Original Research Paper

Applying Data Warehousing in University Information Systems in Kingdom of Saudi Arabia: A Conceptual Discussion

Hisham Al Saghier

College of Computer and Information Sciences,
Majmaah University, P.O Box 66, Al Majmaah 11952, Kingdom of Saudi Arabia

Abstract: The Kingdom of Saudi Arabia education system is on the right way to be one of the finest education systems among the world's education system. The most important aim is to develop the Saudi Arabia education is hub of the global map by improving the educational quality. The kingdom of Saudi Arabia focuses on Information and Communication Technology Systems (ICTS), Information Technology (IT), Computer Science, finance and banking, biotechnology, multimedia content development, advanced manufacturing and industrial design. For this presently, many universities in kingdom of Saudi Arabia depend on information systems to manage their day to day data transactions and information. The data stored in databases are helpful to the universities for conducting daily transactions of the university. However, the data is increasing hugely and to process this data there is a struggle to lead the theory and models of data mining and data warehousing where the data stored in databases are analysed to retrieve and create operative approaches to learn the best possible ways in carrying out the education in universities. In the way to execute the data mining concepts, the fixed data storage databases should be transformed into the informative data marts also known as data warehouse. This research work is focused on applying data warehousing in university information systems in kingdom of Saudi Arabia.

Keywords: e-Governance, Data Mining, Web Usage Mining, Information Systems and Data Warehouse

Introduction

Kingdom of Saudi Arabia universities are trending towards modernization. The universities are in the process of computerization of all the departments in the university. The information is stored digitally in the databases; in some universities certain departments are following both manual and computerized methods of data storage. This leads to create a huge data in databases at all levels of the university for this the information system plays an significant role in the development of the modern university. The university success depends on the university events automation and the level of that automation of departments in the university. Some specific universities frequently comprises many individual systems essential for its department level day to day transactions and processes of many individual subsystems which include the course management system, employees management system, staff system, accounting system, student registration system and many other individual subsystems. All sub information systems are interconnected to different fundamental databases that are used to store daily communications and procedures of the university systems. Though, universities employ information systems for managing prediction, forecasting, data analysis and decision making. The man aim of this research work is to develop a data warehousing architecture where information is stored and processed at all organizational levels, is stored only once to avoid repetitions and is accessible at all levels. To enable this there is a need to be data segregation and integration.

Literature Review

University information systems are created as the integrated systems of the different individual systems,

aiming to automate all the university processes and it uses Information Technology (IT) and internet to link all university departments with faculty, non-faculty and students to give them the best services wherever and whenever they are (Preeti *et al.*, 2011). Generally information systems has made universities more active in today's science and technology development. University Information system allow distinct users to avoid the different analysis of the same data which is stored in a consolidated database and it avoids the data duplication in database (Efraim *et al.*, 2006). Using the information systems the information is collected from different sources and the data is stored in database and managed.

University Information systems are the best suitable applications of data warehousing and data mining, because the information system field is easily suit the different concepts of data warehousing and data mining. The results generated using data mining concepts are speedily transformed into the university actions, at the identical intervals the guidelines of university resulting from the mining can be estimated in time. Data mining, concepts deals with the simplification of unidentified in addition exciting information from the huge amount of data in data warehouse (Laberge, 2011). Data mining is an extensively integrated technique, involving Computer language, informatics, Internet, Artificial intelligence, statistics etc. to process the stored data in databases. Mining data techniques are generally distributed into three different categories like information mining, structure and unstructured mining data and usage mining (Demarest, 2013). Information mining is also part of the data mining and it deals through the information available on information system databases or data warehouses; Structure or unstructured data mining is that part of information mining, it mainly emphases on the construction of data with unstructured data presented on data marts. User's usage mining is also a portion of data mining, which deals using the mining of knowledge on user access forms and user behaviour from data composed from the main data bases like clients including user profile information and registration data servers and proxy servers, using several data mining methods (Ahmad *et al.*, 2013). Both structure mining and content mining are focused on the raw data. In usage mining, the attention is on data describes the usage pattern of data stored in a warehouse; this data can be used to identify the behaviour of the data user and the motivation of user creating these data (Mate *et al.*, 2012). By manipulating the users data can mainly help university to recognise the users or the stakeholder's requirements, requirement needs in addition behaviours etc. to make matching guidelines.

The data warehouse meant for university information system is university data warehouse. The data warehouse can be treasured resource for all the university information system users of the university that need data or information to provide support for their daily operations, performance administration, decisions support and compliance broadcasting strategic planning (Saadatdoost *et al.*, 2011). In today's era, the users are in urgent need for data warehousing in university information systems for the speedy decision making. University has to deal with enormous data and when it is used information systems accurately, it helps the decision makers and planners in making proper conclusions to the growth of the university. Due to absence of data warehouse in university, university have to depend on domain specific information technology staff for generating special reports and charts based on the requirement, it takes long time to produce those reports and to take decisions. Data warehousing and data mining concepts can deliver tactical intelligence to the decision makers and deliver awareness into the complete situation from the historic data in databases. With the help of decision makers, university can easily take their basic level decisions in a timely manner without depending on their IT staff. The dimension design process in this research work follows the methodology of (Bhanti *et al.*, 2011).

This research paper is structured as follows: Firstly, in section 1 a brief introduction is given. In section 2 the literature review related to the need for data mining and data warehouse concepts in information systems are discussed. Then in section 3 the proposed data warehouse model is discussed in detail following in section 4 the benefits of the proposed model are discussed. Lastly, in section 5 the conclusion and future work is discussed.

Proposed Data Warehouse Model

This segment discusses the proposed informational system data warehouse model along with its components and conceptual schema. In Fig. 1 the overall framework is shown.

To build the information system data warehouse, four vibrant stages are mandatory to accomplish the aim of this research work and those steps are correspondingly: Data identification from source data process and data cleansing, transform and Data extraction from the data sources and loader and data index. The Fig. 2 shows the different components that are mandatory to transform the effective data stores into informative database or data warehouse. The data are first mined and then the data is stored provisionally into a buffer of data staging area by using the earlier designed operative databases as a data source and it is pre-processed it includes different stages like cleaning

and integration data, fixing data entry mistakes and altering data into a more standardized data. The altered data are loaded and indexed into the data warehouse once the data is completed the mining stage. In this entire process, new tables are created, columns are discarded, the tables are dropped and new columns are added to alter the data.

The overall design of the proposed data warehouse model for university information system is developed by using the data dimensional modelling method. The traditional object relationship model, unified modelling language, data modelling processes and the analytical design of data marts or data warehouse are defined by the data dimensional modelling method. To minimalize the stored procedures for join operations in the database which slow down user queries, database or table normalization is not the guiding principle in data warehouse design process. Where a data warehouse schema or design is a collection of data objects, including views, indexes, synonyms and tables. A diversity of methods are used in arranging representation of objects in the data schema models designed for data warehousing.

The data dimensional method is relatively diverse from the traditional database normalization concepts followed when scheming a database for daily processes. Each of the dimensions in data warehouse contains at least one level of hierarchy. The hierarchical orders in data warehouse allow users to analyse data combinations using the multi-dimensional analytical methods. Multi-dimensional analytical approach allows associated items to be assembled and briefly explained for high level analysis though recalling the data to minimize the specific product detail. Here the data dimensional tables keep the data that are used for analysing the data warehouse. The Fact tables in the hierarchy contain dimension characteristics and procedures. The attributes in dimension features are foreign keys and these foreign keys are also called as degenerate dimension. When queries are group rows together then the events are the different principles to be accumulated in the data ware house. Fact tables are designed using of two different types of data attributes: Measures and dimension attributes. By applying data dimensional approach the data will be processed and stored smoothly without losing the values in it. Here each component in the proposed model is discussed individually.
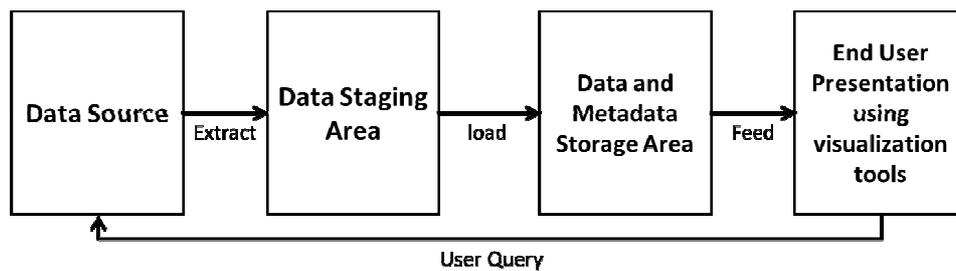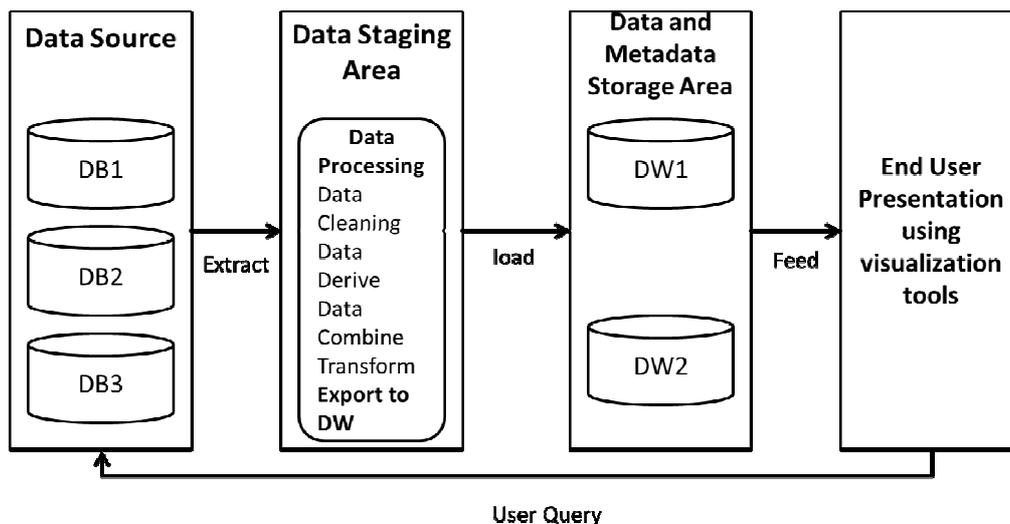

Fig. 1. Overall framework

Fig. 2. Overall proposed data warehouse model

Data Source

In this component, information and data should collect from different sources of the university. Once the data is fetched from the data sources it will sent for the data staging process to process the data. Data set considered in this research work is a student database (Source: https://archive.ics.uci.edu/ml/datasets.html) which contains 50,000 records and 20 attributes in each record with different data types. The data types considered in this study are numerical data and categorical data. Some of the attributes which is considered in this study are REG_NO, NAME, YEAR, SEMESTER, COURSE_CODE, ASSESSMENT PARAMETER, MAX_MARK, MIN_MARK, SPECIALIZATION, PLACE, STAFF_ID, SCHOOL_EDUCATION, FAMILY_DETAILS, MOBILE_NUMBER1, MOBILE_NUMBER2, SOCIAL_NETWORKING_ID, E-MAIL.

Data Staging

The data in university information systems are provided by the primary data sources are usually imperfect, redundancy, unpredictable and obscure. In data processing stage converting the traditional data contained in the various available primary data sources in university information system into the data needed for convenient pattern discovery. Using the data mining concepts like data clean, accuracy and simple abstract data can be used for analysis of user predictions and regulations. In processing stage usually it comprises of data cleaning, transaction identification, data integration and data transformation. Users can take huge benefit of decision makers as they can view frequently asked queries whose result are stored in the database. Here the data extracted using the data mining model from the data source will be processed using extract, convert and data loader tool this data tool gathers the clear information from databases then fit those information with operative needs of the users of the information system and then the data will be load into common warehouse.

Data processing stage refers to eliminating all the data traced in data logs that are unacceptable for information mining purposes and to keep the data that carry appropriate information. Then, the data cleaning stage is used to remove the inappropriate records from the databases or the data logs. The data collected from different sources, which stores the communications between different users and multiple subdivisions of the university; usually there is a need to be data cleaned before use for mining. In divergence, the data collected from data source are appropriately cleaned as of less user intervention. Additionally, the input data given by users should be stored, formatted and restructured for pattern discovery.

Data Warehouse

The processed data along with its metadata will be stored in data warehouse. The data storing in data warehouse depends on the data type and time like the historical information and current information related to the query. Collective data warehouse information gives growth in information distribution between university's subdivisions or individual systems. It allows the communication among users or departments in different universities. Similarly it allows quick response to staff's queries because of universities information saves in huge warehouse.

User Interface

A user interface will be available to all the users of the information system to access the data warehouse which stores the processed information. The major benefits of using data warehouse in university information system are the point of view of providing the information sharing and interaction between departments in the university.

Proposed Algorithm

Step 1: Start
Step 2: Determine the data source from different data sources to begin with.
Step 3: Extract the tables from the data source to process.
Step 4: Create a data mart to store the processed data and table index
Step 5: Process the data related to user query
Step 6: Store processed data in warehouse
Step 7: Store the metadata in data warehouse
Step 8: Iterate the results and present to the user.
Step 9: Repeat step 5 to step 8
Step 10: Stop.

Implementation and Results

The proposed research framework using data mining and data warehouse concepts is implemented using oracle 10 g as database. It contains the advanced individual relations interrelated with the tables in the database. This system is tested by different users of the university to calculate the accuracy and time taken to process the query in Table 1 the results are tabulated and in Fig. 3 these results are compared. In the proposed model the user can give more complex quires which cannot be processed in traditional model also.

Based on the research model constructed above, the information system is used to carry out data warehousing and data mining concepts for the university information model. It will be analysed from few aspects and those are discussed in detail below along with graphs.
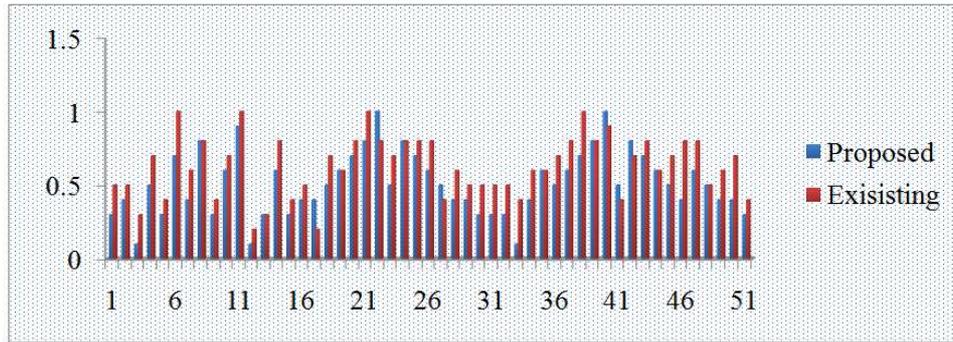
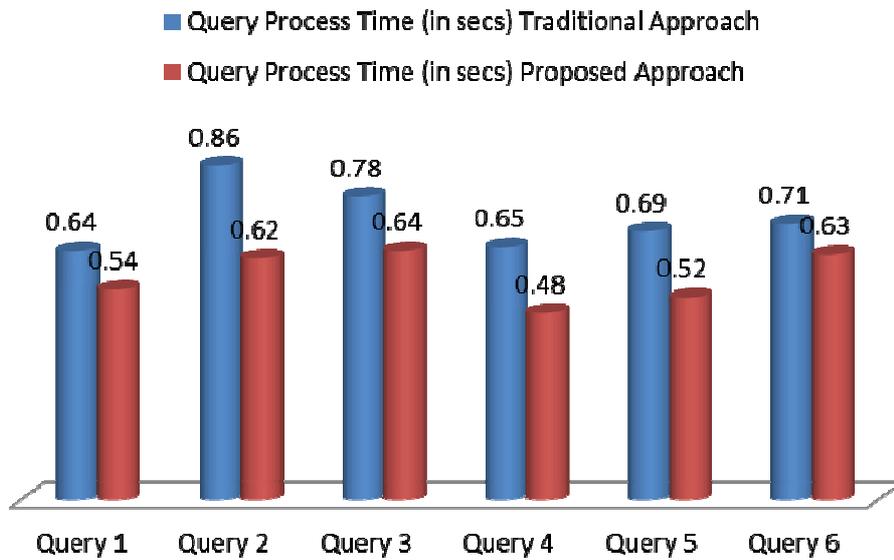Fig. 3. Results comparison between proposed and existing methods (Without Data Warehouse)

Fig. 4. Performance comparison between traditional (Without Data Warehouse) and proposed models
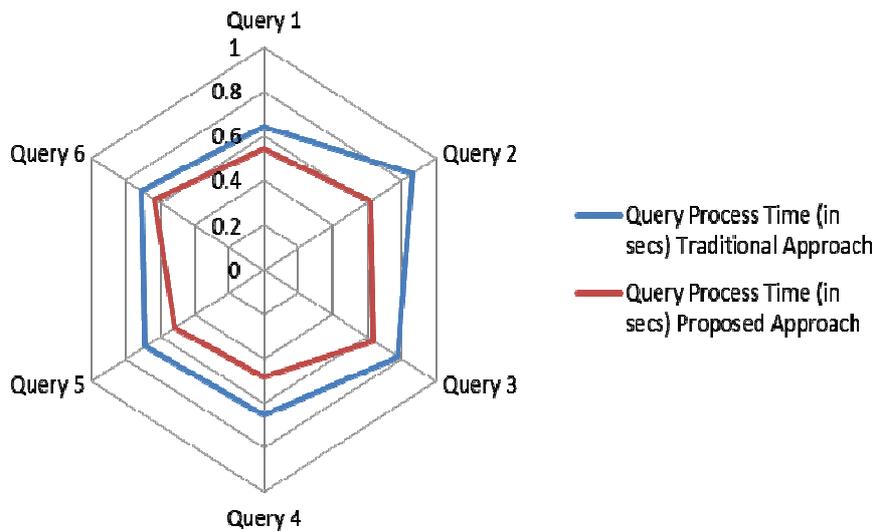

Fig. 5. Performance comparison between traditional (Without Data Warehouse) and proposed models
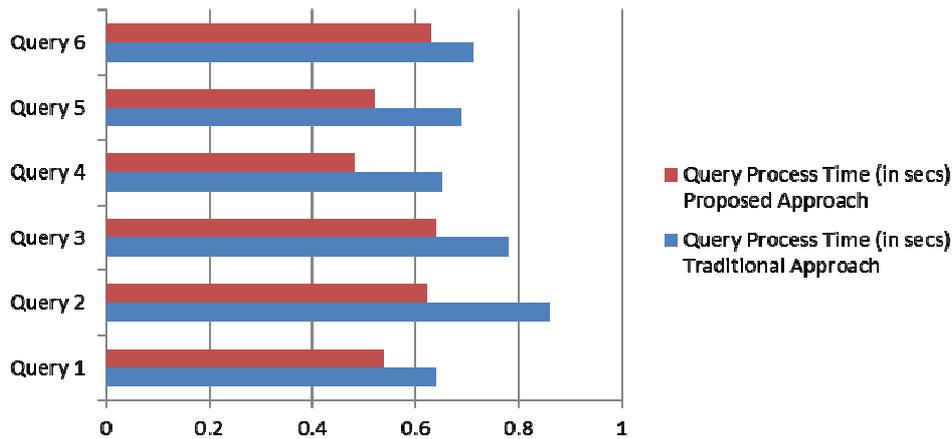
Fig. 6. Performance comparison between traditional (Without Data Warehouse) and proposed models
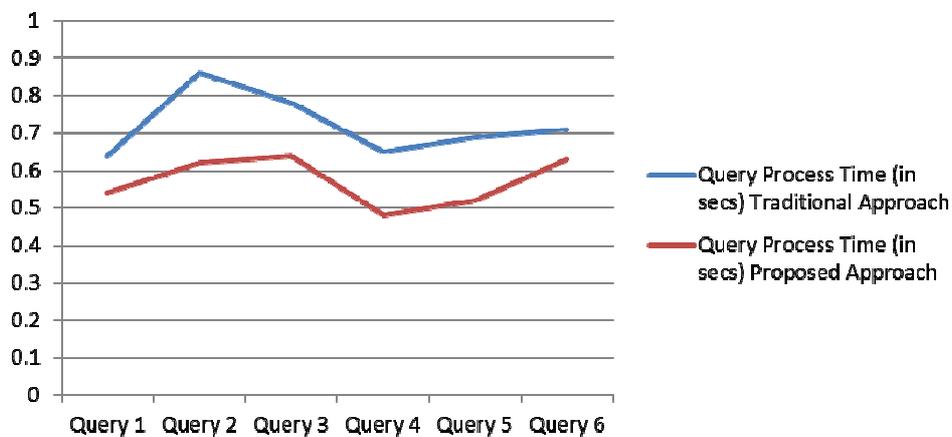

Fig. 7. Performance comparison between traditional (Without Data Warehouse) and proposed models

Table 1. Time taken to process query

User query	Query process time	
	Traditional approach (Without data warehouse)	Proposed approach
Query 1	0.64	0.54
Query 2	0.86	0.62
Query 3	0.78	0.64
Query 4	0.65	0.48
Query 5	0.69	0.52
Query 6	0.71	0.63

Analysis of User Query

The results retrieved using the university information system using data mining and data warehouse is illustrated in Fig. 4 based on the time taken. It is seen that user given query in proposed model occupies the maximum proportion than the traditional model. Here different users from the university are given the quires. It is important for the university to receive more effective and timely results related to the given user query. University stakeholders should adjust the rules and regulations accordingly to provide higher quality of decision making for university growth in advance.

Analysis of User Attention

In Fig. 5, the user attention when entering a query is related to the time has shown. Majority of the users are given complex queries to both systems and in the proposed system the user feel comfort to enter query and the growth rate is about 43.17 percentages.

The Fig. 6 and 7, concludes that the proposed university information system gives more effective results compared to the traditional model. From Fig. 6 the results are absorbed that the user ease of use towards the proposed model rather than using the traditional models are high and based on the time it was driven. In the same way Fig. 7 shows that the user trends over the traditional model the proposed model is high and in the graph also it shows that the time taken to process different user quires are lesser than the traditional model.

In the proposed framework all data stored in the database were processed successfully using data mining

concepts; they are exported from their buffer zone to the data warehouse just built. As the results shows that the data mining can be discover patterns and trends out of the warehouse that are necessary for university development to perform decision making. The expected benefits of the proposed model are outlined below:

- Better student satisfaction
- Benefits the faculty member
- Promotes student academic success
- Enhance efficiency
- Improve overall advising quality

Conclusion

Through a review of the literature, this research paper has described the importance of data mining and data warehousing concepts in Saudi Arabia university information systems and discussed various techniques that can be adopted. The purpose of data mining and data warehousing is to narrow down the data process by eliminating the noisy data related to the given query and to create effective decisions for the university growth. In this research work, a conceptual based framework for university information system using data mining and data warehouse was presented. This framework consists of different components to process the data. This framework can deliver a huge amount of support for the different users of the information system within university as the user needs to extract the unseen and implicit data or information of the students, faculties and stakeholders of the university. Important applications of information mining in university information system were discussed, it gives a systematic logical help for decision making and for observing the revisions of the regulations of a university. Decision makers in university should completely make use of this treasured information which is mined from the traditional database to moderate policies and practice consequently of university information system to satisfy faculty and students and improve the university in a smarter way.

This research work demonstrated that the proposed model performs efficiently than the regular model. Improving the quality and size of the data warehouse and data mining techniques could improve results. Promising results could be achieved with little effort by fixing the data inconsistencies and populating the mining concepts. Another point to consider for future work to improve the system includes correcting the data mart errors using big data concepts.

Acknowledgement

Ethics

This research study complies with the all legal requirements and also follows the international standards.

References

Ahmad, O., I. Hyder, R. Iqbal, M.A.A. Murad and A. Mustapha *et al*., 2013. A survey of searching and information extraction on a classical text using ontology-based semantics modeling: A case of Quran. Life Sci. J., 10: 1370-1377.

Bhanti, P., U. Kaushal and A. Pandey, 2011. E-governance in higher education: Concept and role of data warehousing techniques. Int. J. Comput. Applic., 18: 23-41. DOI: 10.5120/2249-2879

Demarest, M., 2013. Data warehouse prototyping: Reducing risk, securing commitment and improving project governance. CEO of Noumenal, Inc.

Efraim, T., E. Jay, L. Teng-Peng and S. Ramesh, 2006. Decision Support and Business Intelligence Systems. 8th Edn., Prentice Hall, ISBN-10: 0131986600, pp: 850.

Laberge, R., 2011. The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights. 1st Edn., McGraw Hill Professional, New York, ISBN-10: 0071745327, pp: 416.

Mate, A., J. Trujillo, E.D. Gregorio and Y. Song, 2012. Improving the maintainability of data warehouse designs: Modeling relationships between sources and user concepts. Proceedings of the 15th International Workshop on Data Warehousing and OLAP, Oct. 29-Nov. 02, Maui, HI, USA, pp: 25-32. DOI: 10.1145/2390045.2390050

Preeti, S., R. Srikantha and P. Suryakant, 2011. Optimization of data warehousing system: Simplification in reporting and analysis. Int. J. Comput. Applic., 9: 33-37.

Saadatdoost, R., A.T.H. Sim and H. Jafarkarimi, 2011. Application of self organizing map for knowledge discovery based in higher education data. Proceedings of the International Conference on Research and Innovation in Information Systems, Nov. 23-24, IEEE Xplore Press, Kuala Lumpur, pp: 1-6. DOI: 10.1109/ICRIIS.2011.6125693