www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 4 Issue 9 Sep 2015, Page No. 14458-14465

Design & Analysis of Performance of Semantic based approach for Knowledge Extraction in Web mining Environment

Vipra Dave [1]***, Ms. Pooja Parnami*** [2]***, Mrs. Shweta Sharma*** [3]
[1,2,3]Amity School of Engineering & Technology
Amity University Rajasthan, India
[1]**vipra1325@gmail.com**,[2]**pparnami@jpr.amity.edu**,[3]**ssharma@jpr.amity.edu**

Abstract: The stimulating work and determining for social mining of web data, the user interest for individual and non-individual reference systems is an important task. The sharing of the knowledge between the web users has become important in measuring the tradition of web data and marking the information in many social websites as per the requirement of the user. This paper provides an alternative method as an online application reference system for the performance analysis of semantic based methodology. It extracts information from various web application sources and social websites. The information is extracted for the different users to take a proper decision (uncertainty, qualitative and quantitative analyses and risk) at a critical stage. The results of services with conservative web search engines are inadequate in the situation (anomalies, security attacks) and simultaneously reducing the service discovery search space with the attention of mathematical/statistical method to brief the text. The data repository is consisting common data which is reached by the process of obtaining data from different web sources. The semantic data is collected from the ontology based data source and draws it to the user with help of query processor constituent. An online reference system is used to make references for a decision making process of the user. The implementation process of this proposed work is helping the experimental process of an online application reference system for the career guidance of a learner.

Keywords: Knowledge Management; Knowledge Extraction; Web Mining, Knowledge Management.

Introduction: Today the various websites use many social networks (such as Facebook, Orkut, LinkedIn and What's app) to share the huge amount of knowledge between several users from which they can take significant decisions in different fields. For example in field of education, we are choosing the popular college which conduct higher studies, finding the leading and special institution to pursue proposed work. It can also recognize the latest requirement of concern for the learner's recruitment. The wide scope of growth in social web data makes and prepares various new methods, algorithms, events and approaches to measure the large capacity of web data. It is done to recognize the hidden pattern and to quote the information among

numerous web users. In present condition, the existing data is plentiful and requirements for the users are not totally content due to various natures of the web data. The importance of database from a Semantic Web perspective is evident from the multiple advantages and in different formats need a structure to make them accessible. After all, the problem of recording a database into Semantic Web did not appear as a simple workout of change from one sign ideal to different sign ideal. It is significant to recognize many inspirations, problems and associating relations between relational databases and software skills, in order to succeed a clear departure of goals and challenges like data integrity, referential integrity, or keyed data access. The semantic web technologies and data from multiple heterogeneous sources play a dynamic role in information removal and discovery of knowledge from the web papers. It proposes a typical storage for web pages in an organized and structured manner in RDF (Reduced Description Framework format). The methods are related to data extraction and the development of combine data for the field together determines new knowledge. Main goal of semantic based method is to improve the quality and standard of interaction among the web users.

Proposed Approach: The knowledge-based decision support system (KBDSS) is developing for immediate scheduling in FMS (effectively subjective by the management tool concept). It is developing to deliver an important operational regulator tool for an extensive choice of machining cells (where the high flexibility is required, assistances of more efficient cell operation are required, flow control is better of tools). The paper uses three knowledge-based models to ease the decision making process:

a. An expert production scheduling system,
b. A knowledge-based tool management decision support systems
c. A tool management fault diagnosis system.

The implementation of expert system (ES) used as a commercial expert system shell, Knowledge Engineering System (KES) Production System (PS). The recent advancements in Knowledge based decision support system(KBDSS) from different perspectives, especially in terms of the improvement of knowledge management function through multiple knowledge levels, knowledge reuse, knowledge mobilization, critical knowledge, knowledge chain management and knowledge integration are requisite.

DOI: 10.18535/ijecs/v4i9.68

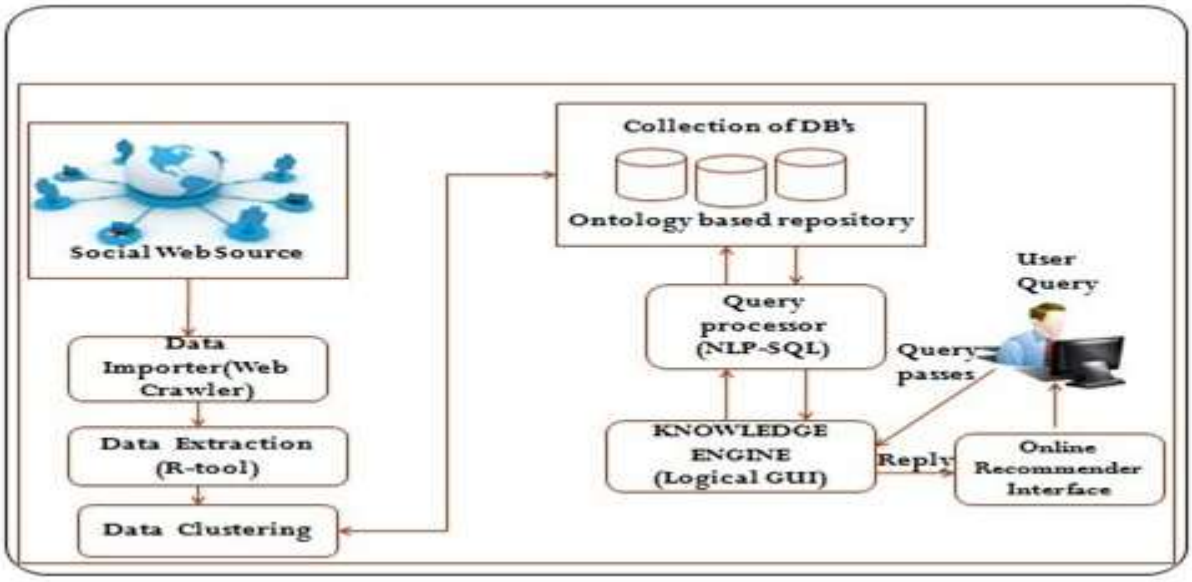

Figure 1: Architecture of KBDSS (Knowledge based decision support system)

KBDSS consists of three components:

A. Database repository
B. knowledge engine and query processor
C. Online recommendation system

A. Database Repository: A database repository is represents a logical group of data. Sometimes it also represents data in physical grouping form but related to discrete databases. The data objects needed to exist on several databases, so a repository is required to get together and is a collection of different kind of data. It may be just the combination of data itself into some available place of storage or it may also suggest some aptitude to selectively extract data. Repository databases are determined by the repository engine, which achieves all transactions and regulates storage structure. The database saves the space. Meta Data Services can sometimes remove redundant data definitions. For example, it may store a single copy of a property value, even if that property value defines many object versions. Sometimes, a single copy of a relationship can be stored in Meta Data Services, even if many unrelated object versions have maintained the relationship. The Repository databases should not be changed directly. An expert database programmer or administrator avoids changes because you can introduce variations that the repository engine cannot manage.

DOI: 10.18535/ijecs/v4i9.68

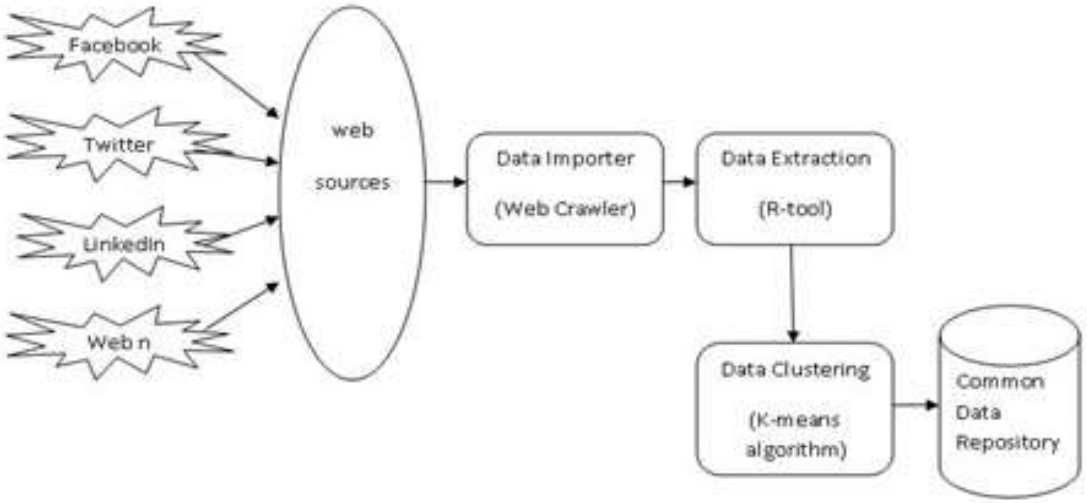

Figure 2: Block Diagram of Data repository Management

B. Knowledge Engine and Query Processor Component: The Knowledge engine represents a logical user interface and this is set to receive the query from the user and deliver it to the query processor (where the natural language query is converted into SQL query). In this process of the query conversion, the natural language queries are reprocessed by many stages, such as morphological analysis, semantic analysis and mapping.

C. Online Recommendation System: The content used for semantic analysis is collected from knowledge engine. The semantic data is interlinked with the fundamental structure of online recommendation system which basically gives the various options to users. This system gathers the information from various web resources and transforms it into common database repository. The user uses various queries and provides appropriate online recommendation through the interface of online recommender.

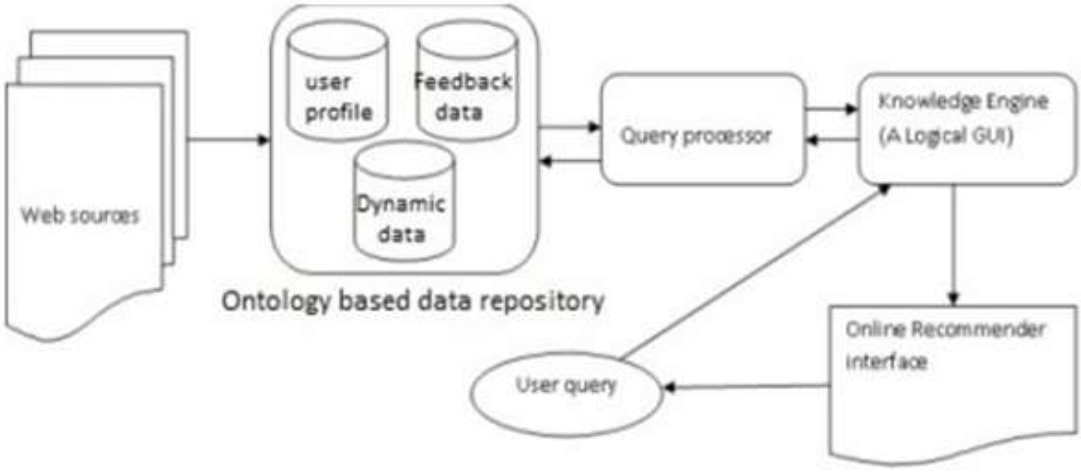

Figure 3: General Architectural Diagram of Online Recommendation System

Data Collection: We have collected the web data from educational sites. Meta data indexing is related with Meta data which is based on the indexed incidence. The web content which is surfed by the most frequent users is compared with the data which is collected from web. In this paper we are using MySQL Database Server. There are certain tables which are going to be used. All the data coming and going out is through this Schema. It keeps the complete records of users, their mails, their discussion and all the study stuff. The web crawling process is used for the huge collection of web data and collection of the links. The web data is collected from the web sources after that with the help of R-tool we can sort and organize for further processing. The collected and processed web data is organized as groups of homogenous and well defined data. After the sequence procedure, common data repository acts as a data pool from where the huge amount of data is processed and kept stable for the fast retrieval of required information. In web document collection process the data is prearranged according to content match. Here we can use the data extraction technique with step by step process based on the correspondence in web data. We are also creating the web data on the basis of many features like user profile, user blogs, user ratings and user logical data, collected from various sources of the web data. Finally the data systematically integrated in data integration phase and stored in shared data repository, which is also use as a data pool from which the query raised by the user is answered. It also provides the details for analysis & design of tested results and performs comparative analysis of all the semantic based approach.

Comparative Analysis of Performance of Generated SQL Results: The results are obtained by submission of an online application in recommendation analysis system to measure the recall and precision.

Recall = number of relevant links retrieved / number of relevant links requested

Recall: Measure the ability of a system to present all attributes.

Precision = number of relevant link retrieved / total number of links retrieved

Precision: Measure the ability of a system to present only relevant attributes.

The measurement of precision and recall are used to measure the quality of an unordered set of retrieved fields. We can measure the list of the rank, plot the precision with the recall after retrieving the semantic fields. We then compute the average presentation over with selected domain with different number of related documents with individual topic precision values. These are then interpolated to make collection of standard recall levels (0 to 1 in increments of .1). The specific rules are used to interpolate the precision at standard recall level. ‘i’ is used to show maximum precision obtained for topic (for any exact recall greater than or equal to ‘i’. The SQL query is used to generate the relevant threshold for measuring the range of recall and precision of each word pair generated in SQL. The high cut off range semantic values extract the relevant

DOI: 10.18535/ijecs/v4i9.68

data from different web sources. This analysis gives SQL results to find the extracted data from different web sources traversed, to make statement wise user query which is extracted in various web sources like Facebook, Twitter and LinkedIn (used as primary sources of data).

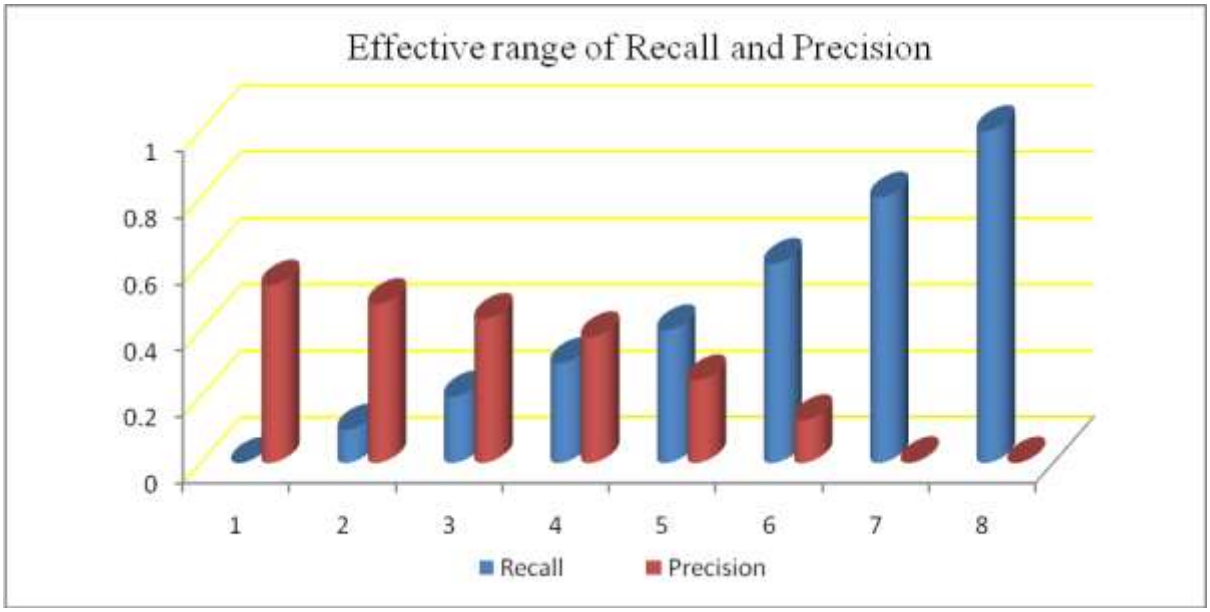

Figure 4: Effective range of Recall and Precision

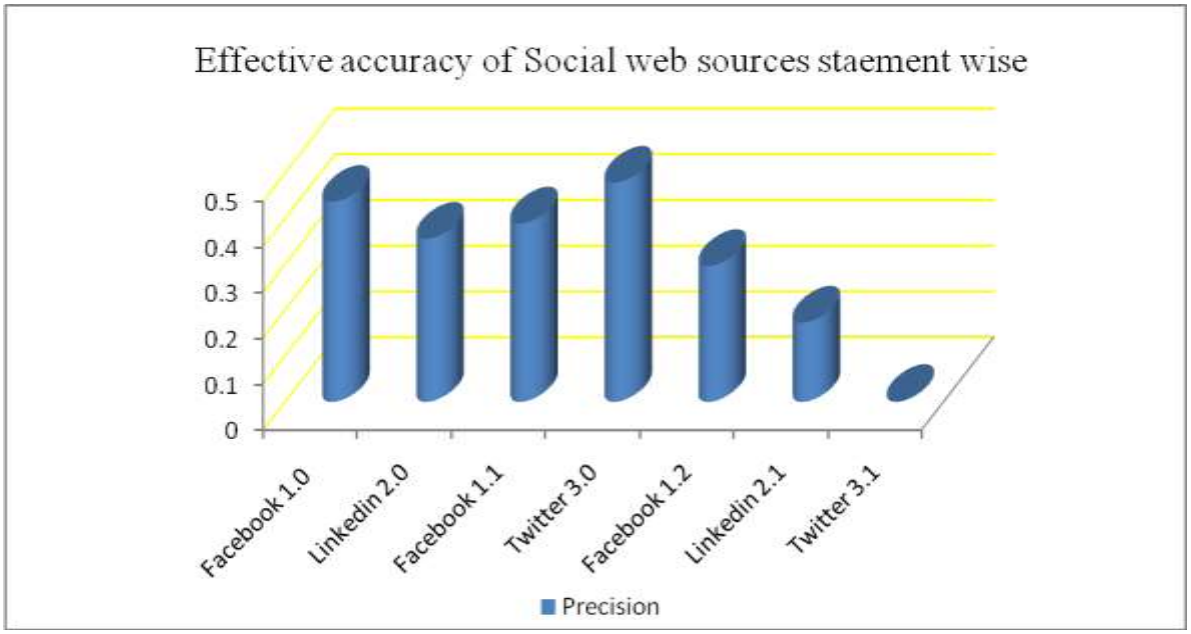

Figure 5: Effective accuracy of Social web sources statement wise

Results & Discussion:

1. Firstly, the main objective of this research work is to providing knowledge among different users in the web to take correct decision at a crucial stage. The proposed structure implements the online recommender application for the learners' career guidance that follows their advancement in expanses and distance education environment.
2. Secondly, this research work the reliable knowledge is conveyed through online social network for interaction among the various users on web.
3. An online application phase is held by examining the data from different web userswhich is used for analysis and predicting the relevancy attained per statement.
4. And proposed online application of recommendation system is used to measure the different ranges of accuracy also extract the semantic knowledge from the various sources of web.

Conclusion: In this research paper, we propose an application of an online recommendation system for processing of decision support system in the ground of data is analyzed semantically. We also measure the list of the rank plot the precision with the recall after retrieved the semantic fields. We computed the average presentation over with selected domain with different number of related documents with individual topic precision values are interpolated to collection of standard recall levels (0 to 1 in increments of .1). The SQL query used to generate the relevant threshold for measuring the range of recall and precision of each word pair in generated SQL. The high cut off range semantic values are extracted as the relevant data from different web sources. Our analysis give SQL results to find and extract data from different web sources, it also give traversed statement wise user query example like Facebook, Twitter, Linkedin (used as primary sources of data).

Future Work:

1. This experimental work can be further extended for other web users in various places such as e-commerce sites, business sites etc.
2. This experimental work can be improved with various precious output generated classification techniques like Naïve Bayesian Classifiers and SVM (Support Vector Machine).
3. The various types of algorithmic approaches are used by knowledge engine for retrieving the large pool of stored semantic data for various types of web users and web sites.
4. In online Media system this proposed work is combined with the computational techniques for extraction of knowledge which is also represent the approaches of knowledge reasoning.

References:

1. G. Vadivelou, E. Ilavarasan,"*Performance evaluation of semantic approaches for automatic clustering of similar web services*", World Congress on Computing and Communication Technologies, 2014.
2. Zhengxiang Pan, Jeff Heflin, DLDB: Extending Relational Databases to Support Semantic Web Queries, the Air Force Research Laboratory, Contract Number F30602-00-C-0188.
3. Yu-Chun Kuo, Andrew E. Walker, Brian R. Bellan, and Kerstin E. E. Schroder, "A Predictive Study of Student Satisfaction in Online Education Programs" *The International Review of Research in Open and Distance Learning,* 2014.
4. G. Vadivelou, E. Ilavarasan,"*Performance evaluation of semantic approaches for automatic clustering of similar web services" ,*World Congress on Computing and Communication Technologies, 2014.
5. ShaojianZhuo, Xing Wu, XiangfengLuo, Chinese Text Sentiment Analysis Based on Fuzzy Semantic Model, 2014 IEEE 13th Int'l Conf. on Cognitive Informatics &Cognitive Computing (ICCI*CC'14) S. Patel, Y. Wang, W. Kinsner, D. Patel, G. Fariello, & L.A. Zadeh (Eds.)
6. S. M. Patil, D. M. Jadhav, Semantic Search using Ontology and RDBMS for Cricket, *International Journal of Computer Applications (0975 – 8887) Volume 46– No.14, May 2012.*
7. Milind R. Naphade, SankarBasu, John R. Smith, Ching-Yung Lin and Belle Tseng, "A Statistical Modeling Approach to Content Based Video Retrieval", Pervasive Media Management Group IBM Thomas J. Watson Research Center Hawthorne, NY 10532, 2002.
8. ShaojianZhuo, Xing Wu*, XiangfengLuo, "Chinese Text Sentiment Analysis Based on Fuzzy Semantic Model", 2014 IEEE 13th Int'l Conf. on Cognitive Informatics & Cognitive Computing (ICCI*CC'14).
9. Mark Zaslavskiy, Dmitry Mouromtsev, "Geocontext extraction methods analysis for determining the new approach to automatic semantic places recognition", PROCEEDING OF THE 16TH CONFERENCE OF FRUCT ASSOCIATION.
10. RanjithaKashyap* and AniNahapetian, "Tweet Analysis for User Health Monitoring", International Conference on Wireless Mobile Communication and Healthcare - "Transforming healthcare through innovations in mobile and wireless technologies" (MOBIHEALTH), 2014.
11. Mrityunjay Singh∗, S K Jain†, and V K Panchal, "Modeling Semantic Heterogeneity in Dataspace: A Machine Learning Approach", 13th IEEE International Conference on Information Technology, 2014.
12. Yunni Xia, XinLuo, Jia Li, and Qingsheng Zhu, "A Petri-Net-Based Approach to Reliability Determination of Ontology-Based Service Compositions", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 43, NO. 5, SEPTEMBER 2013.
13. Xiaocao Hu, ZhiyongFeng, Shizhan Chen, "Analyzing Distribution of Implicit Semantic Information in Web Services", 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops.
14. Michael P. McGuire, Ziying Tang, "Mining Semantic Time Period Similarity in Spatio-Temporal Climate D", IEEE 13th International Conference on Data Mining Workshops, 2013.