ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 3; Issue 3; May-June 2015; Page No. 106-110

Performance Analysis of Semantic based approach for large database system to summarizing the data

Vipra Dave [1], Mrs. Shweta Sharma [2], Ms. Pooja Parnami [3]

[1,2,3] Amity School of Engineering & Technology

Amity University Rajasthan, India

[1]vipra1325@gmail.com,[2] ssharma@jpr.amity.edu,[3] pparnami@jpr.amity.edu

Abstract:

In this paper we provide an unconventional approach for the analysis of performance of semantic based methodology for extracting knowledge from web sources for the end users to take a right decision at a crucial juncture. Today the more business are distributing and their services are available within organization and new and challenge search problem locating the desired services also computing relative importance of textual in large dataset in order to summarizing the text. The searching for the services with conventional web search engines is insufficient in the context and at the same time reducing the search space for service discovery with the consideration of mathematical/statistical method to summarizing the text. The proposed framework implements an online recommender application for the learners online who pursue their graduation in an open and distance learning environment. This framework possesses three phases: data repository, knowledge engine, and online recommendation system. The data repository possesses common data which is attained by the process of acquiring data from various web sources. The knowledge engine collects the semantic data from the ontology based data repository and maps it to the user through the query processor component. Establishment of an online recommendation system is used to make recommendations to the user for a decision making process.It is also optimize the searching quality of the knowledge and services in the future applying for the collection for the services to test the quality of the services.

Keywords: Knowledge management; knowledge extraction; web mining, Knowledge management; knowledge extraction; web mining,

INTRODUCTION

Today the large amount of web data stored in web servers is increasing rapidly. “Online social networks” have exploded in acceptance and now rival the traditional web in terms of usage. Impact of social web in educational domain reflects the magnitude of data accessed on the web for learning online. Particularly in the field of open and distance learning the knowledge transferred online is huge. The social networking sites Facebook, Orkut, LinkedIn and Whatsapp are examples of widely used popular networks to share the enormous amount of knowledge among the various users from which the users take the crucial decision in various domains, for example in educational domain, choosing the best institute to pursue higher studies, finding the premier and special institute to pursue research work, identifying the current requirement of corporations for the recruitment of learners, deciding on best online material available for the different class of academics and getting career guidance information. The web source acts as a resource pool which provides an infinite set of solutions to the users who are looking for useful information on the web. But the fact that information found in web is not relevant and reliable to a great extent is due to large collection of content.Therefore a framework is needed which would consist of structured content from which the user could pick the most relevant and reliable information for usage. Some of the information like historical data, current data, feedback data and dynamic data were analyzed and stored to reduce the complexity of web content for decision making process. The extensive scope of growth in social web data creates and initializes various new techniques, algorithms, procedures and methods to assess the large volume of web data, to identify the hidden pattern and to extract the knowledge among the various web users. In present situation, the existence of data is plentiful in the web and requirements for the users are not completely satisfied due to heterogeneous nature of the web data.The work carried out in this research study is emphasized through the following components: managing data repository, knowledge engine query

Corresponding author: Vipra Dave

processor component, and establishment of online recommendation system. The main objective is to improve quality and performance of semantic based approach for standard of interaction among the web community.

Related Work: A computational method for actionable knowledge extraction from online media is implemented. The approach used is based on mutual bootstrapping and combined with knowledge reasoning. Comparing with the related work, the approach used can acquire more types of action knowledge, and needs much less human labor. However knowledge extraction through conventional method is time consuming which is addressed in this study (AnshengGe, et al., 2013).

Methodology:

The knowledge extraction from the public web source for online recommendation system in open and distance learning environment is carried out with the help of a step by step procedure which is incorporated in this section. The basic architectural layout of knowledge based decision support system shown in Figure 1 is proposed in this research work which provides the generic framework from which the knowledge extraction is initiated from various social web sources.

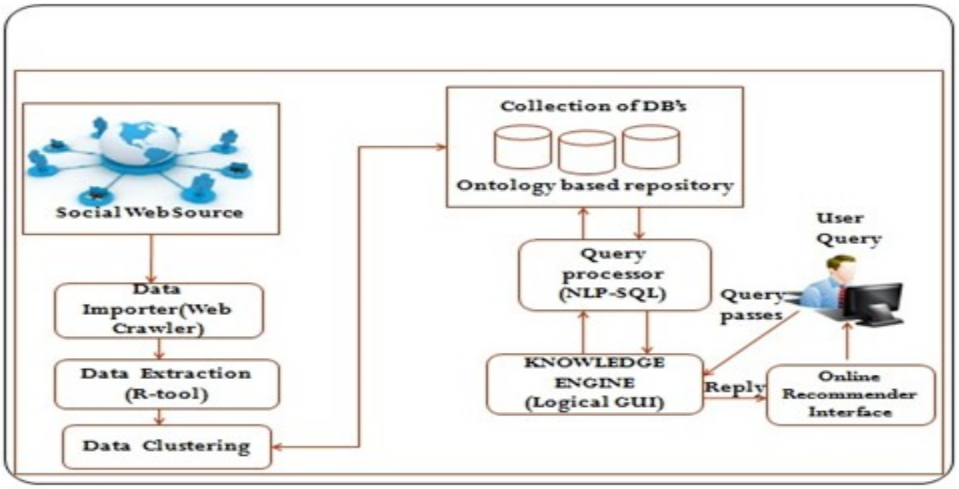

Figure 1: Architecture of KBDSS (Knowledge based decision support system)

Social web source from different collections of web data which are in various web sources such as social networking sites, business oriented sites, e-commerce sites, educational sites, e-learning sites, academic institutional sites, informational sites. The vast collection of web data and its links are collected by using a web crawler. The collected web data is organized and sorted with the help of R-tool for further processing. The organized web data is grouped based on the similarity and stored as well as defined clusters using data clustering technique. After the sequence procedure, a 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.

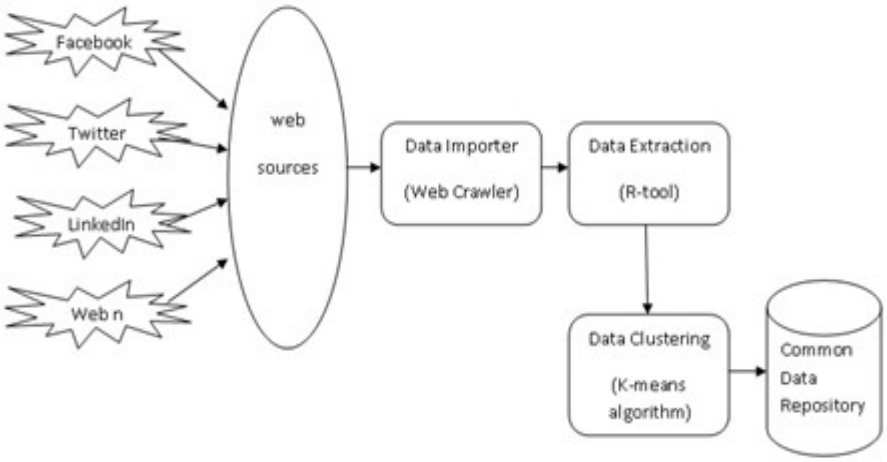

Figure 2: Block Diagram of Data repository Management

In Meta tag indexing method the collected web data from educational sites is compared with indexed based on the Meta data occurrence, that is, the web content which is surfed by the user most frequently is compared with the collected web content.

Table 1: Web Attribute Data fields

Attribute name	Data type
Student Id	Number
Student name	Text
Gender	Text
About me	Text
Place of education	Text(Discrete)
Overall GPA	Number
Softskill rating	Text (Discrete)
Analytical rating	Text (Discrete)
Overall Performance rating	Text (Discrete)

The complete enrichment of data is done using an unsupervised learning technique called 'data clustering'.In data clustering a web document is collected and it is organized according to content similarities. Here K-means clustering technique is used where the extracted data is clustered step by step based on the similarity in web data. The K-means partition algorithm works with the given web data by clustering the web data based on various attributes such as user profile, user blogs, user rating and user logical data from the different web data sources. In this research work educational data of post graduate students is addressed to recommend the career guidance by analyzing the performance of each individual which is based on various criteria. the end the clustered data is combined semantically in data integration phase to be stored as common data repository which acts as a data pool from which the query raised by the user is answered. K-means clustering algorithm devised as follows for data clustering.

K-Means Partitioning Algorithm for Data Clustering

Algorithm prerequisite k-means K,D (k-Abstract of all the attribute, D-most matching data who possesses a similar attribute):

Step 1: Choose *k* data points (web data attribute) to be the initial centroids, cluster centers (user profile attribute) // compares all the attributes by visiting the whole data

Step 2: Assign each data point (X) to the closest centroid (D) (next attribute (performance rating) is targeted as next cluster)// Finding the similarity in attributes which most likely matches the corresponding data for clustering

Step 3: Re-compute the centroids using the current cluster memberships (until similar type of attribute encountered).

Step 4: If a convergence criterion is not met, go to 2.

Table 2: Cluster Form using K-means

Cluster id	Cluster category	Cluster constraint	Description
1	High performer	Overall gpa>8.5 Softskill rating="A"grade Analytical skill rating="Good"	Student who satisfies the required constraints with deserved score for high performer.
2	Intermediate performer	Overall gpa>6.5 Softskill rating="C"grade Analyticalskill rating="Better"	Student who satisfies the required constraints with deserved score for high performer.
3	Low performer	Overall gpa<6.0 Softskill rating="E"grade Analyticalskill rating="Worst"	Student who satisfies the required constraints with deserved score for high performer.

Result

In this paper, an online recommendation system for a decision making process was designed and implemented. The application implemented in this study is centered on the need for decision support system in the field of open and distance learning. The current research work indicates that there are very few applications which emphasize the significance of learner's ability in their academic performance.Here student academic performance, placement eligibility criteria, and student historical data are analyzed semantically to take decision on placement probability.This study will also identify the students who are in need of special attention to improve their performance.

Conclusion

This study will also identify the students who are in need of special attention to improve their performance. With the help of this study the user gets the appropriate recommendation system in educational domain. As this research work targets to improve performance of an educational institute and bring a better result in terms of placement probability, the extension of this research can be applicable for different domains.

References:

[1] Anil M. Bhadgale, Sanhita R. Gavas, Meghana M. Patil, &Pinki R. Goyal.(2013). Natural language to SQL conversion system. *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), Vol. 3*, Issue 2, pp. 161-166.

[2] AnshengGe, Wenji Mao, Daniel Zeng, & Lei Wang (2013).Action knowledge extraction from web text intelligence and security informatics. (ISI) *IEEE International Conference*, pp. 368-370.

[3] Danyllo, W. A., Alisson V. B., Alexandre N. D., Moacir L.M. J., Jansepetrus B. P. (2013). Identifying relevant users and groups in the context of credit analysis based on data from Twitter. *Cloud and Green Computing (CGC), Third International Conference,* pp. 587-592.

[4] EdinOsmanbegović, &MirzaSuljić (2012). Data mining approach for predicting student performance, economic review. *Journal of Economics and Business,volume 10*, issue1.

[5] Hariton A. Efstathiades (2013). *Extract knowledge from social networks*. Thesis Web Information Systems, Department of Software Technology Faculty EEMCS, Delft University of Technology.

[6] Jong Y. Choi, Joshua Rosen, SiddharthMaini, Marlon E. Pierce, and Geoffrey C. Fox (2008). Collective collaborative tagging system. *Grid Computing Environments Workshop GCE'08*, pp.1-7.

[7] Kevin P. Brady, Lori B. Holcomb, and Bethany V. Smith (2010). The use of alternative social networking sites in higher educational settings: A case study of the e-learning benefits of Ning in education. *Journal of Interactive Online Learning,volume no 9,* no.2.

[8] Maina Elizaphan Muuro, Waiganjo Peter Wagacha, Robert Oboko, and John Kihoro (2014).Students' perceived challenges in an online collaborative

learning environment: a case of higher learning institutions in Nairobi, Kenya. *The International Review of research in open and distance learning, volume no 15*, no.6.

[9] Patricia Fidalgo and Joan Thormann (2013).A Social Network Analysis comparison of an experienced and a novice instructor in online teaching. *European Journal of Open, Distance and E-Learning,volume no 12,* no.39.

[10] RabiaBatool, AsadMasood Khattak, Jahanzeb Maqbool and Sungyoung Lee (2013).Precise Tweet classification and sentiment analysis. *Computer and Information Science (ICIS), IEEE/ACIS 12th International Conference*, pp. 461-466.

[11] S.K AlthafHussain, & Y.R Ramesh Kumar (2012). Intellectual knowledge extraction from online social data. *IEEE/OSAIIAPR International Conference on Informatics, Electronics & Vision*, pp. 205-210.

[12] Saravjeet Kaur, Rashmeet Singh Bali (2012). SQL generation and execution from natural language processing. *International Journal of Computing & Business Research*, 2229-6166.

[13] Shreenath Acharya, &Madhu, N. (2012).Discovery of students' academic patterns using data mining techniques. *IJCSE volume 4.*

[14] Verónica Marín Díaz, Ana Isabel Vazquez Martinez, and Karen Josephine Mc Mullin (2014). First steps towards a university social network on personal learning environments. *The International Review of Research in Open a nd Distance Learning, Volume No 15*(3).

[15] VoThi Ngoc Chau & Nguyen Hua Phung (2012). *A knowledge-driven educational decision support system*. Ieee Publications.

[16] Yu-Chun Kuo, Andrew E. Walker, Brian R. Bellan, and Kerstin E. E. Schroder.(2014). A Predictive Study of Student Satisfaction in Online Education Programs. *The International Review of Research in Open and Distance Learning, 15*(3).