STRUCTURAL COMPONENTS OF STUDENTS' DATA-FOCUSED INFORMATION VISUALIZATION

Semiu A. Akanmu[1], Zulikha Jamaludin[2]
[1,2]*Universiti Utara Malaysia, Malaysia,*
[1]*ayobami.sm@gmail.com,* [2]*zulie@uum.edu.my*

ABSTRACT. Information Visualization (InfoVis), as an analytics and visualization tool, had been argued to be befitting in attending to the experience of information overload, and subsequent decision making constraint of higher education institutions' (HEIs) decision makers. This experience is said to be as a result of increase in volume of students' data and the limitations of the available data management tools. However, due to the diversity of domains in which application of InfoVis are demanded, designing domain-specific structural components of the intending InfoVis is compulsory, so as to address its peculiar domain problem. Adapting the generic Design Research method, we employed critical review of documentations of selected InfoVis tools and mapped the findings with the outcome of our previous investigation of the HEIs' decision makers' explicit knowledge preferences. This work therefore highlights the structural components of the HEIs' students' data-focused InfoVis.

Keywords: higher education institutions, students' data, information visualization, structural components

INTRODUCTION

Information Visualization (InfoVis) is a research domain on data analysis and knowledge discovery through visual exploration for the purpose of supporting decision making activities (Sheneiderman & Plaisant, 2010). However, for InfoVis to be functional, its structural components must address the peculiar exploratory data analytics of the applied domain (Koh, Slingsby, Dykes & Kam, 2011). In this paper, the structural components of higher education institutions' (HEI) students' data-focused InfoVis (SDI) is proposed. It highlights visual interactive interface, exploratory data analysis program, and database (table) as its basic structural components, and presents the details of its form. The three-tier components –interface, EDA program, and database of any type –are found to be generic, while the details of each of these components are domain-specific. These details are determined by the explicit knowledge expected to be visualized by the InfoVis. The proposed structural components are to serve as design guide for visualization designers and developers in designing SDI, or its related domain-specific InfoVis.

PAST RELATED STUDIES

Reviewing some of the past studies on InfoVis research: Meyer (2012) and Simon et al. (2011) works centered on visualizing biological data. Wang et al. (2012) worked on designing and implementing interactive InfoVis platform for product lifecycle management (PLM), with aircraft product development as its focus. Stoffel et al. (2012) presented a web-based visual query analysis system to attend to the inadequacy in the analysis and simulation of micro grid energy mix, while Harrison et al. (2010) unveiled the use of InfoVis software in visualizing chemical structure. In each of these studies, and others, the structural components, though generic in terms of the three tiers, the details are determined by the nature of the explicit knowledge to be visualized and the data type. This informs the essence of conducting a domain-specific research to determine the structural components that will serve as guide for the visualization designers.

METHODS

This study adapts the generic design research method (DRM) (Teegavarapu & Summers, 2008) in determining the structural components of SDI. The suitability of DRM is borne out of its primary capability of creating innovative technological product which SDI is undoubtedly one. The phases of the research method are outlined below (a –d).

a. **Critical review of InfoVis tools' documentation:** A review to highlight the structural components of the selected InfoVis tools in view of forming a generic template.
b. **Meta-Data Analysis:** To identify the implications of the earlier investigated HEIs' decision makers' explicit knowledge preferences (published in Semiu & Zulikha, 2014) on the InfoVis structural components of the domain.
c. **Mapping the findings in (a & b):** To align the findings in (1) and (2) as a basis of choosing the structural components of SDI.
d. **Proposing the Structural Components:** To present the chosen structural components with a diagrammatic representation.

A purposively selected fourteen (14) InfoVis tools are reviewed to highlight the general overview of InfoVis structural components and its determining factors. The selection is purely made on the basis of the accesibility of the InfoVis documentations. The details of these tools are presented in Table 1.

Table 1. The Details of the InfoVis Tools Analyzed for the Structural Components of HSDI

Code Tag	Name	Description	Uniform Resource Locator (URL)/ Reference
1	Spotfire	It is commercial business visual analytic software. It has versions for desktop, web-based (cloud), and mobile.	http://spotfire.tibco.com/demos, accessed on 04/02/2015.
2	VisDB	A pixel-based visualization system for multidimensional datasets	Keim, D. & Kriegel, H-P (1999), in Card, Mackinlay, & Shneiderman (Eds).
3	InfoCrystal™	A visualization tool for visual query language developed by Center for Educational Computing Initiatives, MIT.	Spoerri, A. (1999), in Card, Mackinlay, & Shneiderman (Eds).
4	SeeSys™	A system that used statistical techniques to visualize code inform of hierarchy to divide the entire program code into subsystems	Baker, (1999), in Card, Mackinlay, & Shneiderman (Eds).
5	TennisViewer	A visualization system to show the dynamics of Tennis match	Jin & Banks (1999), in Card, Mackinlay, & Shneiderman (1999), (Eds)

6	SeeSoft	A system used to visualize about 400 line of program code	Eick (1998).
7	TextArc	A textual visualization tool for a book called *Alice in Wonderland*	www.textarc.org, accessed on 04/02/2015.
8	PresidentMap	This is a visualization tool used by New York Times to present the result of the election	http://elections.nytimes.com, accessed on 04/02/2015.
9	ThemeView	A 3D map that is used in the representation of results of a search of a large document corpus	Wise, et al. (1995).
10	Tableau	This is a commercial visual analytics tool. It can be used interactively to construct displays by dragging the variable name into the elements of the display	http://www.tableausoftware.com/, accessed on 04/02/2015.
11	TableLens	It provides a spreadsheet-like view of table data.	http://www.businessobjects.com/, accessed on 04/02/2015.
12	Baby NameVoyager	A visualization that allows users to type in a name and see the graph of its popularity over the past century	http://www.babynamewizard.com/voyager, accessed on 04/02/2015.
13	Gapminder	A visual platform for world's social, economic and political data.	http://www.gapminder.org/, accessed on 04/02/2015.
14	LifeLine	A visualization tool for the presentation of the health record of the patients	http://www.cs.umd.edu/hci/lifelines/, accessed on 04/02/2015.

STRUCTURAL COMPONENTS OF SDI

From the critical review of the documentations of the selected InfoVis tools (presented in Table 1), InfoVis is found to be of a three-tier structure. The first tier, which is the graphical user interface (front end), can be called visual exploration interface because it serves as the medium that users employ to perform the exploratory data analysis (EDA). The second tier (mid-tier) is a program code where the implementation of EDA model, mapping information to visuals, and data binding using varieties of appropriate programming languages is done. The third tier is the database (back end) which can either be for structured, semi-structured or non-structured data, and populated with the problem-focused data type. Table 2 presents the details of the structural components of the InfoVis tools analyzed.

Table 2. The details of the structural components of the InfoVis tools analyzed

Tier	Components	Description	Freq. (in %)
Visual Exploration Interface	Data layer panel	This is the space within the GUI where the information about the InfoVis and the data attributes to be visualized are held	79
	Visualization span	This is space within the GUI that holds the statistical graphics (e.g. Line chart, Radial graph, etc.) used in conveying the explored data analytics to the user	100

	Interaction Mechanism	This is the collection of interaction events handling programs executed at the browser. Examples are Pliant response hinting, dynamic hinting, etc.	92
	Distortion tools	These are tools used in rearranging, enlarging, and reducing the information displayed (through the statistical graphics) on the interface. Examples are Zooming, Scrolling, etc.	100
Visual-Exploratory Data Analysis System	Implementation of Association rules	Association rule shows that the occurrence of some items signifies the occurrence of the associated items with a certain probability, defined as confidence. If the set of items is represented by $I = \{i_1 \ldots i_n\}$, D is the set of transaction where T represents a set of items, such that: $T \subseteq I$. An association rule implies $Y \Rightarrow Z$, where $Y \in I, Z \in I$; $Y, Z \neq \emptyset$.	86
	Implementation of Classification rules	Classification technique is used for classification analysis. This is the process employed to develop a classification model from the data set using known class names. Classification is the task of learning a target function f that maps each attribute x to a predefined class; y.	86
	Implementation of Clustering rules	Clustering technique is used for finding partitioning of the dataset into subsets - *clusters*. Cluster analysis is used for dividing data to meaningful and useful groups – of objects that share common characteristics.	71
	Colour Coding	This characterizes the explored data analytics using colour scheme, i.e. different colour to distinguish the information displayed.	100
Database	Hierarchies	This is the data that highlights a hierarchical organization of the data sets. This is best represented using the tree diagram	50
	Temporal	This is data about time, highlighting the valid and transactional time.	14
	Discrete	This is the data type that characterizes data based on count (ordinal) or label (categorical)	64
	Networks	This is the data type that provides about nodes in a internetworked computers, IPV4, IPV6, and Mac addresses.	7
	World Wide Web	This is information on the web which are often in tags and markup languages	21
	Spatial	This is the data that support Euclidean geometry. It is used in geographical value representation.	14

The frequency, weighted in percentage (%), defines the occurrence of the components as shown in the reviewed tools. This serves as the basis of justifying the applicability rate of the highly ranked components i.e. components with 75% and above weighted frequency, except in the case of the database. The highly ranked components are Data layer panel, Visualization span, Interaction Mechanism, Distortion tools, Implementation of Association rules, Implementation of Classification rules, Implementation of Clustering rules, and Colour Coding are proposed to be parts of the structural components of SDI. The component of the database is exclusively determined by the nature of the data that the domain is dealing with, not by its assumed applicability. So, it is exempted from the 75% and above rule that guided the choice of other components. The 75% is the calculated median –a measure of central tendency – of the recorded weighted frequencies. Notably, the choice of the implementation of association, classification and clustering rules as the components of the visual exploratory data analysis system of SDI is further supported by the earlier identified explicit knowledge preferences of the HEI decision makers (published in Semiu & Zulikha, 2014). The database component which is to be populated by the befitting dataset is exclusively chosen based on the data type to be visualized. It does not depend on any applicability rate of the component. In SDI case, a database housing data tables(s) is proposed because of the structured nature of the dataset and its multidimensionality. The proposed structural components of SDI are illustrated by Figure 1 below

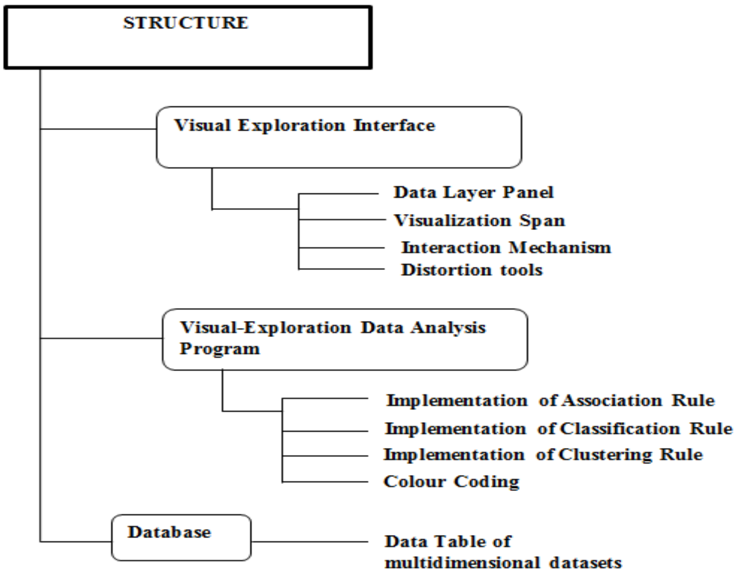

Figure 1. Proposed Structural Components of SDI

Notably, the primary function of SDI is exploratory data analysis. This is done by attending to exploration of the domain datasets through the collaborative roles of its structural components. The choice of these components, as earlier highlighted, depends on their suitability. The proposed structural components are generally applicable in InfoVis design of students' data, having considered a broadened scope and generic educational institution decision making processes. Our future work is basically the validation of the proposed framework which will be done through prototyping, as a proof of concept. In furtherance, heuristic evaluation of the prototype will also be conducted.

CONCLUSION

In view of designing an InfoVis which will specifically address the HEI policy makers' constraints in handling the students' data, this study proposes SDI structural components, using an adapted DRM. The structural components are outlined to be of three main tiers, namely; visual exploration interface, visual exploratory data analysis system and database. These three main tiers have their respective detailed components that specifically address InfoVis importance and HEI specificity. These structural components are to serve as design guide for visualization designers and developers in designing SDI or any related domain-specific InfoVis.

REFERENCES

Eick, S. (1998). Maintenance of large systems. In Stasko, J., *Software Visualization: Programming as a Multimedia Experience*, MIT Press, Cambridge MA, 315-328

Jin & Banks (1999), in Card, Mackinlay, & Shneiderman (1999), (Eds). *TennisViewer: A Browser for Competition Trees*, pp. 183 – 185

Keim, D. & Kriegel, H-P (1999), in Card, Mackinlay, & Shneiderman (Eds). *VisDB: Database Exploration Using Multidimensional Visualization*, pp. 126 – 135

Koh, L.C., Slingsby, A., Dykes, J., & Kam, T.S. (2011). Developing and Applying a User-centred Model for the Design and Implementation of Information Visualization Tools, *Proceedings of 15th IEEE International Conference on Information Visualization*, IEEE Computer Society.

Lam, H., Bertini, E., Isenberg, P., Plaisant, C. & Carpendale, S. (2012). Empirical Studies in Information Visualization: Seven Scenario, *IEEE Transaction on Visualization and Computer Graphics*, 18(9), 1520-1536

Meyer, M. (2012). Designing Visualization for Biological Data, *A Keynote talk at 3rd Leonardo Satellite Symposium of Arts, Humanities and Complex Networks (*NetSci, 2012)

Rogers, Y., Sharp, H. & Preece, J. (2011). *Interaction Design: Beyond Human Computer Interaction (3rd Ed)*, United Kingdom. John Wiley & Sons, Limited

Semiu, A., & Zulikha J. (2014). Students' Data-driven Decision Making in HEI: The Explicit Knowledge Involved. *Proceedings of 5th International Conference on Future Information Technology*, Bangkok, Thailand. Oct 10 – 12, 2014

Sheneiderman, B. & Plaisant, C. (2010). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 5th Edition, U.S.A. Pearson Higher Education.

Simon, S., Oelke, D., Landstorfer, R., Neuhaus, K. & Keim, D. (2011). Visual Analysis of Next-Generation Sequencing Data to Detect Overlapping Genes in Bacterial Genomes, *Proceedings of IEEE Symposium on Biological Data Visualization*, 47- 54. October 23-24, Providence Island, USA.

Spence, R. (2007). *Information Visualization: Design for Interaction*, 2nd Edition, England. Pearson Education Limited.

Spoerri, A. (1999), in Card, Mackinlay, & Shneiderman (Eds). *InfoCrystal: A Visual tool for information retrieval*, 140 – 135.

Teegavarapu, S., & Summers, J.D. (2008). Case Study Method for Design Research, *Proceedings of IDETC/DTM 2008, ASME 2008 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, 1-9.

Wise, J. A., Thomas, J., Pennock, K., Lantrip, D., Pottier, M. Schur, A., & Crow, V. (1995). Visualizing the non-visual: Spatial analysis interaction with information from text documents, *Proc. IEEE Symposium on Information Visualization*, 51 – 58.