Interactive Poster: Towards an XML Toolkit for a Software Repository Supporting Information Visualization Education and Research

Jason Baumgartner and Katy Börner
Indiana University, SLIS
Bloomington IN 47405, USA
jlbaumga@indiana.edu, katy@indiana.edu
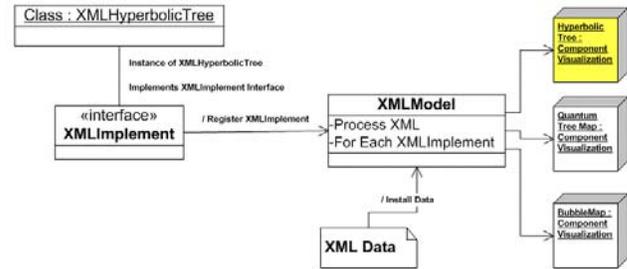
Motivation

Many information visualizations (IVs) are highly interactive and dynamic. In order to design effective information visualizations, theoretical knowledge on human perception as well as available approaches, techniques and their evaluation is essential. In addition, practical skills are needed to select appropriate algorithms, interaction metaphors, and visual designs that best fit a user group, their tasks, and their data. A number of excellent IV textbooks have been published recently by Card et al. (1999), Chen (1999), Ware (2000), Spence (2000), and Dodge & Kitchin (2000). However, 2-dimensional printouts on paper often cannot convey the true visual appearance and interactive performance of IVs. While several textbooks come with accompanying web sites containing screen-sized snapshots of user interfaces as well as animations and movies, none of them facilitates the exploration, application, evaluation, and comparison of algorithms.

XML Toolkit

Since 2000, our research group is working on the design of an *Information Visualization Data and Software Repository* (Börner & Zhou, 2001) for teaching and research purposes. Currently, the repository[1] provides access to diverse data mining and dimensionality reduction techniques such as Latent Semantic Analysis, Multidimensional Scaling, and Pathfinder Networks, as well as to spatial layout algorithms such as the GRIDL GRaphical Interface for Digital Libraries, diverse Treemap algorithms, Force Directed Placement, and Hyperbolic Tree algorithm.

This poster introduces an XML toolkit that extends the repository to create a unified architecture in which multiple data analysis and information visualization algorithms can be incorporated and combined. The toolkit provides an XML-based interchange format to unify data input, interchange, and output formats. Factory and interface classes allow all software packages to implement and to use the standard XML

---

[1] http://ella.slis.indiana.edu/~katy/L697/code/

format and ensure that packages can be easily interchanged, compared, and combined. Simple configurations of the XML input format suffice to use different algorithms in a wide variety of applications. All Java-based algorithms can be run in stand-alone

mode as an applet or application.

Figure 1: General Architecture of the XML Toolkit

The general structure of the IV repository XML toolkit, depicted in Figure 1, relies on the use of factory and interface classes to instantiate and populate the various visualization algorithms. The algorithm classes must implement the XMLImplement class to be registered with the toolkit. The XML data and the interfaced objects are managed by a model that controls access of the data and its population into the objects. Thus a Hyperbolic Tree, a Quantum Treemap, and a Bubblemap can all implement this interface, register with the model, and then concurrently visualize and interact with the same dataset.
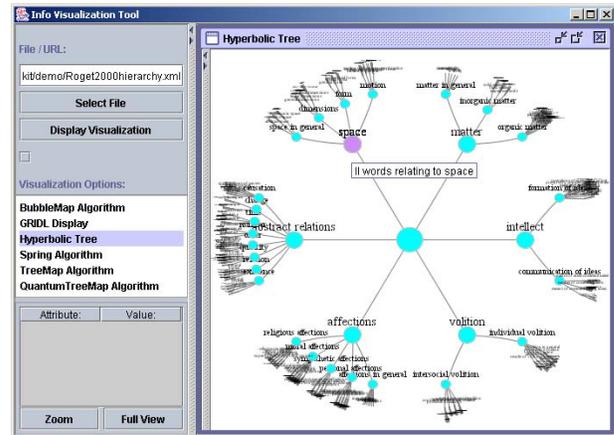

Figure 2: Graphical User Interface to the Current Software Repository

Non-programmers with a basic understanding of the included algorithms can use a graphical user interface to visualize their data sets. Figure 2 shows a snapshot of the interface with a visualization of Roget's Thesaurus using a hyperbolic tree layout (Baumgartner & Waugh, 2002). The left control panel can be used to load XML

formatted data files, e.g., "Roget2000hierarchy.xml", and to select the preferred "Visualization Option", e.g., "Hyperbolic Tree". Upon clicking "Display Visualization" the data set is displayed for interactive exploration in the right window. The control panel can be minimized and a control panel specific to the Hyperbolic Tree can be opened to choose among diverse display options. Alternatively, a different layout algorithm suited to a hierarchical data set, e.g., Treemaps, can be selected. Multiple visualizations of one data set can be generated simultaneously for comparison.

The repository was used in the Information Visualization course taught at Indiana University to complement the theoretical study of information visualization techniques and to augment the critique and evaluation of existing applications. It facilitates the design, comparison, and customization of IVs for different user groups, visualization tasks, and data sets. Students can concentrate on interface and interactivity aspects of visualizations. In addition, the repository enabled students to take on class projects of considerable complexity. Final projects completed in 2001 resulted in multiple conference and workshop papers (Baumgartner & Waugh, 2002; Feng & Börner, 2002; Mongin, Mostafa, & Fieber, 2001; Paolillo & Heald, 2002). Paper write-ups based on the 2002 final project results are currently under review.

The toolkit is freely available for non-commercial purposes. Our hope is that it will facilitate easy sharing, comparison, and evaluation of existing and new IV algorithms and help to collectively understand the underlying issues of differing visualizations, and to pool together existing and future IV efforts.

Outlook

In the near future, users will be able to use the XML toolkit to save interaction data such as manipulation changes in the data, the state of the visualization, etc. that could be advantageous to compare visualizations of different data sets among others.

Metadata schemas, like the Dublin Core and the Resource Description Framework (RDF), will be employed to provide an interoperable way to represent meaning with data. There will be two schema sets, one for retaining information about a data resource and another one to provide information for a particular service. The Dublin Core tag set will be used as the framework for the data schema, which will store information about a given resource. There will also be a schema set that extends the Dublin Core and RDF to represent information about displaying content in a given IV algorithm. The schema set will be registered with the Open Archives Initiative (OAI) protocol (Lagoze & Sompel, 2001) to allow the greatest interoperability of the data.

Furthermore, the existing schemas will all be centered on exclusive document description, vector graphic markup, geographical layout, etc. The full schema set, that is currently under review, defines simple format objects, encapsulates the more complex format structures through importing the Scalable Vector Graphics 1.0 (http://www.w3.org/TR/SVG/) specification and uses it within a given structure that subclasses the top-level container class for that specification. A description of the full schema set is available for review and comments at http://ella.slis.indiana.edu/~jlbaumga/infovis/infovis.rdf.

Acknowledgements

References

Baumgartner, J., & Waugh, T. (2002, January 20-25). *Roget2000: A 2D hyperbolic tree visualization of Roget's Thesaurus.* Visualization and Data Analysis. Proceedings of SPIE, San Jose, CA, USA,

Börner, K., & Zhou, Y. (2001, July 25-27). *A Software Repository for Education and Research in Information Visualization.* Information Visualisation Conference, London, England: IEEE Press, pp. 257-262.

Card, S., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). *Readings in Information Visualization: Using Vision to Think*: Morgan Kaufmann.

Chen, C. (1999). *Information Visualisation and Virtual Environments*. London: Springer Verlag.

Dodge, M., & Kitchin, R. (2000). *Mapping Cyberspace*: Routledge.

Feng, Y., & Börner, K. (2002, January 20-25). *Using Semantic Treemaps to Categorize and Visualize Bookmark Files.* Visualization and Data Analysis. Proceedings of SPIE, San Jose, CA, pp. 218-227.

Lagoze, C., & Sompel, H. V. (2001). *The Open Archives Initiative: Building a low-barrier interoperability framework.* First ACM+IEEE Joint Conference on Digital Libraries, Portland, Oregon, USA: ACM Press,

Mongin, L., Mostafa, J., & Fieber, J. (2001). *Using a Spring Embedding Algorithm to Display Term Relationships from a Medical Concept Discovery System.* Annual Conference of the American Medical Informatics Association, pp. 661-662.

Paolillo, J., & Heald, D. (2002). *Demographic Participation in the Discursive Management of Usenet.* 35th Hawaii International Conference on System Sciences, Wailea, Maui, Hawaii: IEEE Press,

Spence, B. (2000). *Information Visualization*: Addison-Wesley.

Ware, C. (2000). *Information Visualization: Perception for Design*: Morgan Kaufmann.