MetaVis: Metadata Visualization using JUNG'S Library

Nurul Amelina Nasharuddin[2], Jamaliah Abdul Hamid[1], Mohd. Hasan Selamat[2], Hamidah Ibrahim[2], Rusli Abdullah[2], Muhamad Taufik Abdullah[2] and Wan Malini Wan Isa[2]

[1]*Faculty of Educational Studies*
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
Tel: 03-89468177, Fax: 03-89468246
E-mail: aliah@putra.edu.my

[2]*Faculty of Computer Science and Information Technology*
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
Tel: 03-89466555, Fax: 03-89466576
E-mail: nurulamelina@gmail.com, hasan@fsktm.upm.edu.my, hamidah@fsktm.upm.edu.my,
rusli@fsktm.upm.edu.my, taufik@fsktm.upm.edu.my, wanmalini84@gmail.com

ABSTRACT

In this paper we present a prototype for a new interface called MetaVis for search engines through graph drawing representation of interconnected metadata. This interface represents the metadata as connected nodes in a graph. The result will be displayed in a network with the keyword as the center and interrelated metadata as its connected nodes. Users will then be able to select any of the interrelated metadata nodes to navigate further all details connecting titles, sources, emails and addresses to the author. This fluidity of navigating related information from the interconnected nodes overcomes the rigidity of single layer search which is often time consuming. This interface which implemented Java Universal Network/Graph Framework (JUNG) has a few functions that enable users to manipulate the concepts graph.

Keywords

Metadata visualization, interface, visualization, Java Universal Network/Graph

1.0 INTRODUCTION

Researchers and students especially in academic world depend greatly on journals and articles' databases to search for information. Search engines dedicated to these databases often experience overload caused by the wealth of information. Generally the search engines for those academic papers will take the users' queries, search from the databases and produce a listing of results. But, in this paper we argue that search displays of articles with respect to authors, co-authors, related papers, research areas and publication metadata can be made visually more appealing, and more efficient, rather than through a listing.

The results of current searches are presented usually in a single layer presentation of information comprising information which pertains only to the selected search. Today however, with a proliferation of writings by authors and co-authors in multiple publications, there need to be a quick online search result view that displays connected networks of metadata linking numerous titles to authors, co-authors, their emails and publication details.

This current project described in this paper, called the MetaVis, develops a prototype for a new interface for search engines through graph drawing representation of interconnected publication metadata. The interface enables the navigation of layers of details connecting multiple titles, sources, emails and addresses to the author through the network. In this way, MetaVis aims to provide a more efficient visual interface of a search engine for researchers.

In this interface the user first begin a search the usual way by entering a keyword. The result will be displayed in a network with the keyword as the center and interrelated publication metadata as its connected nodes. Users will then be able to select any of the interrelated metadata nodes to navigate further all details connecting titles, publication sources, emails and addresses to the author. This fluidity of navigating related author and publication information from the interconnected nodes overcomes the rigidity of single layer search which is often time consuming.

This interface will implement the Java Universal Network/Graph (JUNG) library (O'Madadhain, Fisher, White & Boey, 2003) to visualize the whole model in a graph and the metadata in vertices or nodes. This project will be tested on 20 documents where the metadata of each document will be stored in a database. To that end, MetaVis expects the following data as input to the database:

1. List of papers (title, year, publication venue and authors).
2. List of publication journals (name, volume, issue and pages).
3. List of authors (name, address and email address).

The interface can be used by students, researchers and anyone who wants to search a journal or article and other related articles written by the same author, or published by the same publishing company in a specific number of years.

It should be very user-friendly for the user regardless of their computer literacy level.

2.0 RELATED WORK

Visualization is the process of transforming data, information and knowledge into visual form making use of humans' natural visual capabilities (Gershon, Eick & Card 1998). An effective visual interface enables users to observe, search, navigate, filter, discover, understand and interact with small and large volumes of data rapidly and users can reveal hidden characteristics and patterns of the data. Visualization somehow changes the way we present and understand complex data sets from the conventional way. The impact of visualization leads to new insights and more efficient decision making.

Information visualization is better known as visual representations of the semantics or meaning. Visualization of information using visual forms has been researched for many years. It combines aspects of scientific visualization, human-computer interfaces, data mining, imaging and graphics (Robertson, Card & Mackinlay 1993). On the other hand, scientific visualization focuses more on physical data and is generally served for specialized and highly technical scientists (McCormick, DeFanti & Brown 1987). Information visualization focuses more on abstract information that can come in huge quantities and be broadly deployed among a diverse and potentially nontechnical community. By using information visualization, users must be able to get the information they need, to make sense of it, and reach the decisions in a relatively short time.

One of the topics under Information visualization is the visualization of publication-related metadata. This type of metadata visualization presents visual representations of the metadata in a journal or article comprising the title, author, email and source of publication. In a usual search, students and researchers will typically begin by using the digital library or the Internet's search engines to search for articles or journals as references. Some of the online digital libraries are ACM Portal and IEEE Xplore. The articles are usually searched by titles, authors, publication years and sometimes the journals' name. The result pages of the search engines basically contain only the listings of the related journals that the students have been searching (see Figure 1).

In order to view the details of any article, the users need to open each of the articles' description pages. Only then will the details such as the authors and co-authors related to the article, the full citation of the journal, and the publication year will be available. This is a tedious process, especially when one wishes to compile bibliographies or to do a quick scan of authors doing related work. Moreover, other than the above list of information, other useful details such as information about an author and co-author (s), and other related work which they might have published either singularly or collectively in other journals are not available.

Figure 1: The result page of the IEEE Xplore search digital library.

There is a need to develop a new interface that can be implemented to the search engines of online digital libraries. The new interface must be able to overcome the difficulties in finding information about author- and co-authorship and navigation of details pertaining to related work they might have published in other publications. MetaVis interface focus on simplicity and efficiency when displaying journals publishing related work of authors. Users are thus able to search not only for related articles, but also for the authors who have published articles across various journal sources. Rather than having the search results in a listing, the results will be displayed in a graph form. An efficient visual interface like this enables users to interact with a large volume of data rapidly.

2.1 Existing Systems and Applications for Metadata Visualization

Visualization of metadata using visual forms is not a new research area. There are a quite number of systems and applications that have been developed and many of them provide good results for visualizing complex network either in two dimensional (2D) or three dimensional (3D) view. BiblioViz (Shen, Ogawa, Teoh & Ma 2006) and VisualLink (White, Lin & Buzydlowski 2004) are examples of those systems.

BiblioViz is a visualization system for bibliography data. It is developed based on the requirements analysis made from the other systems that have been presented in the InfoVis 2004 (Fekete, Grinstein & Plaisant 2004) contest. This contest was dedicated to visualizing the bibliography of the field of information visualization. BiblioViz is developed after examining the solutions for the bibliography visualization's problems from the solutions presented in the contest.

The interface of BiblioViz consists of five parts; table view, network view, paper details, data menu and user control

panel (see Figure 2). Network view visually conveys relationships between the data and table view showing time-related information. The relationships can be viewed in several different layout algorithms which are self-organizing maps (SOM) (Kohonen 1997), force-directed (Kamada & Kawai 1989) and centroid, in which each of them has their own advantages. It also provides two possible views of the network which are in 2D or in 3D. The user can interactively operate the system as it allows user to pick data entities, make dynamic queries and specify visualization parameters for the table view and network view.
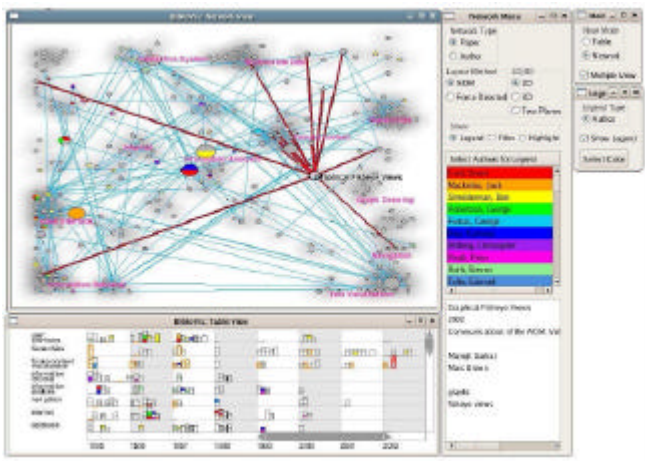

Figure 2: The screen shot of five parts of the BiblioViz

However, BiblioViz is a complex stand alone system and user requires some time to understand and be familiar with this system. Its ability to render 3D graphics and the use of lighting for depth cues may affect the loading time and the quality of the drawing, if a user's computer has a low capability. It can also becomes quite cluttered if too many visualizations are in the view.

One of the tasks that BiblioViz system has accomplished is to show the relationships between two or more of the authors in the database. The system includes a plane containing author nodes. Users can select the specific author they want to look into either by clicking a node or from a list in the network view. Users can see the other authors who have collaborated with earlier selection through the highlighted edges. With its interface that contains lots of nodes, the resulting display is often cluttered and it is quite hard to view the relationships.

VisualLink, an associative information visualizer have been developed to visualize the dataset from the InfoVis 2004 (see Figure 3). This is the continuation of work in the mapping of terms from the bibliographic records (White, Lin, Buzydlowski & Chen 2004). In VisualLink the collaboration between the authors were being addressed as "cross-map" (see Figure 4). There are a few key principles that the authors of VisualLink have considered in developing the visualizer that are innovative. Some of the principles include key word start, and associations between metadata sets.

Some of the VisuaLink principles have also been used this MetaVis interface. Users are able to create the visualization's graph with a single seed or phrase (keyword). Next, the resulting graph should reveal interesting associations hidden in the database, while the user's momentum is maintained by making the graph appear quickly after the keyword is entered. The MetaVis interface focus on the simplicity of design, thus limits the view in 2D only, so that user can easily view the interconnected metadata without clutter. The MetaVis interface also includes a search area for user to query and a panel that will list the details of selected paper or selected authors.
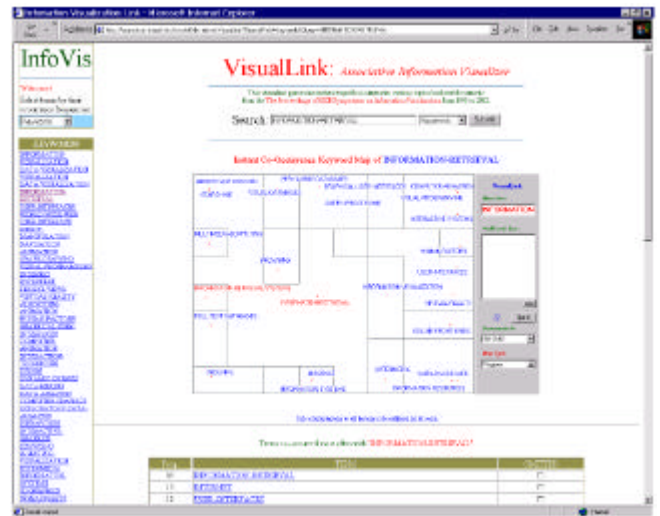

Figure 3: The screen shot of VisualLink interface

In MetaVis the navigation task is achieved through layers. The first layer navigation can be realized by displaying the details of an author or articles in nodes, which upon selection (by clicking at the nodes) will then enable the second navigation to the details of the other authors. A new graph will be drawn which contains the selected author as its centre. When a year is being selected, all the related work published by the related author will be displayed with the co-authors and the names of those journals in which those articles were published.
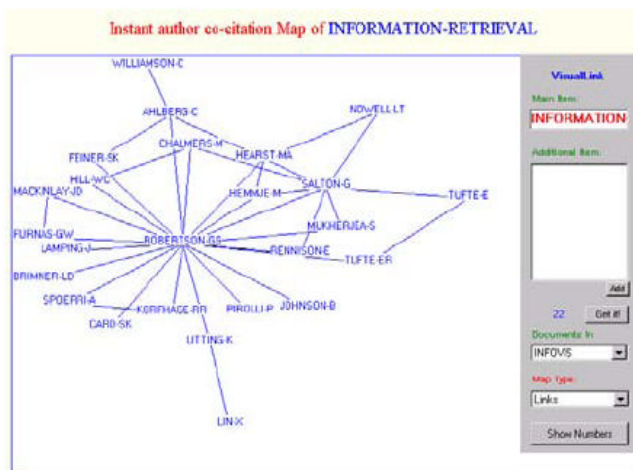
Figure 4: Network of co-cited authors in VisualLink

2.2 Tools and Packages for Visualizing and Analyzing Networks

For manipulating, analyzing and visualizing the metadata in the graph forms, there are several tools and packages that can be implemented. Some of them are UCINET (Borgatti, Everett & Freeman 2004), Pajek (Batagelj & Mrvar 2004), JGraph (Alder 2005) and JUNG. UCINET and Pajek are examples of standalone applications that each provides a number of tools for visualizing and analyzing networks. These software can successfully visualize the network and provide the user with some powerful visualization features such as vertices manipulations and rich collection of graphical primitives and output formats. Still, they cannot conveniently be addressed programmatically by other applications, and so they are not well-suited to process large numbers of graphs. Furthermore, they are applications rather than libraries. Other applications can employ these tools but user cannot write their own routines that take advantage of the capabilities of existing codes to handle especially complex data sets. In other words, these applications do not allow for the inclusion of other methods that are not provided as part of the application.

In addressing these rigidities, a third party library is chosen so that the functionalities of the existing code can be extended. One of them is JGraph, which is an open-source library. Although JGraph is oriented more toward design layout, it barely meets the needs in terms of graph operations. Moreover, although JGraph comes in three separate packages, only the first one is a free version. Unfortunately, the first package, JGraph comprises only the basic JGraph swing components.

Finally, the Java Universal Network/Graph (JUNG) is a completely free, open-source Java software library that provides a common and extensible language for the modeling, analysis, and visualization of data that can be represented as a graph or network. It can be implemented in any Java-based applications and it makes use of the extensive capabilities of the Java API. One of the

functionalities in JUNG that is suitable for visualizing networks is that it provides a mechanism for annotating metadata to the graphs, entities and relations. This facilitates the creation of analytic tools for complex data sets that can exa mine the relations between entities as well as the metadata attached to each entity and relation.

JUNG currently provides many of the tools and elements that are mostly required for writing network graph's applications. JUNG also supports a variety of representations of data sets and their relations, and includes a number of algorithms from graph theory, exploratory data analysis and machine learning. JUNG provides for a visualization framework that offers layouts and algorithms that user can choose to use or adapt to create their own custom algorithms.

These capabilities make JUNG framework a good platform for this project. It can be used along with Swing Application Programming Interface or API (Sun Microsystems 2004), to create a stunning interface. With this API, the interface was designed with functionalities that can assist users in navigating through the network.

3.0 MetaVis

MetaVis helps researchers, educators and learners to view the metadata of an article or journal in a graph of connected nodes. The interface uses graph to enable users to easily view and understand the connection of the metadata in the network. The interface is created in Java Applet using Java language. The visualization of the metadata network was created using JUNG framework.

A MySQL database was developed to store the required data that were needed to build the metadata networks. Some of the inputs are the author's details (name, email and address) and the journal's details (title, name of the authors, year published and publisher). Each of the data were manually added into the database since the purpose of this paper is only focused on the development of the interface, not on how to automatically extract the metadata from articles and stored in the database. The automatic extraction is written in another paper. The interface draws the metadata graph based on the query input by the user and the graph is redrawn with every new input.

3.1 How to use MetaVis

The interface of MetaVis contains a graph viewer, an attribute pane at left side and a text pane at the lower side. The nodes of the graph are the metadata and the edges (also called the connecting nodes). Metadata nodes contain author's details and journal details. Connecting nodes are the nodes that connect two metadata. Nodes representing connecting nodes are yellow in color and other colors represent metadata nodes. The interface also includes a set of toolbar and attributes panes that allow access to various operations.

The first operation when using the interface is **'Search'**. A user needs to enter a word in the **'Search'** field. There is also a radio button where user needs to select either **'author'** or **'title'** when searching for the metadata. For example an author named "Robert Shaw" is being searched and button **'author'** is selected. By clicking the **Find'** button, a list of names which are all related to the word "Robert Shaw" is displayed in a table. Users can select any content by clicking on it and click button **'View.'** The MetaVis will search in the database the entire network related to the searched metadata and the network is visualized as graph in the graph viewer with the selected metadata as the centre of the graph. The graph is in a radial layout form. The attribute pane will also contain all the details of "Robert Shaw" in a listing.

When users search for an author, the metadata that will be displayed are the author's name, email, address, list of years that the author have published, and list of other authors that he/she have collaborated with in writing related articles. By clicking a publication year node, for example "2005", a popup menu displaying a list of all the articles produced in that year will be displayed. When users click a title that they want to view further, the previous graph be erased and a new graph will be drawn with the selected title as its centre. These link nodes can be identified by colored nodes with a dashed outline. The text pane at the below will contains two tabs; the description panel and legend panel. The description panel will display all the metadata and details that contains in the graph on the graph viewer. This is an alternative way to view the metadata details of an author or an article. The legend panel will list the explanation for each of the colored nodes.

The nodes in the graph can be manually rearranged by the users if they find that the graph is cluttered. The graph is not the same every time it is drawn because the positioning is automatically calculated by the interface. In this interface the way the graph is drawn is not very important. More important is whether the metadata are related accordingly to each other.

In the toolbar area, there is toggle buttons which represent the **'Radial Tree Layout'** and **'Top-down Tree Layout'**. By default, when the users click the **'View'** button, the graph will be drawn in a radial layout. By clicking the **'Top-down Tree Layout'** button in the toolbar area the graph will automatically change to a top-down tree layout. The last function provided to users is the **'Help'** button where a pop up screen will be displayed which will briefly describe this interface and the functions it have.

4.0 CONCLUSIONS AND FUTURE WORK

Currently the interface can display a network of metadata of an author or an article. Later on, we will enhance this interface to be an efficient interface for metadata visualization. This prototype interface provides a simple and easy tool to view the connection of metadata for researchers. This interface is still new in the field of information visualization and the development and refinement of MetaVis is an on-going project.

Some of the unresolved issues that still need attention are the undo function and ability for the user to add or delete metadata to a network. The undo and redo function are also a must since there are a lot of functions that can manipulate the graph and which constantly changes it. If the users have the ability to manually manipulate the metadata either by adding, deleting or correcting them, the interface can have the potential of becoming a powerful and efficient interface while still maintaining its simplicity.

As we continue to receive more feedbacks, we will possibly add in new functions and conduct experiments that will enhance the effectiveness of the interface of MetaVis. Once ready, this interface will be embedded in the project website of an automatic knowledge search engine currently being developed concurrently. If extended further this interface can be incorporated in search engines to trace the details of an author and the co-authorship among the articles over a period of time, and within or across domains of knowledge disciplines.

REFERENCES

Alder, G. (2005). JGraph: The Java Graph Visualization Library. Retrieved February 4, 2008 from http://www.jgraph.com.

Batagelj, V. & Mrvar, A. (2004). Pajek: Program for Large Network Analysis. Retrieved February 4, 2008, from http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Borgatti, S., Everett, M. & Freeman, L. (2004). UCINET: Software for Social Network Analysis. Retrieved February 4, 2008, from http://www.analytictech.com/ucinet.

Fekete, J.D., Grinstein, G. & Plaisant, C. (2004). IEEE InfoVis 2004 contest, The History of InfoVis, Washington, DC, USA.

Gershon, N., Eick, S.G., & Card, S. (1998). Information visualization. *interactions, 5(2),* 9-15.

Kamada, T. & Kawai, S. (1989). An algorithm for drawing general undirected graphs, In *Information Processing Letters, 31(1)*, 7-15.

Kohonen, T. (1997). Self-organizing Maps, Springer-Verlag.

McCormick, B.H., DeFanti, T.A. & Brown, M.D. (1987). Visualization in scientific computing. In *Computer Graphics, 21(6)*.

O'Madadhain, J., Fisher, D., White, S., & Boey, Y.B. (2003). The JUNG (Java Universal Network/Graph) framework. Retrieved February 4, 2008, from http://www.datalab.uci.edu/papers/JUNG_tect_report.html.

Robertson, G.G., Card, S.K. & Mackinlay, J.D. (1993). Information visualization using 3D interactive animation. In *Communications of ACM, 36(4),* 57-71.

Shen, Z., Ogawa, M., Teoh, S.T. & Ma, K.L. (2006). BiblioViz: A System for Visualizing Bibliography Information. In *Proceedings of Asia-Pacific Symposium on Information Visualization 2006,* 93-102.

Sun Microsystems (2004). Creating a GUI with JFC/Swing. Retrieved February 4, 2008 from *http://java.sun.com/docs/ books/tutorial/uiswing/.*

White, H., Lin, X. & Buzydlowski, J. (2004). An associative information visualizer. In *The History of Infovis (Infovis 2004 Contest),* IEEE Computer Society, Washington, DC, USA.

White, H., Lin, X., Buzydlowski, J. & Chen, C. (2004). User-controlled mapping of significant literatures. In *Proceedings of the National Academy of Sciences 101 (suppl. 1),* 5297-5302.