A Generic Framework for Extraction of Knowledge from Social Web Sources (Social Networking Websites) for an Online Recommendation System

Javubar Sathick and **Jaya Venkat**
B.S. Abdur Rahman University, India

Abstract

Mining social web data is a challenging task and finding user interest for personalized and non-personalized recommendation systems is another important task. Knowledge sharing among web users has become crucial in determining usage of web data and personalizing content in various social websites as per the user's wish. This paper aims to design a framework for extracting knowledge from web sources for the end users to take a right decision at a crucial juncture. The web data is collected from various web sources and structured appropriately and stored as an ontology based data repository. 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. This research work is implemented with the help of an experimental case study which deals with an online recommendation system for the career guidance of a learner. The online recommendation application is implemented with the help of R-tool, NLP parser and clustering algorithm.This research study will help users to attain semantic knowledge from heterogeneous web sources and to make decisions.

Keywords: Knowledge management; knowledge extraction; web mining, decision making system; R tool, natural language processing; query mapping; k-means clustering; knowledge engine; social web source; online recommender interface

Introduction

Nowadays the amount of web data stored in web servers is increasing rapidly. "Online social networks" have exploded in popularity 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. Especially 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.

In this era, 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 analysed and stored to reduce the complexity of web content for decision making process.

 The growth of technology in educational domain has evolved into a new level where all the information is communicated to the end user through a web such as e-learning environment, virtual classes, simulation tools, common web forum, etc. These infrastructural enhancements provide a valuable reason for upgrading in sharing information effectively to all users. Therefore a framework is essential to address the feature of the web environment in the field of education specifically for open and distance education.

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 current scenario, the existence of data is abundant in the web and requirements for the users are not completely fulfilled due to heterogeneous nature of the web data. So the data which is available at multiple web sources and in different formats need a structure to make them accessible. This research work aims to provide a reliable framework for using web data efficiently. Since, the web data

collected from the different web sources possess different criteria for assessing them efficiently a standard framework is needed as immediate solution to assist the end user for an online recommendation.

This framework aims to provide a personalized and non personalized recommendation system for the web user with the help of common data repository. The proposed framework shows the typical interaction between the user and the web through the knowledge engine which acts as a logical interface for the user to fetch useful information from the web.

The work carried out in this research study is emphasized through the following components: managing data repository, knowledge engine query processor component, and establishment of online recommendation system. In the present investigation, an experimental methodology was adopted to generate a user interface where a user enters his web based queries which targets the performance analysis of a student for his career guidance and the solution obtained in this study provides a suitable recommendation system for the student in terms of providing him suitable career guidance. The main goal is to improve quality and standard of interaction among the web community.

Related Works

This section emphasizes some of the existing studies in this domain. Online social networks serve a number of purposes, but three primary roles stand out as common across all web sites. Firstly, online social networks are used to maintain and strengthen existing social collaboration or make new social connections. The targeted websites allow users to "articulate and make visible their social networks", thereby "communicating with people who are already a part of their extended social network". Secondly, online social networks are used by each member to upload his/her own content. It is observed that the content shared often varies from site to site, and sometimes it is only the user profile. Third, online social networks are used to find new, interesting content by filtering, recommending, and organizing the content uploaded by users. Jong Y. Choi et al. (2008) discussed a collective collaborative tagging (CCT) service architecture in which both service providers and individual users can merge folksonomy data (in the form of keyword tags) stored in different sources to build a larger, unified repository. The unified repository acts as a data pool from which only a similar kind of query is addressed, whereas this research framework is applicable for different collections of queries.

Althaf Hussain and Ramesh Kumar (2012) present a systematical data mining architecture to mine intellectual knowledge from social data. Here the author used social networking site Facebook as primary data source. He collected different attributes such as 'about me', 'comments', 'wall post' and 'age' from Facebook as raw data and used advanced data mining approach with the help of Weka tool to excavate intellectual knowledge. He also analyzed the mined knowledge with

comparison for possible usages like human behavior prediction, pattern recognition, job responsibility distribution, decision making and product promoting which is enhanced in this research 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 (Ansheng Ge, et al., 2013).

Rabia Batool et al. analysed the tweet data and information from tweets.The system is tested on a collection of 40,000 tweets for finding semantic content in the tweet. The knowledge enhancer and synonym binder module is applied on the extracted information which increases information gain in a range of 0.1% to 55 which can be expanded in this research work, since the current research study deals with web data.

Edin Osmanbegović and Mirza Suljić (2012) proposed the factorization method which describes the recommended system technique for predicting student performance. The matrix factorization method is used for predicting student performance which is handy information to advance this research study.

Vo Thi Ngoc Chau and Nguyen Hua Phung (2012) used the technique to evaluate the faculty performance by signed and unsigned student feedback using linear regression technique. It shows the significance of the student feedback in making certain decisions in learning environment which is addressed in this research work. Apriori algorithm is implemented where the student and placement based data is mined. The outcome of the study provides suitable analysis of student performance and placement possibilities. It is evident that the various factors are involved in determining the overall performance of a student. The factors which affect a student are encountered and proper solution is provided in this research study (Shreenath Acharya and Madhu, 2012).

Danyllo et al. discussed the method of collected data from the social network Twitter, and compared them with data from a financial institution in order to model the network and analyze the similarities. The result revealed that most of the users have more credit restrictions than neighbours, and users with no restrictions normally have neighborhoods with no credit restriction as well. Here social network analysis was done with reliable metrics on the Twitter database. However, the knowledge extraction in social network differs in large extent which is highlighted in this research work.

Hariton A. Efstathiades (2013) describes knowledge extraction from social network. The data was retrieved from online social network mainly through a manual programming procedure with the use of application programming interface. However, the majority of social science researchers

usually have low programming experience. Thus the procedure from data retrieval to knowledge extraction is not a trivial task for them.

In the online recommendation system, a user interface is very much required to get the query from the user. The query posted by the user is of normal conventional statement which must be converted into official structured query, therefore Anil M. Bhadgale et al. (2013) describe a system that will convert English statement given by user to all possible intermediate queries so that user can select appropriate intermediate query and the system will generate SQL query from intermediate one. Finally the system will fire SQL query on database and provide output to user.

An interface module converts user query given in natural language into a corresponding SQL command. Asking question to database in natural language like English is a very convenient and easy method to access data from database system especially for normal users who do not understand complicated database query languages. The complete semantic conversion is not attained due to complex sentences as a query statement (Saravjeet Kaur and Rashmeet Singh, 2012), which is addressed in this research work.

Kuo et al. (2014) intended to investigate the degree to which interaction and other predictors contribute to student satisfaction in online learning settings. The effects of student background variables on predictors were explored. The results showed that learner-instructor interaction, learner-content interaction, and Internet self-efficacy were good predictors of student satisfaction while interactions among students and self-regulated learning did not contribute to student satisfaction. Therefore a better interaction has to be incorporated for positive e-learning environment.

DIPRO 2.0 is an educational social network for university professors to develop their training in the area of personal learning environments through collaborative learning and production of knowledge. Here web 2.0 social network tool is used efficiently to extract the knowledge from various web sources. Members from various educational institutions interact through social network tool. The ideas and thoughts were shared among the various users. The level of knowledge transmitted through social network tool is in high-end since the collaborative environment is provided with the help of web 2.0. The knowledge sharing phenomena is also discussed in this framework (Verónica Marín-Díaz et al., 2014).

A study on collaborative learning system speculates the importance of online distance learning system where the students are provided with social interaction to share knowledge among the various group of people. Social interaction session among the various users is initiated by examining the challenges involved in social network environment which is encountered in this article (Maina Elizaphan Muuro et al., 2014).

Kevin P. Brady et al. (2010) emphasize the significance of social network in distance learning environment. Tutor has choice of picking the right interface to transfer the knowledge through social network such as Facebook, Myspace, and so on. Different forums were used to collect the

data of a student to analyse a suitable mode of distance learning which is discussed in this article in terms of analysing the career guidance of a student who completes their education in open and distance learning environment.

Patricia Fidalgo and Joan Thormann (2013) investigate the enhancement in learning through distance education online by analysing a traditional online teaching and social network forum. The level of interaction between the tutor and learner were measured and analysed in terms of frequency in which the interaction occurs. Knowledge sharing among various web users emphasizes the need of proposed framework which is discussed in this article.

Thus, the existing work and strategies were discussed in detail in this section which provides sufficient reason to expand and enable the current framework. On the whole all the existing work provides complicated and different types of framework which are not suitable for many cases which holds up the reason for the extension. The methodology and experimental case study is discussed in the next section.

Methodology

The knowledge extraction from the social 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 1.0 is proposed in this research work which provides the generic framework from which the knowledge extraction is initiated from various social web sources.
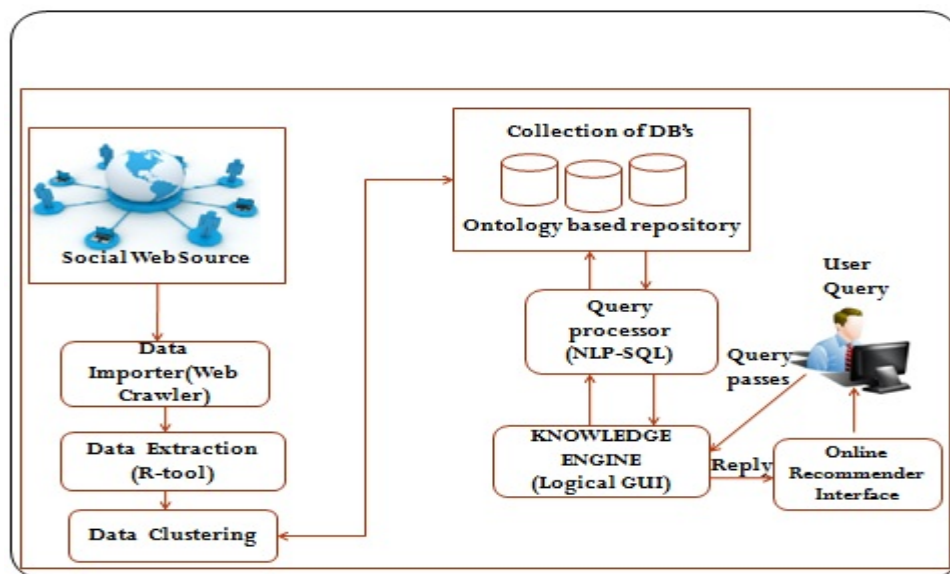

Figure 1. Architecture diagram for KBDSS.

KBDSS (Knowledge based decision support system) consists of three phases:

- managing data repository,

- knowledge engine and query processor,

- online recommendation system.

Managing Data Repository

Social web source is the origin of 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, general blog forum, internet forum, authors forum and few more inevitable 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. Managing the data repository consists of three phases: data importer, data extraction and data clustering.
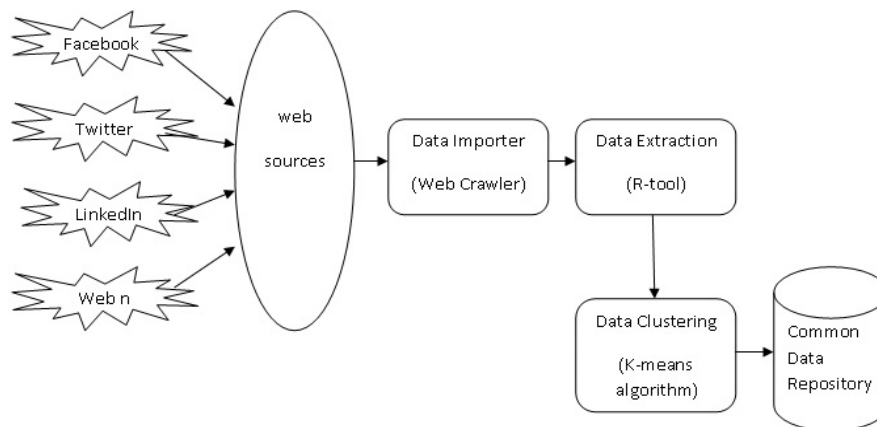
Figure 2. Block diagram for data repository management.

The typical architectural layout of managing data repository is shown in Figure 2.0. The data is collected from various web sources such as Facebook, Twitter, LinkedIn. In this research work the data is collected from the social web community BSAU professional group which is developed and maintained to process the student professional data for determining career guidance and

253

placement opportunities. BSAU professional group is an exclusive and private forum in which the students enroll themselves and post their professional data where various people communicate to share their valuable thoughts about career and placement opportunities.Therefore the data is collected from the BSAU forum back-end database, which is illustrated in Figure 3.0, and imported with the help of a web-crawler in the data importer phase.

Figure 3. BSAU forum back-end data .

The collected web data is in unstructured form which is converted as structured form. The web content is indexed with the help of meta tag indexing method. The attributes targeted for indexing are web document id, frequency of term in web document and position of term in document. The attribute name listed in Table 1 occurs as randomized group of data from BSAU forum from which the id of the web document, frequency of term and position of term is measured. 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 Data Attribute List

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)

Meta tag indexing technique structures the web data to a large extent, the data is extracted by using R tool in the data extraction phase, the complete conversion from unstructured form to structured form is done here. With help of R tool the collected data from the previous phase is extracted and organized for further analysis of data to be stored in the common repository. Figure 4 illustrates the data extraction phase of sample student data; the data extracted in this phase defines the organized attribute of the collected data from the previous phase.
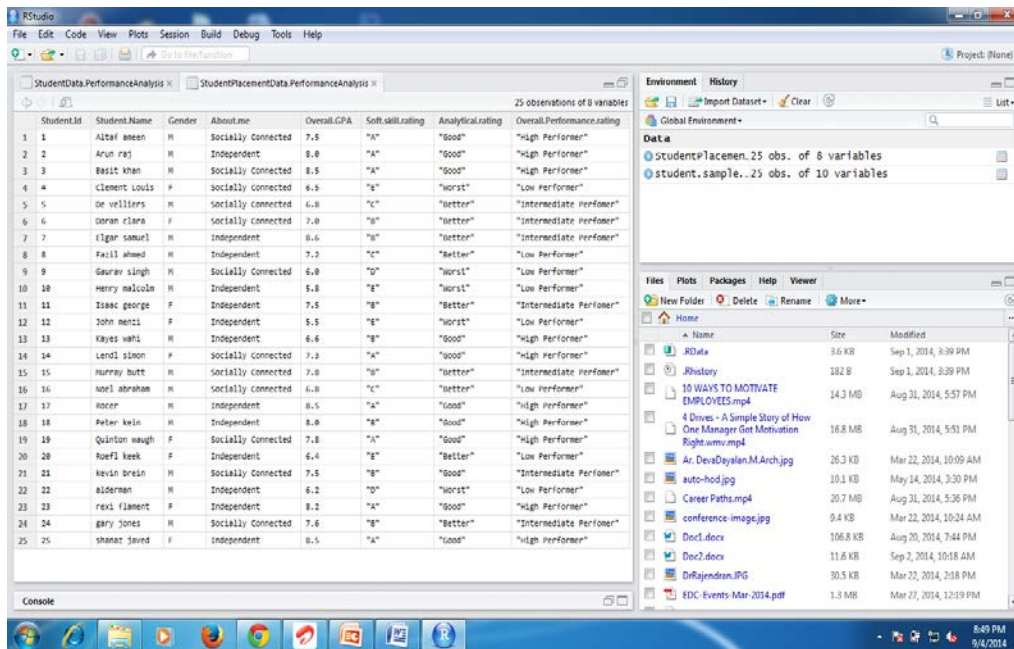
www.manaraa.com

Figure 4. Screenshot for data extraction.

The data is collected and organized at this juncture, but 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. 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 analysing the performance of each individual which is based on various criteria. The collected web data is clustered based on the performance rating of a student such as high performer, intermediate performer, and low performer. In the later part of the work the placement and career guidance criteria for a student is addressed and mapped to the clustered data which is discovered from the educational data in this phase. Each cluster is analyzed individually and computed semantically for further processing in the next phase. At 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

Clusters Formed Using K-Means Algorithm

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.

Knowledge Engine and Query Processor Component

Knowledge engine is a logical user interface. The knowledge engine, a logical user interface is set to accept the query from the user and the query is passed to the query processor where the natural language query is converted into SQL query. Figure 5 shows the query conversion process where

the natural language query is reprocessed with several phases such as morphological analysis, semantic analysis and mapping.
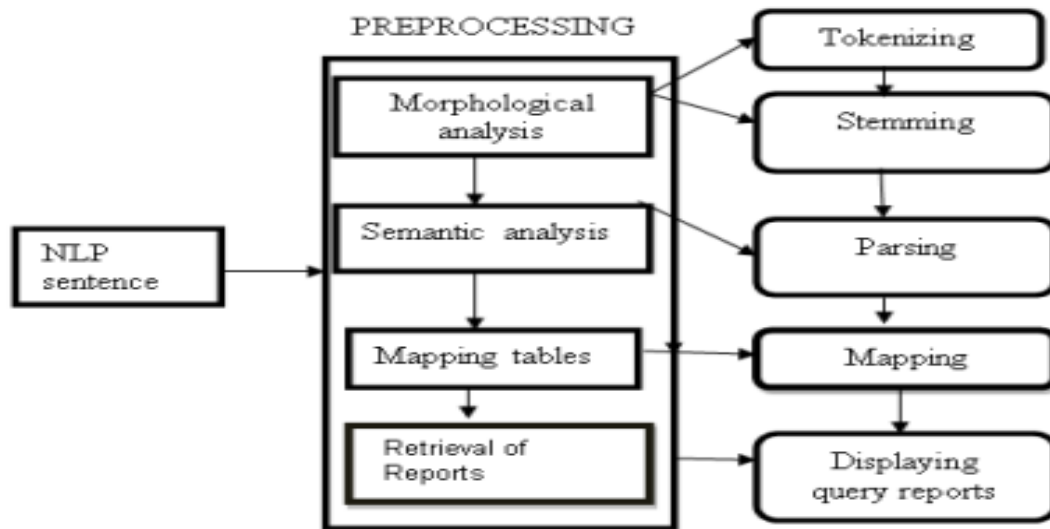

Figure 5. Block diagram query conversion process.

In Morphological analysis the user gives an NLP sentence as input and it is sent to tokenizer. The tokenizer splits the sentences into words based on whitespace character. The tokenized words are taken to extractor for stemming process. In the stemming process, the extractor maintains the collection of predefined words which are used for comparison with the incoming new words. Predefined words are most used words in the document for querying. It compares the tokenized words with the predefined and extracts the main keywords; the keywords are the words that are present in the predefined list of words. Then from the extracted words, the root words are identified. In the semantic analysis, the identified set of words will be given as input. The parse tree is generated through parser; object and verb present in the set of words is identified. The output of this analysis is collection of identified words. In Mapping phase, a mapping table consists of predefined set of SQL queries along with the maximum possibility of NLP words. The collection of identified words are mapped with best suitable query. The SQL query is generated at the end as a report, from which the query is picked and answered semantically. Knowledge engine also acts as a storage place for the knowledge base where the large pool of semantic data is stored after fetching the semantic data content from the query processor zone. The meaningful data which is stored in knowledge engine is used for the online recommender system. The query processor such as link parser, Stanford parser and masque/SQL interface were cross-examined initially and appropriate technique is used for query conversion. The parser used will pick the user query and deliver it to the right DB to fetch the solution for the requested user.
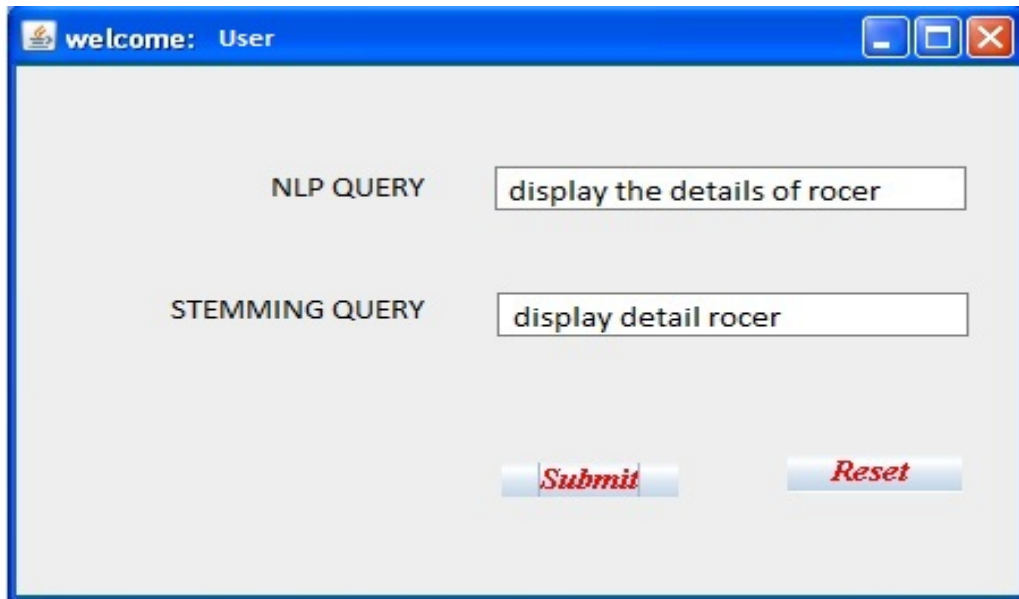
www.manaraa.com
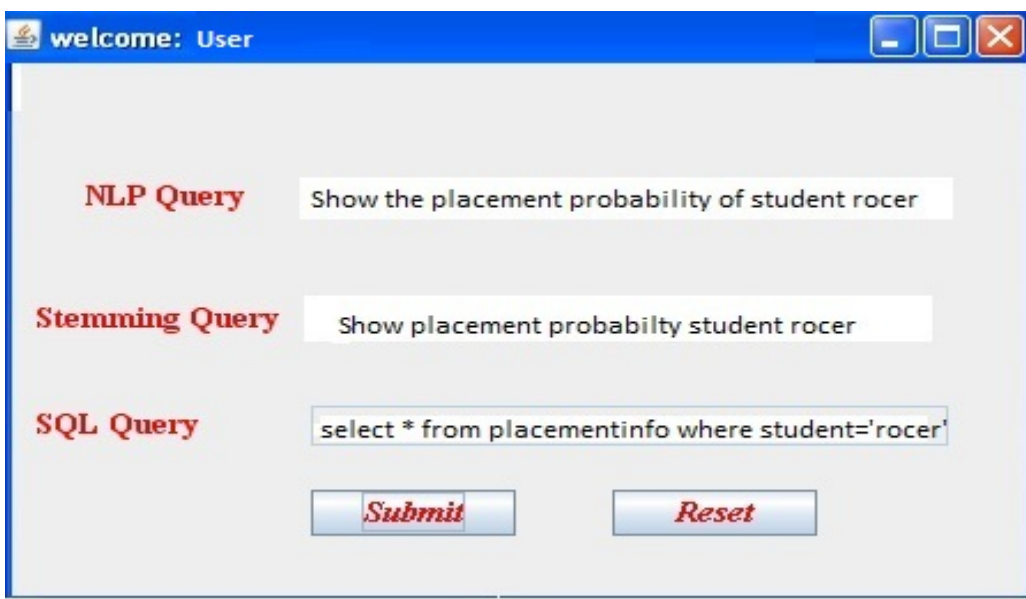
Figure 6. Query conversion.

Figure 7. NLP - SQL.

Figure 6 and Figure 7 represent the query conversion from natural language to typical structured query language, where the natural language query undergoes a series of steps in which the stop words and root words are removed initially and the stemming process is carried out to convert the natural language to structured query.

259

Establishment of Online Recommendation System

As discussed, the semantic content is collected in the knowledge engine. This semantic data is interlinked with the basic framework of online recommendation system which is used to provide a user with various options.
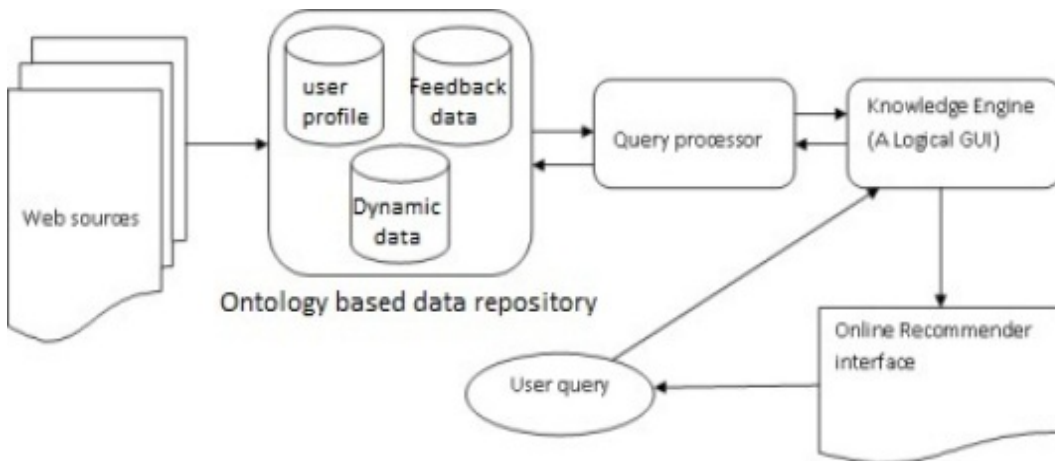

Figure 8. Block diagram for online recommendation system.

Figure 8 shows framework of online recommender system where the data is gathered from various web sources, and converted into a common data repository. User posts a query and appropriate online recommendation is provided to the user through online recommender interface.

In this research work, the student's professional data is collected and stored as common data repository from which the data is fetched for analysis. Here the data from the BSAU forum is collected which consists of student's professional data which possess both student data and professional corporate data. Therefore both the data is compared and mapped appropriately to recommend to the user the right career guidance and placement opportunity.

The user triggers a query from the GUI; this query processes and invokes the semantic data from the common data repository. This semantic data is analyzed with the help of knowledge based recommender system strategy and suitable recommendation is provided to the user.

Some of the aspects covered with the help of recommender system is to cope with information overload, to analyze academic performance of a student, to find the placement probability of student in an institute, to help all customers (new, frequent, and infrequent) to make decisions what products to buy, which news to read next, which movie is worth watching, etc., to convert

www.manaraa.com

observers to buyers, to build credibility through community and maintain the loyalty of the customers, to invite customers to come back, to enhance e–commerce sales and cross–sell.

Implementation

The implementation work is based on the collected web data from the educational forum referred as 'BSAU forum' where the student community share their academic and personal data. A simple case study is conducted by collecting the educational data from the BSAU forum which describes various attributes in an academic institute such as academic performance of a student, extracurricular achievement attained by the student, placement probability of a student, overall performance of an institution in terms of web rating, and so on.

This research work centered around the recommendation for the career guidance and placement probability for a post graduate student which is based on the current grade point average (GPA) score of a student, undergraduate score, analytical skill rating, softskill rating, overall performance rating and various other norms followed in the recruiting corporate sectors. When a query entered by the user such as 'Show the placement probability of the student rocer' this natural language query is converted to structured query with the help of query parser as discussed earlier.The converted structured query will be directed to the exact database by the knowledge engine where the collected data will be cross examined with the requested query. After a query conversion the exact data will be fetched.

On the other hand the desired data base is kept intact which is referred to as ontology based common data repository. After the implementation, the results exhibit the compatibility of the proposed framework. The placement criteria of various corporations are listed in Figure 9, which is based on the user ranking in terms of different recruiting procedures by different corporations.
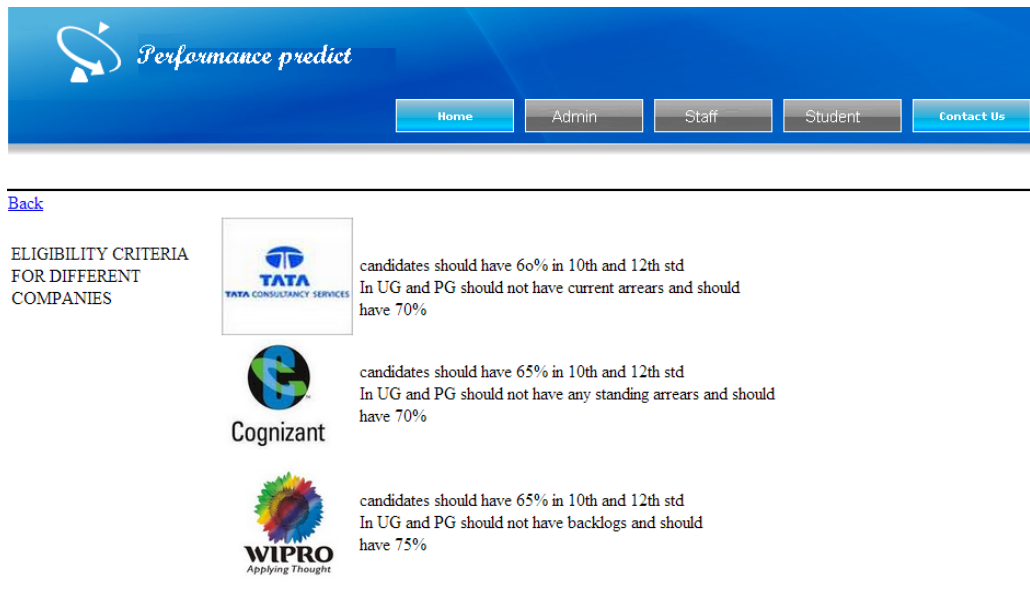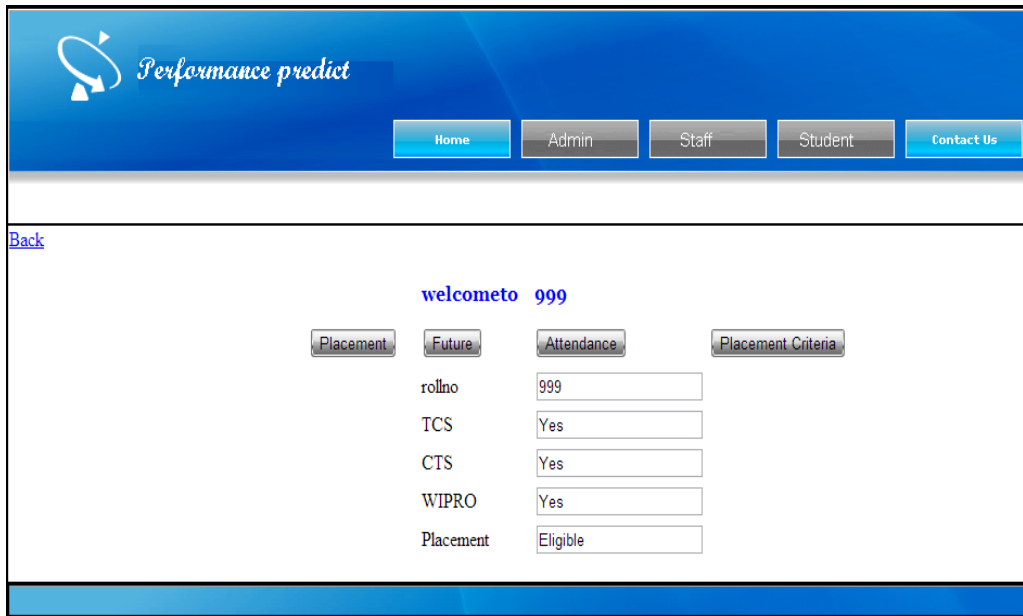
www.manaraa.com

Figure 9. Placement criteria.

The student eligibility criteria which must be attained by the student for the successful participation in the placement process is shown in Figure 10. The eligibility criteria is listed based on the placement standards which is maintained in all the corporate sectors which shows the statistical information about eligibility *criteria* of various companies which pursues the *student's performance* in their curriculum. The actual status of the student is also viewed through Figure 10 which represents whether a student matches the eligibility criteria of various corporations.

Figure 10. Student eligibility criteria.

The performance ranking of a student indicated in Figure 11 reveals the various factors such as academic performances, extracurricular activity and placement criteria. The ranking is based on the accumulation of all the performances. The performer can be of various classes such as high performer, intermediate performer, and so on.

Figure 11. Performance ranking.

263

The overall performance shown in Figure 12 represents the performance chart or performance graph which reveals the effective data which can be used by the institution for the efficient improvement in the performance of the student and placement probability in the future. Thus this simple case study with 117 students reveals the factors involved and user intervention in deciding the placement probability and academic performance of postgraduate students. The result shows the performance graph report for an individual which can be generated for multiples. Finally the requested user gets the appropriate reply through dedicated GUI with the help of the generated report which displays the placement probability of a student 'x'. The result reveals that the current system is worthwhile for knowledge sharing among the various web users for making a reasonable decision at crucial junctures.
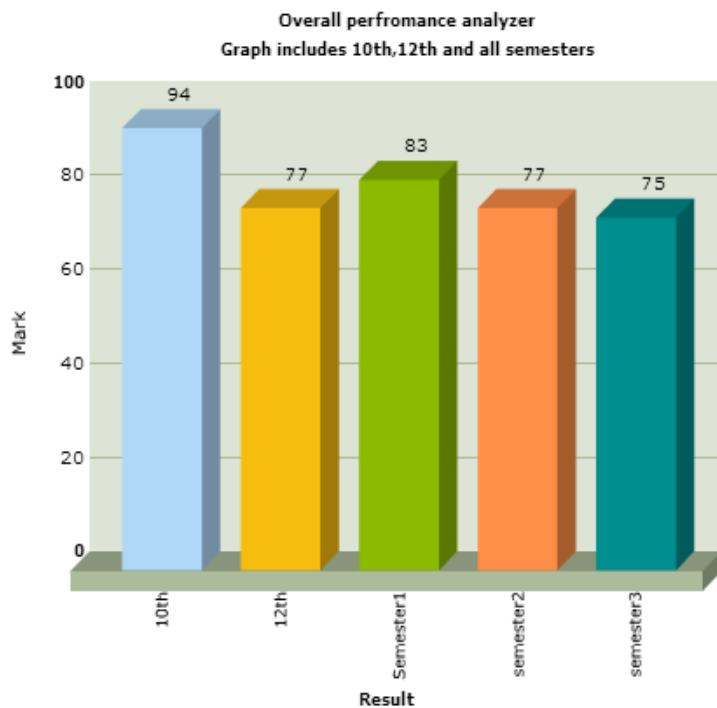

Figure 12. Overall performance graph.

Discussion

Firstly, the generic intention or objective of this research study is to provide knowledge among different users in the web to take right decision at a crucial juncture. The proposed framework

implements the online recommender application for the learners' career guidance who pursue their graduation in open and distance learning environment.

Secondly, the requirement of the objective is carried out in developing an application to predict the career probability of the students who pursue their graduation. To achieve the objective, the data is collected from common forum which is referred to as BSAU forum where the transactional data, such as current GPA, softskill rating, are extracted from the various social networks. The collected data is optimized with the help of R-tool and semantic interpretation is done by intiating the NLP parser.

Thirdly, the implementation phase is handled by investigating data from 117 students, from which the analysis is done and the career probability is calculated.The graphical report is generated which represents the overall performance and career probability of the student who enrolled in an open and distance learning environment.

Finally, the result obtained in query conversion and online recommendation is analysed with the help of common evaluation measures such as recall and precision.

Recall

A measure of the ability of a system to present all relevant attributes

$$recall = \frac{number\ of\ relevant\ query\ retrieved}{number\ of\ relevant\ query\ in\ sentence}$$

Precision

A measure of the ability of a system to present only relevant attributes.

$$precision = \frac{number\ of\ relevant\ query\ retrieved}{total\ number\ of\ query\ retrieved}$$

Precision and recall are set-based measures. That is, they evaluate the quality of an unordered set of retrieved attributes. To evaluate ranked lists, precision can be plotted against recall after each retrieved semantic attribute. To assist computing average presentation over a set of selected domain each with a different number of relevant documents individual topic precision values are interpolated to a set of standard recall levels (0 to 1 in increments of .1). The particular rule used to interpolate precision at standard recall level i is to use the maximum precision obtained for the topic for any actual recall level greater than or equal to i. Note that while precision is not defined at a recall of 0.0, this interpolation rule does define an interpolated value for recall level 0.0. The example takes up the query used in the above experiment, that is, the query based on placement enquiry "show the placement probability of student rocer" is measured with the recall and

265

precision and the graph is plotted for the total. The actual threshold range of precision and recall is plotted and illustrated in Figure 13, which displays the suitable range of word occurrence in the generated query. The values are measured precisely and generated as report.
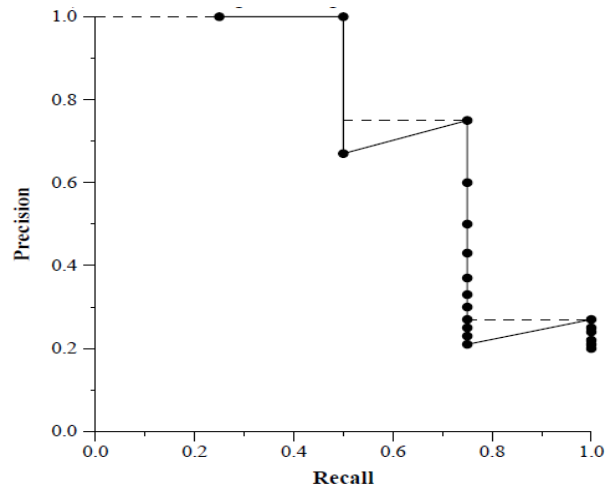
Figure 13. Interpolation of precision and recall threshold range.

The recall and precision level table is generated which provides the exact evidence for evaluating the semantic measure of the collected social web data which is gathered out of a user query in terms of natural lingual statement. The generated SQL has the relevancy threshold which is evaluated and listed in Table 3, which indicates the recall and precision range of each word pair in the generated SQL . The values are in the high cut-off range which enables the semantic nature of the extracted content from the different web sources.

Table 3

Recall and Precision Table for Generated SQL

Recall level precision averages	
Recall	Precision
0.00	0.5349
0.10	0.4789
0.20	0.4345
0.30	0.3790
0.40	0.2491
0.60	0.1284
0.80	0.0023
1.00	0.0014

The analysis of generated SQL results in finding the accuracy of the extracted data from various web sources and Table 4 represents the accuracy table for social web sources by traversing statement wise of a user query which is extracted from the various web sources such as Facebook, Twitter and LinkedIn which is the primary source of data.

www.manaraa.com

Table 4

Accuracy Table for Social Web Source Statement-Wise

Generated SQL averages/statement	
Per statement in web sources	Precision
At source 1.0 (Facebook)	0.4375
At source 2.0 (LinkedIn)	0.3567
At source 1.1 (Facebook)	0.3902
At source 3.0 (Twitter)	0.4793
At source 1.2 (Facebook)	0.2972
At source2.1 (LinkedIn)	0.1736
At source3.1(Twitter)	0.0023
R-Precision(Precision after R-source retrieved(where R is the relevant web source from which query is structured)	
Accuracy	0.3052

The relevancy attained per statement is calculated and measured for the accuracy range which is used to extract the semantic knowledge from the various web sources.

Conclusion and Future Enhancement

In this research paper, an online recommendation system for a decision making process was designed and implemented. The application implemented in this study is centered around 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. On that aspect this research article is quite helpful for open and distance learning arena. Here student academic performance, placement eligibility criteria, and student historical data are analyzed semantically to take decision on placement

268

probability**.** A graph is generated which indicates basic requirement, efficiency to help the students achieve better placement in their career. 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. Learner-content interaction explains the largest unique variance in student satisfaction. So with the help of healthy interaction, the knowledge sharing will take place to a large extent. The interaction among the users on the web through online social networks for reliable transmission of knowledge is conveyed through this research work.The performance of the learners is identified to a maximum extent, and the results obtained through this research work reveals the positive outcome of students' placement percentage in an educational environment.With the help of this study, the various factors which affect the performance of the students are identified and valid recommendation is offered for betterment in the future. This experimental study can be further expanded which can be applicable to other web users in various sites such as e-commerce sites, business sites etc. This research work can be enhanced with a few new techniques and algorithms which can be applicable for different types of web users and web sites. So future work could include the system which can meet different user requirements and improvised interfaces with sound technical definition which suits all kinds of environments.

www.manaraa.com

References

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.

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

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.

Edin Osmanbegović, & Mirza Suljić (2012). Data mining approach for predicting student
performance, economic review. *Journal of Economics and Business,volume 10*, issue1.

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

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

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.

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.

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.

Rabia Batool, Asad Masood 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.

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

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

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

Verónica Marín-Díaz, Ana Isabel Vazquez Martinez, and Karen Josephine McMullin (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).

Vo Thi Ngoc Chau & Nguyen Hua Phung (2012). *A knowledge-driven educational decision
support system.* Ieee Publications.

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).

www.manaraa.com