Turkish Online Journal of Distance Education-TOJDE July 2010 ISSN 1302-6488 Volume: 11 Number: 3 Article 8

INFORMATION VISUALIZATION AND PROPOSING NEW INTERFACE FOR MOVIE RETRIEVAL SYSTEM (IMDB

Ronak ETEMADPOUR
Mona MASOOD
Centre for Instructional Technology and Multimedia,
Universiti Sains Malaysia, MALAYSIA

Bahari BELATON
School of Computer Sciences,
Universiti Sains Malaysia, MALAYSIA

ABSTRACT

This research studies the development of a new prototype of visualization in support of movie retrieval. The goal of information visualization is unveiling of large amounts of data or abstract data set using visual presentation. With this knowledge the main goal is to develop a 2D presentation of information on movies from the IMDB (Internet Movie Database) as our movie search engine. The aim is to amplify the perception of users over the retrieval environment while preserving the output quality at an acceptable meaningful level. Visualization ideas deal with visualization of information seeking results. These visualization techniques have been collected from the literature on how to map the results of the information retrieval process. We propose a 2D visual interface for mapping collections of movies and exploration of their related information to maximize density of needed information in a single page. For this purpose, we introduce a movie categorization scheme to help users in navigating through the movies information.

The traditional style (interface) of clicking a link to view details of documents in most of the search engines like the IMDB (Internet Movie Database) as our movie search engine to find information is very uninteresting and tedious. This is because when clicking a particular link the user's focus is shifted to the new page, and if the information presented is not to their interest, they will need to switch back to the movie's search results. The key contribution is thus a reasonable mapping result of a query on an actor/actress movie database displayed in just one page that can amplify visual perception of retrieved movies. This theory of data graphics interface focuses on maximization of the density of useful pertinent information with respect to users query in a screen page. This method of grouping of information needs some data extraction algorithms by parsing and crawling the IMDB web pages that are useful to retrieve important movies information. Our graphical-based visualization provides a correct understanding of information that users can view information without reading them. A qualitative experimental test comparing the classic (traditional) interface of the IMDB and visual interface was conducted.

Keywords: **Visualization, Retrieval system (IMDB), 2D visual interface, Contribution, Extraction algorithm, experimental test, IMDB Movies' Information Visualization**

INTRODUCTION

Information is one important need in our private life, thus World Wide Web has become a necessary important information source as part of our life. Technology has done a tremendous job of enabling the world to capture, store, and transfer huge amount of data. However, to help people make good decision, large amount of data is used to uncover important meanings in an accurate, clear, and useful way. Finding the interested data is one challenge for users, therefore "Visualization" leads us to first use graphical elements carefully in representing large amount of information and datasets; second, attempting to display the datasets graphically, in two or three dimensions, grouped by topic, categories or clusters.

This ability of the human mind to rapidly perceive visual information makes information visualization not only useful, but also powerful and a necessary tool for information discovery. Visualization tools are strongly related to the users and their tasks. For example in printed form, visualization has included the display of numerical data (bar charts, plot charts, pie charts) and geographic data (encoded maps). Users, Tasks, Data, and Basic visualization interactions are four important issues for visualization implementation that focus on understanding the user's needs.

As Tufte (1990) says the main goal of information visualization is to amplify a deeper level of understanding and insights into the underlying process. "The application of the visualization is a technique to information retrieval really broadens the horizon of information retrieval" (Zhang & Korfhage, 1999). In fact visualized data is much better, more creative and obviously an interesting way for presentation of datasets because data visualization is a modern approach and highly efficient for people to directly perceive data and discover knowledge and insight from it. Traditionally valuable information has been achieved through charts and graphs. The model that is shown in Figure 1 is the data transformation from the input in the form of raw data to data tables and then visual mapping to visual structure and the final interface. The whole processes and tasks are manipulated by human interaction (Mann, 2002).

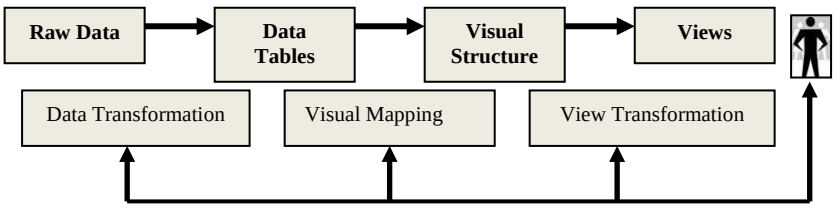

Figure: 1
Human Interaction

Traditional Retrieval Systems Problems

Large collections of documents are accessible to users via powerful search engines and retrieval algorithms; in fact Search engines are very useful because they allow the user to retrieve documents of interest from the World Wide Web. Since the computer is a powerful tool for searching, most conventional search engines like movie search engines actually return thousands of hit documents as their output and follow traditional interface of search engine that provide a linear list of results matched with query. Various searching mechanism and algorithms are becoming useful to allow the user to retrieve documents of their interest.

Search engines retrieve a ranked list of potentially relevant documents and usually result of retrieved information displays in text list of titles. The only way to browse the content on these sites is by choosing from a long linear list and scrolling through them to find useful information. This process is tiresome and tedious, simply because it is impossible to navigate such a long list. Based on this logic movie results of Movie search engines like Internet Movie Database (IMDB) are typically displayed in an alphabetic organization that is difficult to use. They employ a traditional approach to database queries with textual response in forming scrolling list. Due to limited screen space, the results are displayed in more than one page.

Objectives

Therefore, the main aim is to produce an interactive 2D visual interface in one page that reduces the needed screen space. It is a visual interface for visualizing the results of movie retrieval system that summarizes the results using grouping techniques, features that facilitate user navigation through displayed information. Using visual, attractive and understandable abstract shapes amplify users' perception and this theory of data graphics interface focuses on maximization of the density of interested information in just one screen page.

Related Works on Information Visualization

An enormous amount of work has been done in the field of Information Visualization in the last few years that ranges from geometric techniques (e.g., scatter plots, parallel coordinates), spiral- and axes-techniques to interactive visualizations for hierarchical information (e.g., cone), graph-based techniques (e.g., curved line) and icon-based techniques (e.g., Chernoff faces). Multivariate data visualization techniques are often limited in terms of data records to the limited size of the screen. One presentation technique is the tree map.

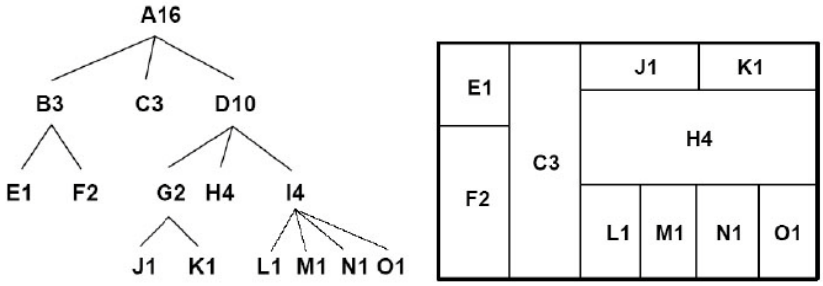

Figure: 2
Treemap display in a hierarchical data set

The tree map is a rectangle space-filling technique which separates rectangles into smaller rectangles repeatedly to represent a hierarchical structure. Each rectangle represents a node. Vertical and horizontal lines are used alternatively to separate the rectangles to small segments. Figure: 2 shows a sample of traditional tree view and its equivalent visualization technique in a tree map (Schneiderman, 1990).Simple shapes like faces are effective graphical representation to reduce screen space when presenting huge complex datasets. Colors can be used to represent strategy through mapping; the more similar the strategies, the closer the colors. For example, each segment (rectangle) in a tree map can be used to represent different score. The presentation of search engine result and representations of abstract data can be improved in order to facilitate cognition of users specifically to achieve pertinent information by using interactive graphics to reduce time in finding the corresponding data.

Previous Works on Visualization of Retrieved Results

Finding relevant information becomes increasingly difficult by the growth of the web. In 1997, Zimmer noted in a German newspaper that "the chance to find certain information decreases drastically with the increase of information possibilities". His conclusion was that information overload is the key word; information rejection is already the necessary action. The availability of large collections of documents coupled with powerful search and retrieval algorithms provides the opportunity for people to access large sets of relevant documents in electronic form. Visual information retrieval technique has two distinct design genres: the cognitive strategies of the human mind to solve problems and observed interaction patterns with existing information retrieval systems.

There are many challenges to visualize information including choosing between 2D and 3D interfaces, navigation and interaction methods, and selecting an appropriate level of detail. Animation is used more and more in information visualization systems to help users keep their orientation when transformations or changes of mappings occur. Cat-a-Cone (Hearst & Karadi, 1997) also uses 3D animation for presentation of hierarchical categories.

The second technologies that we discuss are 2D visualization using simple HTML with JavaScript supports. This has recently gained popularity primarily due to certain applications mapping search engine results. Some Visual Search assistant like oSkope with a highly intuitive visual interface is producing interactive 2D visual interface that browses quickly through a large number of images to let us preview information with minimal paging regarding reduced needed screen space.

Accordingly, in our study, the list of retrieved movies from IMDB can be represented visually by using images or shapes in order to avoid long textual list of movies. 2D presentation is handled using a display graphic data structure from a shape. However 3D-ideas have an important role in the area of documents visualization; but they have no potential components in most systems because the technical environment of the target users is standard PCs and input devices.

PROPOSED METHOD

The issue here is to present data on limited screen and to amplify user cognition of retrieved movie results. In general though, a 2D mapping interface for search engine results handles the comprehension aspect (producing variant category by using different colors and abstract shapes).

Thus, construction of a graphical display will become important in representing the retrieval interest of texts or documents that consist of movies/videos. IMDb (http://www.imdb.com) is a popular site cataloging almost every movie ever made.

Most people know about movies and can relate to movies and actors that are presented with a visualization of the movie data that makes them eager to find their favorite movies and actors, check movies and explore actor/actress of their interest. That is why the study of IMDB as a movies engine is interesting.

The dataset has rich information on each movie and actor scope. Figure 3 illustrates the position of our proposed methods in the search engine visualization based on 2D interface.

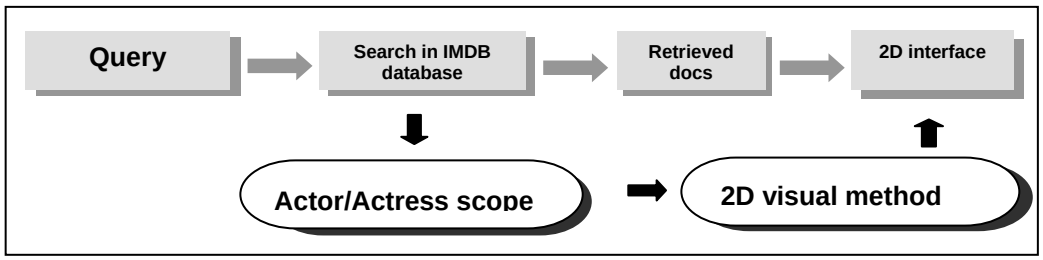

Figure: 3
Architecture of proposed method

In this study, we are concentrating on the visualization of movie search results. The specific features are Data Presentation and Data Exploration that focus on the maximization of the density of useful information in a single page; these are two fascinating aspects in user comprehension. Information Visualization can be used as the solution to present information by providing an interactive environment for people to discover information about movies.

Problems in existing interface of IMDB that have been identified as follows:

- **The searched results of IMDB are represented in the typical linear fashion.**
- **The standard style of click the link to view details of particular movie is practiced here while it is useful for normal text based web navigation. Because by clicking a particular link the user focus is shifted to the new page, and if the information presented are not to their interest, need to switch back to the first page.**
- **The first page doesn't show movies information such as genres, pictures, and actors of each movie.**
- **Principles of our visual interface are organized as follows:**
- **Phase 1 introduces the data preparation that consists of a list of over 20,000 Actors/Actresses. We extracted 26000 names of famous actors and actresses from the IMDB using a surface crawling procedure from www.IMDB.com. Surface crawling method enable searching of 10,000 movies in the IMDB web site.**
- **Phase 2 explains the data analyses and results and this trend is further complemented by extracting some summarized information from the IMDB, mainly to speed up our visualization progress. In fact we reduced and limit our data scope to a list of all actor/actress names, movies of per actor/actress, year that actor/actress play in a movie, genres of movies, and any documentation related to each movie.**

Phase 3 discusses the design of the visualization. The aim is to provide 2D visualization interface with one search box to retrieve the name of actor/actress of interest from the IMDB. Thus these retrieved texts will be presented using special 2D interface with different colors and different shapes or icons. A graphical model communicates with users through iconic presentation of documents that users have more control over the information retrieval process. Each document offers a 2D view of the dataset to yield information for user decision-making with regard to item selection. The entire document set is presented to make an environment to accommodate users' needs in an interesting way. The most functional rules for information retrieval visualization include proximity, closure, and continuity.

These principles provide a framework of visualization-based information retrieval system interfaces from the users view.

- **Proximity refers to perception of people on the grouping of elements; accordingly it is one of the best principles in interface design area.**
- **The law of closure is clear when the brain perceives boundaries of objects or when interface has contours that separate spaces.**
- **Text may be minimally used. Instead icons and abstract shapes are used while it is essential to make sense of the display and interpret the meaning of the icons or any shapes that is used (symbols to represent concepts).**
- **Color is used to sort, group or categorize such as year and name of director are used as popular design features (here, colours and textures highlight or differentiate elements).**

Graphical presentation of results are easy on the eye, colors are powerful element in visualization and have greatest psychological impact on visitors, and at the same time the computer screen supports RGB color model. Furthermore, movie genres are sorted and color coded by genre that give them a distinct color each as shown in Table: 1.

Table: 1
Sorting genres based on colors

List of Genres		List of Genres	
Action	Light Blue	Animation	Dark orange
Comedy	Orange	Horror	Red
Family	Green	Crime	Black
Thriller	Yellow	History	Blue
Short	Violet	War	Grey
Romance	Pink	Documentary	Light green

Genres are represented in a circle that is divided into 12 slices; each slice is dedicated for one of the genres in Table 1. As Figure 4 demonstrates since each movie may have more than one genre, using the slice technique for one genre is specified.

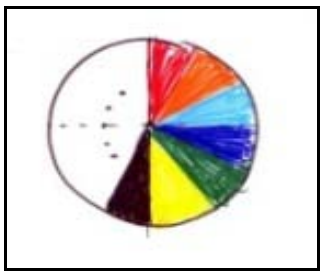

Figure: 4
Movie is divided into 12 slices

Circles are arranged in a year block and a star symbol is chosen to represent the actor/actress who plays the main role in the specific movie (Fig: 5).

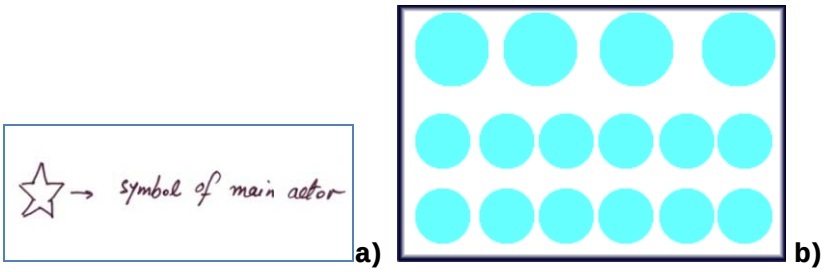

Figure: 5
a) Star Symbol, b) Revolution size of circles

An information triangle card provides more information about attributes such as actors/actresses, documents, pictures is displayed in order to give tips to users in an effort to obtain an overview of a text or pictures that is directly tied to a normal reading view. As figure 6 displays the symbol of each segment in the triangle visualization represents the movie information.

Figure: 6
Triangle card Information

As Figure 5 shows for handling of large volumes of data we determined our threshold 3.16 close to 4 based on the average number of movies that an actor may play a role in a year.

Movies more than 4 can be mapped in a year square with changing the size of circles to handle large volumes of data in a limited space of blocks that guarantees all relevant movies with the same year located in a square block. Size of the small diagonal is 2/3 of current diagonal.

D1 = Size of current circle diagonal
D2 = Size of small circle diagonal
D2 = 2/3 * D1

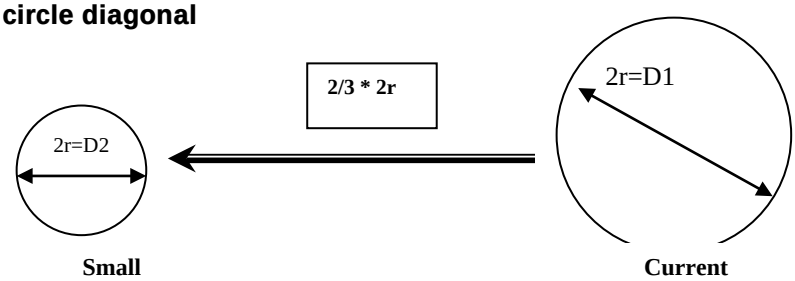

Figure: 6
Revolution sizes of circles

The final 2D proposed visual user interface is demonstrated in Figure 7.

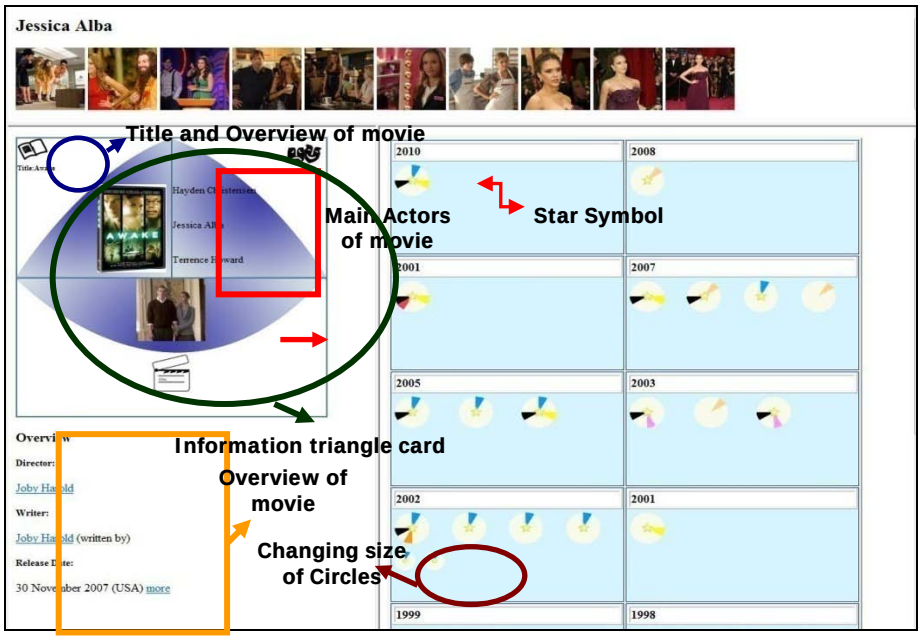

Figure: 7
Final 2D Proposed Visual Interface

EXPERIMENTS, RESULTS AND PERFORMANCE ANALYSIS

A qualitative comparative survey between existing the IMDB interface and proposed visual interface was developed. Twelve participants were given 5 minutes of training session individually because each user has unique skills and experience. All participants perform the same two tasks during a 20-30 minutes session. Each session includes a search on traditional interface of the IMDB and another with our proposed visual interface. T-test Statistical method is applied to show significant differences in mean values. Our evaluation setting had the goal of comparing visualization and non-visualization condition. Two groups of people, one group with computer science knowledge (6 participants: 3 males, 3 females) another without computer science knowledge (6 participants: 3 males, 3 females) were chosen to complete both parts ("viz" & "noviz"). All participants had no experience related to information visualization techniques.

As illustrated in Figure 8, visualization has an increased trend in "wonderful" and "easy complete task". Overall the graph shows fluctuations regarding the user satisfaction parameters in the visual interface. The Graph has a dip in "easy" parameter; again indicating that traditional interface with the difference rate of 0.1509 is much easier than visualization environment. Zero T-value in the logical item shows both methods of presentation have an equal trend and has the same logic. This is good enough for our purpose since it shows the qualitative growth between the two interfaces. The visual results of movies is facilitated by different parameters such as navigation using grouping, satisfaction and high density of interested information but do not say about the user's ability to work with structures. Therefore, it is not guaranteed as an easy interface. A correct understanding of search results is a successful visualization technique that our graphical visualization provides useful hints to solve the information overloading.

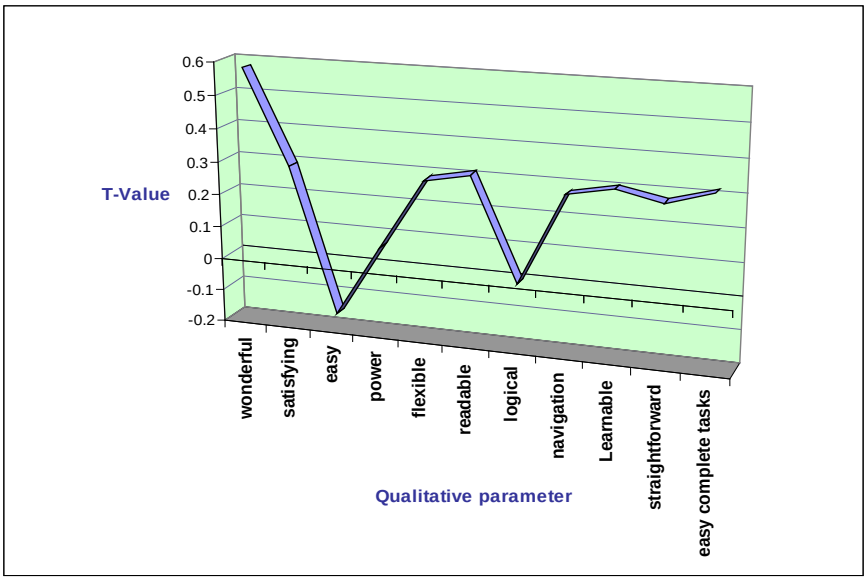

Figure: 8
T-value Graph

As discussed before our proposed method of visualization is a pre-developing stage that aims to amplify cognition of users in a touchable way in work via some search engines like IMDB while preserving all retrieved information from the huge IMDB database. The design of an effective information visualization system is still more an art than a science.

We have used information visualization techniques as an aid to the understanding of the IMDB environment as a model of a search engine. In order to perform the task successfully, we can suggest a set of structural description rules and legend in visual working. In order to generalize the results to other information retrieval applications, the performance of all factors were evaluated with qualitative measures. In order to generalize the results to other information retrieval applications the performance of all factors were evaluated with qualitative measures. We convert textual information into graphical representation that can be processed visually rather than reading. Therefore, user can see information without reading the information. But as evaluations shows most of the users were unfamiliar with this kind of result (Movie) representation.

This research implicates to improve the visualization of items interactivity. We need more extensive usage studies and evaluations to investigate and observe user interaction to determine a suitable implementation enhancing better support for data exploration.

BIODATA and CONTACT ADDRESSES of AUTHORS

Ronak ETEMADPOUR, she is a Ph. D. student in Instructional Systems Technology and Multimedia from Universiti Sains Malaysia (2009), an M.S.C Computer Science from Universiti Sains Malaysia (2008), and B.S. in Computer software engineering from Central Tehran Azad university of Iran (2003). She is currently a PHD fellowship student at Universiti Sains Malaysia and researches on Information visualization and visual literacy.

She had 1 year work experience as web designer (2005-2006) and she was a person in charge of IT in Blue Gulf Shipping lines Company (2003-2005). Her interests are the use of visualization in Education and psychological aspects of visual interfaces. She has participated in 3rd international computer and Instructional technology Conference.

Ronak ETEMADPOUR
Centre for Instructional Technology and Multimedia,
Universiti Sains Malaysia, MALAYSIA
Email: r_etemadpour@yahoo.com

Mona MASOOD, she has a Ph. D. in Instructional Systems Technology from Indiana University (2004), an M.S.Ed from Arizona State (1987), and B.S. in Mathematic and Economics from Oklahoma State (1986). She is currently a Senior Lecturer at Universiti Sains Malaysia and teaches both graduate and undergraduate levels including Visual Communication and Human Computer Interaction. Her interests are in the use of ICT in Education, e-learning, instructional systems technology, trends in educational technology, and multimedia design and production. She has published 40 papers in journals and proceedings in the field of instructional technology and presented at both national and international conferences, seminars and workshops. Presently she is actively involved in several research grants that look into issues and standards of ICT and Instructional Technology. Association: Member of the South East Asian Association for Institutional Research (SEAAIR) and Member of the Malaysian Educational Technology Association.

Mona MASOOD
Centre for Instructional Technology and Multimedia,
Universiti Sains Malaysia, MALAYSIA
Email: msmona@usm.my

Bahari BELATON, he is Assoc. Prof. Dr in Computer science school of Universiti Sains Malaysia, He has BAppSc.(Comp.Studies) from South Australian Inst. of Technologu, Australia, B.Sc.(Hons) Flinders and PhD. Leeds, U.K. his research interests are Scientific Data Visualization, Computer Graphics & Network Security. Currently he is Acting Research Dean and Information & Communications Technology Research Platform and he is Head of Visualization And Graphics Group.

Bahari BELATON
School of Computer Sciences,
Universiti Sains Malaysia, MALAYSIA
Email: bahari@cs.usm.my

REFERENCES

Hearst, M. A. & Karadi, C. (1997). Cat-a-cone: an interactive interface for specifying searches and viewing retrieval results using a large category. *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval.* **Philadelphia, Pennsylvania, United States, ACM, 246 - 255.**

Shneiderman, B. (1990). Tree Visualization with tree-maps: A 2D space filling approach. *ACM Trans. Graphics,* **11, 92–99.**

Tufte, E. R. (1983). *The visual display of quantitative information*. **Cheshire, CT: Graphics Press.**

Tufte, E. R. (1990). *Envisioning Information*. **Cheshire, CT: Graphics Press.**

Tufte, E. R. (1997). *Visual explanation*. **Cheshire, CT: Graphics Press.**

Zhang, J. & Korfhage, R. (1999). Dare, distance and angle retrieval environment: A tale of the two measures. *Journal of the American Society for Information Science, 50*, **779–787.**