# Improve Knowledge Visualization through an Interactive Graph-based Dashboard System with Key Performance Indicator: A Case Study of University Dashboard for Higher Education.

[1]**Wan Maseri Binti Wan Mohd,** [2]**Abdullah Embong** [,3]**Jasni Mohd Zain,**

**Faculty of Computer System & Software Engineering**
**University College of Engineering & Technology Malaysia**
[1]**maseri@kuktem.edu.my,** [2]**aembong@kuktem.edu.my,** [3]**jasni@kuktem.edu.my,**

## Abstract

We propose a novel approach of knowledge visualization technique by generating knowledge through a graph-based clustering technique, incorporating tacit knowledge and finally visualizing the knowledge in dashboard form. Three aspects are emphasized: clustering technique, knowledge visualization and domain knowledge interaction. Clustering technique with the feature vectors helps user to mine data from the enterprise database for pattern recognition. The knowledge visualization helps users to analyze the comprehensive characteristics of the objects under study. Subsequently, domain knowledge interaction shall enable the human's perceptual capabilities to be transferred, thus increases the intelligence of the system. The knowledge visualization is enhanced through the dashboard concept which incorporated with Key Performance Indicator where it provides significant patterns of knowledge on real-world. The framework consists of the knowledge visualization model and a prototype system design for higher education environment is presented in this paper.

Keywords: Data Mining, Knowledge Discovery, Data Visualization, Knowledge Visualization, Higher Learning Dashboard

## 1.0 INTRODUCTION

For many tasks of exploratory data analysis, visualization plays an important role. Data visualization is widely accepted by the medical, scientific, engineering, entertainment and business industries due to its ability to add value to existing data, but have not been widely studied for applications in higher education data analysis. It is critical for the higher educationist to identify and analyze the relationships among different entities such as students, subjects, lecturers, environment and organizations to ensure the effectiveness of their important processes. Furthermore, higher education will find larger and wider applications for data visualization than its counterpart in the business sector, because higher education institutions carry three duties that are data mining intensive: scientific research that relates to the creation of knowledge, teaching that concerns with the transmission of knowledge, and institutional research that pertains to the use of knowledge for decision making [1]. All the above tasks are well explained within the boundaries of data visualization, which drives the need for better visualization methods and tools.

## 2. RELATED WORKS

### 2.1 Visual Data Mining

The main reason for data visualization is the limitation of human beings to absorb the large amount of information. The volumes of data are overwhelming and the human visual systems and brain are not equipped to work with the data in this form [2]. Using data visualization, we allow much faster processing of the data and the ability to see the patterns in the data. On the other hand, data visualization should be designed to simplify visualization as much as possible to help human make more correct decisions. Less effort should be spent on visualizing data, but more on visualizing the results of data mining and helping the users to understand them. The visualization tools should not confuse an average user with an overwhelming set of choices for visualization. Instead, they should guide the user towards the most appropriate visualizations for the task. In brief data visualization is a method of presenting the output so that the entire problem and the solution are clearly visible to domain experts [3]. Data visualization uses graphical and numerical tools to reveal the information contained in data [2]. It is more effective approach for understanding and communication than the use of common numerical tools. Graphical methods hold a key for visualization. The best visualization method is one, which supports the most insight into the phenomenon under study.

### 2.2 Data Visualization Technique and Software

Visual representation of data has evolved rapidly in recent years. Ankerst (1999) gives a comprehensive overview over existing visualization techniques for large amounts of multidimensional data [4]. In general, visualization techniques are powerful tools that are frequently employed in knowledge discovery processes. Visualizations can make complex relationships easily understandable and stimulate visual thinking [5]. Especially, tools which visualize the cluster structure of data are valuable for exploring and understanding data. Such tools include data histograms for one-dimensional data as well as algorithms which project high-dimensional data to a two-dimensional visualization space trying to preserve the topology [5].

Recently, several techniques of data visualization have been introduced. Among the popular techniques are independence diagrams [6], equiwidth histograms, correlation coefficients and scatterplots, geometric techniques, icon-based techniques, pixel-oriented techniques, hierarchical techniques, graph-based techniques and dynamic techniques [7].

## 2.3 Graph-based Visualization Technique

Over the years, numerous approaches for visualizing non-quantitative, structural, set-theoretical, and relational information have been proposed. Not surprisingly, the diagrammatic paradigms employed are essentially topological rather than geometrical in nature. The term *topovisual* has been used to describe these formalisms [9]. Two of the best known topovisual formalisms are the **graph** and the *Venn diagram*. Both have their roots in the work of Swiss mathematician Leonhard Euler (1707-1783). The graph topovisual formalism is particularly well-suited for depicting relational data. A graph', in its most basic form, is simply a set of vertices connected by edges (undirected) or arcs (directed) and is commonly used to represent a (single) set of elements and its relation. Graphs can be modified to support several types of vertices and edges, each representing different types of elements and relationships. Bondy and Murty (1976) provide an excellent introduction to graph theory [10]. The basic data structures and algorithms for manipulating graphs are out-lined in [11]. A general introduction to database theory can be obtained from [12].

An example of graph application is the re-presentation of a relational data consisting of entities and the relationships among them which can be depicted by the vertices and edges, respectively, of a labeled graph. Edges are used to represent non-directional relationships between two entities. Multiple relationships can be represented within a single graph by displaying the name of each relationship as the label of an edge or arc, or by using different colors or line styles to render the edges or arcs associated with each relationship [13].

Beside the graph structure, interactive graph-based interfaces have been discussed in various studies. Interactivity is desirable to browse and edit data. However, as graphs increase in size, interactive interfaces risk information overload and low responsiveness.

Subsequently, Freire (2004) highlighted the challenge of preserving the mental map during graph navigational actions. Performing incremental layouts after each navigational or editing step implies that there will be greater differences between successive views, and therefore greater care is required to preserve the mental map during interaction. Freire (2004) have also identified a series of general "factors" that contribute to mental map preservation which are predictability, degree of change and Traceability.

## 2.4 Interactive Data Mining

The importance of user interaction and data visualization in the knowledge discovery and data mining process to help in finding understandable patterns for humans has been identified by Fayyad et al. [23]. However, most existing tools are developed for limited interaction with users and without human domain knowledge. Generally, whenever the user is required to choose datasets and sets of some parameters related to algorithms, the selection of these parameters is tedious and difficult to determine a priori. Recently, Ankerst et al. (1999) propose strategies to assimilate human domain knowledge and integrate data visualization and user interaction into knowledge discovery procedure for three important reasons: (1) With the help from data visualization, the capacities of human to find useful patterns can be greatly improved; (2) The users will trust the created patterns from this interaction process; (3) Domain knowledge of users can steer the data mining process [24].

Subsequently Jiang ( ) highlighted that the interpretation of co-expressed genes and coherent patterns mainly depends on the domain knowledge, which presents several challenges for coherent pattern mining and cannot be solved by most existing clustering approaches [21].

Another reason for user interaction is that different users at different occasions may like to mine on different portions of data, for different kinds of knowledge, and with different requirements and constraints, and moreover, they would like to interactively refine their mining queries and perform drilling/dicing to progressively deepen their mining process [39].

Among the Data Mining products which incorporating the interactivity features are CHARM, MOBIMINE, GENEX, JRV system.

## 2.5 Digital Dashboard

A digital dashboard, also known as an enterprise dashboard or executive dashboard is defined as a business management tool used to visually ascertain the status (or "health") of a business enterprise though key business indicators. Digital dashboards use visual, at-a-glance displays of data pulled from disparate business systems to provide warnings, action notices, next steps, and summaries of business conditions (33).

Historically, the idea of digital dashboards follows the work in the 1970s with the study of decision support systems. In the late 1990s with the surge of the web, digital dashboards as we know them today began appearing. Many systems were home built as the emphasis on efficiency became a passion. Today, digital dashboard technology is available "out-of-the-box" with many software providers on the scene.

The concept of "information overload" has been the driving force behind the creation of Digital Dashboards. In a typical organization there are so many different sources of pertinent information from which to make decisions, yet it can be so difficult to obtain the information into a format that is truly useful [17]. The term Dashboard has transcended the automotive arena to enter the business world hand in hand with visualization.

The main function of a digital dashboard technology is to synchronize information in one visible and easy-to-read place. Hence a digital dashboard is a graphic representation that contains a series of gauges and depictions that summarize the state of the company, be it financial, sales or more generally of any indicator

that allows you to know the situation, possibly in real time, of your business [18]. This way an executive or anyone else who needs it in the organization, can instantly see how the business is evolving, deciding in which areas it is necessary to act in order to correct some behavior that could potentially deviate the expected results. Furthermore, Dashboard provides internal and external information with single-click access to analytical and collaborative tools. Best of all would be dashboards that could provide significant patterns of knowledge based on real-world and theoretical modeling. We would call that wisdom. Typically, business dashboards transform data from other, more complex, industrial-strength data-analysis engines, into information, providing key financial summaries.

Thus, Dashboard provides everyone in the organization with insights for managing organizations large or small. Adding organization goals to contrast with the reported actual results in instant accountability – people are held to standards; the goals are revealed to be attainable or not before it's too late to change course. Securely exposing dashboards across the organization, so that many levels can compare their performance against objectives, results in a system that quickly and consistently conveys corporate goals, unfiltered, across the enterprise.

### 2.6 Key Performance Indicator (KPI)

KPIs are the key elements of dashboard visualizations since there are the most effective means of alerting users on their goals and objectives achievement. A KPI is simply a metric that is tied to a target. A KPI is usually tied with a metric. Metrics are referring to a direct numerical *measure* that represents a piece of business data in the relationship of one or more *dimensions such as* "gross sales by week". Most often a KPI represents how far a metric is above or below a pre-determined target. KPIs usually are shown as a ratio of actual to target and are designed to instantly let a business user know if they are on or off their plan without the end user having to consciously focus on the metrics being represented [28].

### 2.7 Data Visualization in Higher Education

Major applications of data visualization identified for higher education institutions are alumni, institutional effectiveness, marketing and enrollment management; for example, data visualization which integrated with KPI allows the higher institutions to act before a student drops out, to plan for resource allocation from knowing how many students will transfer or take a particular course, to identify those who are most likely to donate or participate in alumni related activities, to identify how do students learn best, what courses are often taken together, what learning experiences are most contributive to overall learning outcomes [1]. Although there are several studies done on data visualization in higher education, the effectiveness of the visualization system is still an open area to be explored. The study proposes a holistic approach of visualizing the data in higher education environment.

The motivation for this paper is to develop a framework of Graph-based Dashboard Visualization system, a visualization tool, for knowledge discovery in higher education environment which integrates the dashboard, key performance indicator and graph-based concept of visualization.

## 3. THE FRAMEWORK OF GRAPH-BASED DASHBOARD DATA VISUALIZATION SYSTEM FOR HIGHER EDUCATION ENVIRONMENT

As the computer system become more advance, the demand for better visualization of data increases. While most of the existing research in data visualization especially in data mining area focus more on the visualization of data through graph, which is still does not give the overall meaning of data to the users. In this paper, several concepts and techniques has been combined and integrated to improve the data visualization, hence improve the understanding of the data. The proposed framework has the following characteristics: (1) User friendly interface, (2) Representing the overview of the data dimensions using dashboard concept or single visual cue, (3) Key Performance Indicator which reflect the current status of the organization compared to predefined target and (3) Providing graph-based visualization interface using cluster and feature vectors view with the color scheme to understand the insight of the data while preserving the mental-map.

### 3.1 Framework Design

Figure 1.1 illustrates the framework of the new visualization system for higher education. The system is designed in 3 structured layers to provide a systematic, natural and convenient visualization of data [9]. In a layered graph, each layer corresponds to an abstraction level that contains a view of a collection of data starting from overview and drill down further into the detail to zoom insight the data. The first layer represents the summary or overview of data which implies the health of the institution in dashboard form, while the second layer provides the users with the insight of the data to further understand the internal pattern through the cluster view in graph form, and the third layer provide the behavior of the cluster through feature vectors view for further action. In higher education system, each level of organization's hierarchy needs a specific view of the data. Thus, the proposed framework is designed to fulfill each level's need by providing different view for each level of user in the institution. The first layer which is the dashboard layer is designed for management members of the organization to visualize the overall health of main subject in the organization by comparing the current state of the organization with the predetermined Key Performance Indictor in the dashboard. The idea of summarization of data from cluster view to dashboard view is motivated from several streams of previous work. The concept of summarizing data through aggregation and roll-ups concept has been applied in this study. Aggregation of nodes and edges according to various criteria is a well-known technique for simplifying complex graphs [16], and there has been recent work on the idea of letting the user dynamically control the level of detail by aggregating based on topological criteria [16]. The second layer which shows the cluster of the object such as student is designed for implementer such as lecturers in higher education institution. The cluster of student is important to help the lecturers to understand the overall pattern and identify clusters need to be further investigated for necessary action. The clusters are generated by a data mining process with clustering technique namely MaxD K-means clustering [19].

MaxD Clustering technique is identified as the most appropriate clustering technique for the study because it does not need any parameter to be input for the clustering engine. In higher education environment, the need to explore the unknown is highly demanded. The last layer which is feature vectors layer explains the behavior of each cluster to give more information to the implementer to take further action. The idea is motivated from the study done by jonyer (2001), whereby the data is presented as feature vectors and represented as a collection of small, star-like, connected graphs. Furthermore, the interaction with the domain knowledge at dashboard layer shall allow them to change the key performance indicator and the system shall trigger the clustering process to reprocess the clustering algorithm based on the input. Subsequently in the cluster layer, the interactivity shall improve the clustering process based on domain expert's inputs. At the third layer which is feature vector view, the domain expert shall facilitate the system to define more attributes for each particular cluster. These interactions shall improve the intelligence and accuracy of the clustering process and at the same time, the system facilitates the knowledge forecasting.
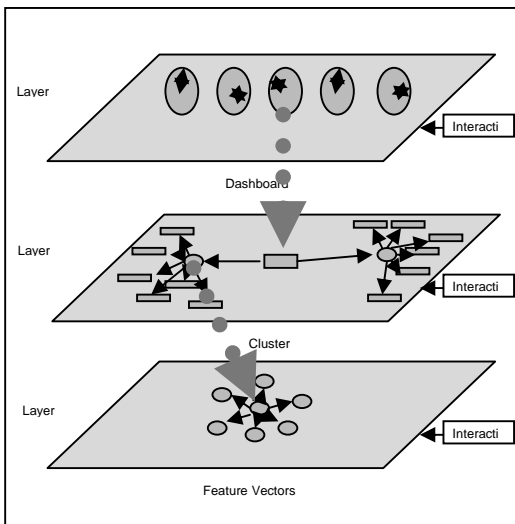


Figure 1.1 Framework of Graph-based Dashboard Data Visualization System

## 3.2 A Prototype of Higher Learning Dashboard - Forecasting First Class Student

The prototype of First Class Dashboard System is designed to visualize data from the student's result database by recognizing the pattern through clustering algorithm, and at the same time capturing the knowledge and experience of educationist such as professors, lecturers, administrator and students themselves. The process of preparing the Dashboard is divided into three steps. The first step is to capture the initial inputs. The first input is to identify the default factors that help students to get first class's result. The default factors are captured from education experts as tacit knowledge and stored in the factor's table to be used by clustering engine. Then, the second input is the Key Performance Indicator (KPI) of first class student. The second step will be the processing part. The first process starts with deriving the current state of student's result by clustering the students based on their

result using MaxD clustering engine. During each clustering process, the new factors which are identified as feature vectors will be generated and added in factors' table together with the existing factors captured from education experts. Subsequently, the forecasting engine shall calculate the forecasted clusters of students by applying the factors in factors table. Then, from the generated clusters, the number of forecast first class student is generated through aggregation and roll-up engine. Finally, the forecasted number of first class student is visualized in dashboard form together with KPI and current number. The dashboard then can be drilled down to visualize the clusters, and for each cluster, the feature vectors or cluster's behavior can be visualized. The dashboard can further drilled into list of students associated with the behavior and finally the detail profile of the student can be visualized. The third or final process is the process of improving the number of first class student. In this step, first, the forecasting engine shall derive and visualize list of proposed actions to be taken by each entities. The system shall use the factors table to derive the proposed actions. Then, at the same time, the system shall capture the knowledge and experience as tacit knowledge from education experts, lecturers and students themselves on the actions to be taken by related entities in order to improve the forecast number of student by improving each student's result. At this stage, we are combining the auto-generated and tacit actions to improve the accuracy and correctness of the actions. All the steps are illustrated in Figure 1.2 to Figure 1.8.

### 3.2.1 Capturing Default First Class Factors for Data Clustering Process

Figure 1.1 illustrates the screen to capture the initial factors affecting the student to get first class result. This screen shall be used by education's expert such as professors, senior lecturers, deans and administrators. The tacit knowledge consists of factors, justifications, case study and supporting publication of the specific factors.
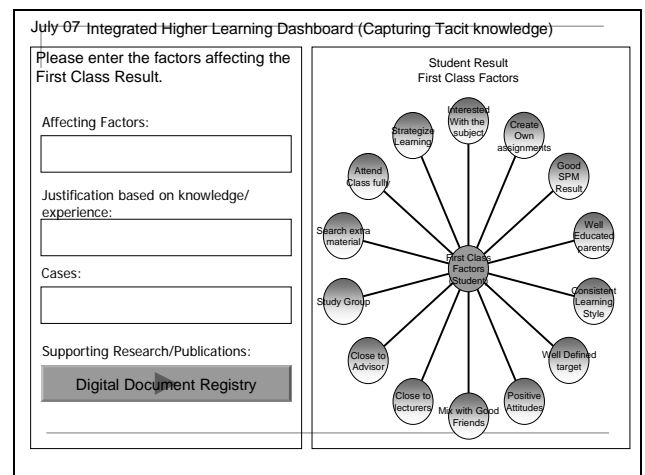


Figure 1.2 Screen to capture tacit knowledge on the factors

### 3.2.1 Visualizing the Clustering Result in Dashboard Form

Figure 1.3 demonstrates the result of clustering and forecasting engine, where the generated clusters are transformed into dashboard by applying the aggregation and roll-up algorithm. The

figure also illustrates the drilling down from dashboard to clusters, and from clusters to feature vectors.
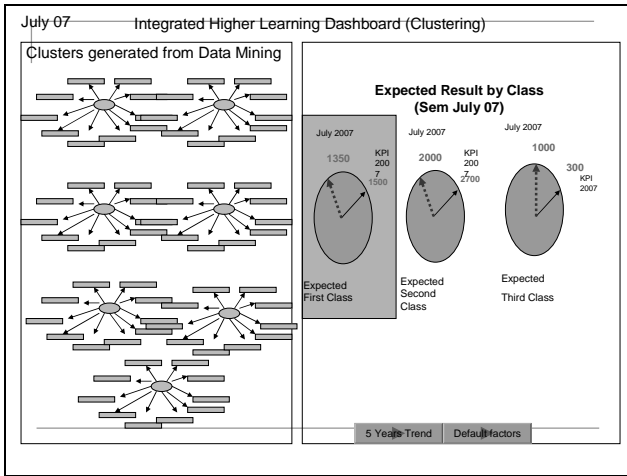


Figure 1.3 Screen to Visualize the Dashboard of Forecast Number of First Class Students with the associated clusters.

### 3.2.3 Zooming the Cluster's Behavior

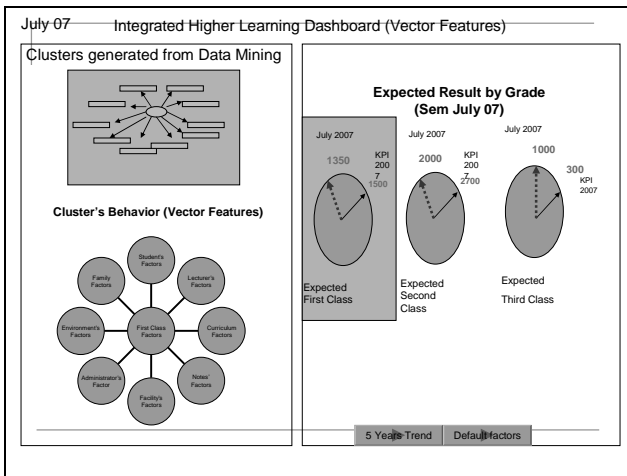Figure 1.4 shows the selected clusters and the associated factors.



Figure 1.4 Screen to visualize the factors for each cluster.

3.2.4 Drilling Down to Each Factor

Figure 1.5 illustrates the drilling down from each factor to the associated behavior of the factor. Figure 1.5 shows the example of student's behavior of first class students.
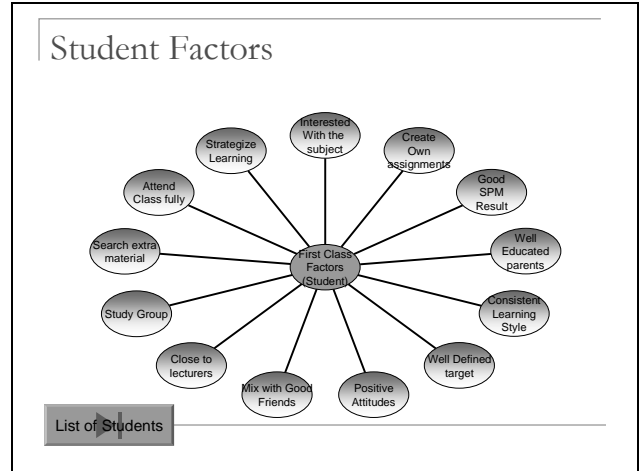


Figure 1.5 Screen to visualize the student's behavior\

Figure 1.6 then shows the list of students in the selected cluster with the behavior presented in the screen. The detail profile of the student can be reached by clicking the student name.
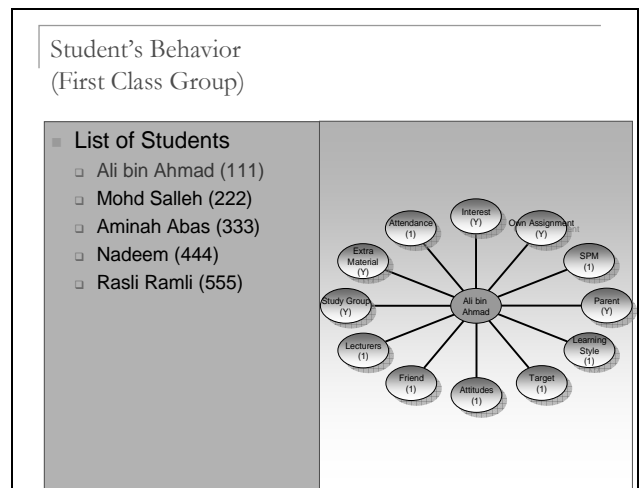


Figure 1.6 Screen to Visualize List of Students associated with the behavior.

3.2.5 Visualizing the proposed actions to be taken generated by Intelligent Engine

Figure 1.7 demonstrates the screen which zooms into each particular student with the proposed actions to be taken by the respective entity in order to improve the student's result. The proposed actions are derived by the forecasting engine based on the factors table.
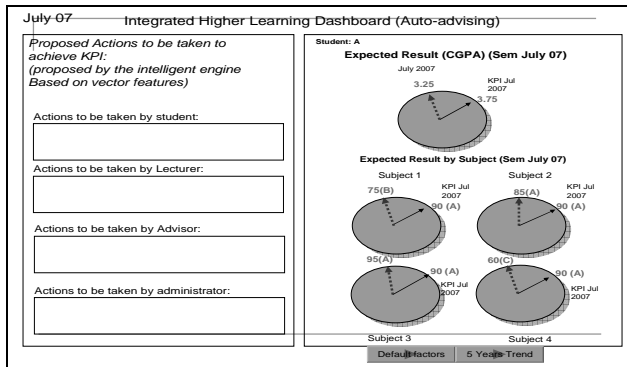
Figure 1.7 Screen to Visualize the Proposed Actions to be taken by each entity to achieve KPI

3.2.6 Capturing Tacit Knowledge in Post Data Clustering Process

Figure 1.8 represents the screen to capture tacit knowledge of education's expert, lecturers and students on the new factors and reasons that affect the student's result. The new factors shall be added up in factors table for future clustering and forecasting process.
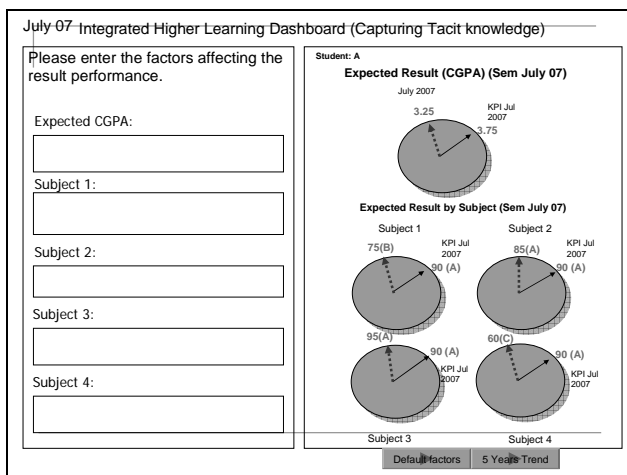


Figure 1.8 Screen to Capture Tacit Knowledge from Education's Expert, Lecturers and Students on the reason and new factors.

## 4. CONCLUDING REMARKS

In this paper, we addressed problems with the existing visualization system, in which the understanding of data plotted in current system is difficult to understand. We propose a holistic framework of data visualization system which shall improve the visual impact of the system. By designing the system to be multi-layered, it shall improve the visualization and personalization level. The framework of the system combines dashboard and graph-based concept to ensure that data is visualized in appropriate view for each group of users. A prototype of the graph-based dashboard system design is illustrated for higher education environment focusing on first class student.

## 5. FUTURE WORK

The research shall be continued to experiment the framework through the development of graph-based dashboard software and shall implemented in higher education environment.

## 6. REFERENCES

[1] Luan, J., "**Data Mining and Its Applications in Higher Education" in A. Serban and J. Luan (eds.) Knowledge Management: Building a Competitive Advantage for Higher Education**. New Directions for Institutional Research, No. 113. San Francisco, CA: Jossey Bass, 2002.

[2] Anand M., Bharath B.N., Chaitra, Kiran Kumar M.S., Vinay C., Mr.T.L.Bharatheesh, " **Visual Data Mining Application in Material** ",http://www.geocities.com/anand_palm/

[3] Sarabjot S. Anand, David A. Bell and John G. Hughes,"The Role of Domain Knowledge in Data Mining", **Conference on Information and Knowledge Management** (CIKM '95), **Proceedings of the fourth international conference on Information and knowledge management,** Baltimore MD USA, ACM Press, 1995, pp. 37-43.

[4] Mihael Ankerst, Christian Elsen, Martin Ester, Hans-Peter Kriegel, "Visual Classification: An Interactive Approach to Decision Tree Construction", **Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '99**, San Diego CA USA, ACM Press, 1999, pp. 392 – 396.

[5] Elias Pampalk, Werner Goebl, and Gerhard Widmer, "**Visualizing Changes in the Structure of Data for Exploratory Feature Selection",** SIGKDD '03, August 2003, Washington, DC, USA.

[6] Berchtold S., Jagadish H.V. and Ross K.A.:"Independence Diagrams: A Technique for Visual Data Mining", in Proc. **KDD '98, 4th Intl. Conf. on Knowledge Discovery and Data Mining**, New York City, 1998, pp. 139-143.

[7] Jianchao Han, Nick Cercone, RuleViz: "A Model for Visualizing Knowledge Discovery Process", ACM Press, **Conference on Knowledge Discovery in Data**, **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining,** Boston, MS, USA, Pages: 244 – 253.

[9] Emanuel G . Noik, "Challenges in Graph-Based Relational Data Visualization", **Proceedings of the 1992 conference of the Centre for Advanced Studies on Collaborative research - Volume 1 CASCON '92**, IBM Centre for Advanced Studies Conference , Toronto, Canada, 1992, Vol. 1, pp. 259-277

[10] J .A. Bondy and U .S .R. Murty, "**Graph Theorywith Applications**", North Holland, Amsterdam,1976 .

[11] R.E. Tarjan, "Data Structures and Network Algorithms", Number 44 in CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, 1983 .

[12] J .D. Ullman, "**Principles of Database and Knowledge-base Systems**", volume 1. Computer Science Press, 1803 Research Boulevard , Rockville, Maryland, 20850, 1988 .

[13] Manuel Freire, Pilar Rodr´ıguez, "A graphbased interface to complex hypermedia structure visualization", **Proceedings of the working conference on Advanced visual interfaces ,**ACM 2004, Gallipoli, Italy, 2004, pp. 163-166.

[14] P. Eades and K . Sugiyama . How to Draw a Directed Graph. **Journal of Information Processing,**13(4), 1990.

[15] J .D. Mackinlay, G.G. Roberson, and S .K.Card," **The Perspective Wall : Detail and Context Smoothly Integrated** ", In ACM CHI '91, ACM Press, New York , NY, April 1991, , pages 173-179.

[16] www.wikipedia.com

[17] Todd Abel, "Microsoft Office 2000: Create Dynamic Digital Dashboards Using Office, OLAP, and DHTML", MSDN Magazine, http://msdn.microsoft.com/msdnmag/issues/0700/Dashboard/, posted on 2006.

[18] Aaron Marcus. Dashboard in your Future. Interactions, January-February http://www.denniskennedy.com/blog/2006/06/is_there_a_digital_dashboard_in_your_future_a.html,  posted on June 22, 2006.

[19] Wan Maseri Wan Mohd, Jasni Mohd Zain, Abdullah Embong, "Parameterless K-means: Auto-generation of Initial Number of Cluster and Centroid based on Distance of Data Points", **Industrial Conference on Data Mining**, ICDM 2007 Symposium Poster.

[20] Istvan Jonyer, Diane J. Cook, Lawrence B. Holder, "Graph-Based Hierarchical Conceptual Clustering", **Journal of Machine Learning Research** 2, 2001, pp. 19-43.

[21] Daxin Jiang Jian Pei Aidong Zhang, **Towards Interactive Exploration of Gene Expression Patterns,** SIGKDD Explorations, Volume 5,Issue 2 – pp. 79.

[22] Mihael Ankerst, "**Report on the SIGKDD-2002 Panel The Perfect Data Mining Tool: Interactive or Automated?",** SIGKDD Explorations, Volume 4, Issue 2, pp. 110

[23] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM 39, **Communications of the ACM ,** Volume 39, Issue 11, November 1996, pp. 27-34.

[24] Data Sets, Proc. Visualization '96, Hot Topic Session, San Francisco, CA, 1996.

[25] Luminita Dumitriu, "Interactive Mining and Knowledge Reuse for the Closed-Itemset Incremental-Mining problem", **ACM SIGKDD Explorations Newsletter,** Volume 3 ,  Issue 2  (January 2002), 2002, pp. 28-36.

[26] Hillol Kargupta, et. al., MobiMine: Monitoring the Stock Market from a PDA, **ACM SIGKDD Explorations Newsletter,** Volume 3, Issue 2  (January 2002), 2002, pp. 37 – 46.

[27] Danyu Liu, Alan Sprague, Upender Manne, "JRV: An Interactive Tool for Data Mining Visualization "**ACM Southeast Regional Conference**, **Proceedings of the 42nd annual Southeast regional conference,** Huntsville, Alabama, 2004, pp. 442 – 447.

[28] Thomas W. Gonzalez , Dashboard Design: Key Performance Indicators & Metrics. http://www.xcelsius.com/index.html, posted on 2006.