and Son or Indivisible Twins" *IEEE Internet Computing*, 2006, vol. 10, no.2, pp. 85-88.
- [7] Galib S M. et al., "Clustered and Smarter Web mining using Semantic Web," *International Conference on Computer and Information Technology* 2015.
- [8] Rettinger A, et al. "Mining the Semantic Web Statistical learning for next generation knowledge bases" *Data Mining Knowledge Disc* (2012) 24:613–662 DOI 10.1007/s10618-012- 0253-2.9
- [9] Berners-Lee T. "Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor" 1999, Harper San Francisco.
- [10] Yong-gui W, Zhen J. "Research on semantic Web mining" 2010 *International Conference on Computer Design and Applications (ICCD)* , 2010, vol. 1, pp.67-70.
- [11] Segaran, Toby, Evans, et al. "Programming the Semantic Web" O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. p. 84. ISBN 978-0-596-15381-6.
- [12] Chakravarthy A. "Mining the semantic web" In: *Paper Proceedings of the First AKT Doctoral Colloquium* (2005).
- [13] DENNAI A, BENSLIMANE S M, "Semantic Indexing of Web Documents Based on Domain Ontology" *I.J. Information Technology and Computer Science*, 2015, 02, 1-11.

- [14] Singh R, Sharma S A. "Semantic Web and application in ERP" *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 2013, Volume 2, Issue 1.
- [15] Sharma K, Shrivastava G and Kumar V, "Web Mining: Today and Tomorrow," in proceedings of the *IEEE 3rd International Conference on Electronics Computer Technology*, 2011.
- [16] Bhatia C S and Jain S, "Semantic Web Mining: Using Ontology Learning and Grammatical Rule Interface Technique," in *IEEE 3rd International Conference on Electronics Computer Technology*, 2011.
- [17] Meirong T and Xuedong C, "Application of Agent Based Web Mining in E-business," in *IEEE Second International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2010, pp. 192-195.
- [18] DENNAI A, BENSLIMANE S M, "A New Measure of the Calculation of Semantic" *I.J. Information Technology and Computer Science*, 2015, 07, 48-56.

Authors' Profiles



Md. Motiur Rahman: Working as lecturer in the department of Physical and Mathematical Sciences, Chittagong Veterinary and Animal Sciences University. His research interest is in the field of Big Data, Cloud Computing, Data Mining and Machine Learning.



Ferdusee Akter: Served as assistant professor and Head of the department of Physical and Mathematical Sciences, Chittagong Veterinary and Animal Sciences University. She has interest in research in the areas of Computational Mathematics, Modeling and Simulation, Fluid Mechanics, Statistics.

How to cite this paper: Md. Motiur Rahman, Ferdusee Akter, "An Efficient Approach for Web Mining using Semantic Web", *International Journal of Education and Management Engineering(IJEME)*, Vol.8, No.5, pp.31-39, 2018.DOI: 10.5815/ijeme.2018.05.04