Fusion: Interactive Coordination of Diverse Data, Visualizations, and Mining Algorithms

Chris North, Nathan Conklin, Kiran Indukuri, Varun Saini, Qiang Yu
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
north@cs.vt.edu
http://infovis.cs.vt.edu/

ABSTRACT
Fusion is a web-based system that enables end-users to rapidly and dynamically construct personalized visualization workspaces without programming. Users first use advanced *data schemas* to link diverse data sources. Then they use *visualization schemas* to coordinate visualization components and data-mining algorithms according to the unique needs of their data and tasks. They create a custom interactive visualization workspace that can be published on the web. This is accomplished through the Fusion model and user interface that is based on schema concepts that are easy to learn and simple to use.

Keywords
Information visualization, multiple views, data mining.

INTRODUCTION
When working with large and complex data, users often need custom designed information visualization workspaces that integrate diverse data sources and coordinate diverse visualization and data mining tools. Custom programming of such workspaces is time-consuming and expensive. For example, in bioinformatics, biologists need to integrate their experimental data with data from collaborating labs, then analyze it using a combination of in-house and external tools, and finally publish results on the web. A coordinated *multiple-view* visualization [1], such as the workspace shown in Figure 1, provides users with the integrated interface needed for their analysis.

The Fusion model and user interface enable users to dynamically construct the needed visualization workspace. Users specify integrated data structure using data schemas and visualization structure using visualization schemas.

DATA FUSION
Relational data schemas enable users to structure data based on the relational data model. Many database systems provide a visual language for organizing relations into data schemas in the form of a diagrammatic, direct manipulation user interface. Tools such as Visage/VQE [2] enable users to perform operations on tuples and attributes in different relations that share joined entities. We expand on the data schema model by enabling users to associate data across separate data sources and integrate data mining algorithms. The user visually specifies a unified data schema by establishing associations between relations.

In the example in Figure 1, a biologist connects to micro-array experiment results from the lab's database. She then connects to a public database containing a hierarchical gene categorization. In the data schema, she associates the relations between the separate databases, forming a personalized database. She can now use this integrated data schema for visualization.

VISUALIZATION FUSION
Users then specify visualization schemas to coordinate multiple-views based on the Snap-Together Visualization model [4]. Visualization schemas establish a direct correspondence between data schema primitives and visualization schema primitives:

- *Visualization component* = data relation. A visualization component is a view that displays a data relation or query result.
- *Visual item* = data tuple. Data tuples are displayed as visual items in a visualization component.
- *Visual property* = data attribute. Visualization components map data attributes to visual properties of the visual items.
- *Visualization coordination* = data association join. A coordination between two components tightly couples them according to the join between their relations.

Visualization schema diagrams are represented visually as a graph structure similar to data schema diagrams (Figure 1). A node represents an instantiated visualization components along with the data relation displayed in the visualization. Edges represent coordinations between visualizations. Coordinated visualization components are tightly coupled so that interacting with tuples in one visualization causes immediate visual feedback on joined tuples in the other visualization.

CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.
ACM 1-58113-630-7/03/0004.

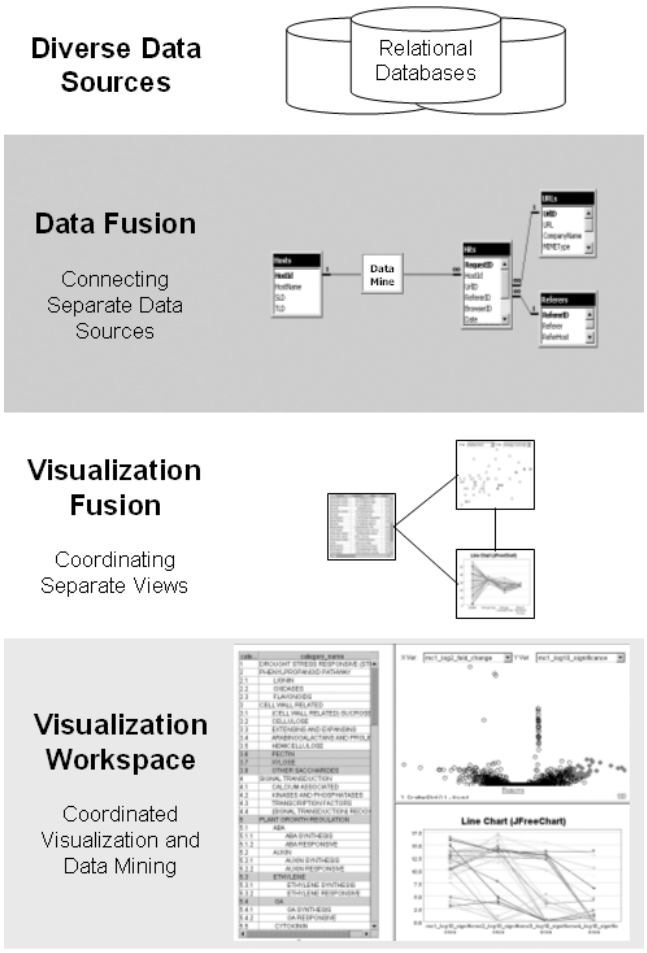

Figure 1: The integration of data fusion and visualization fusion enable interactive coordinated visualization.

Visualization schemas extend data schemas by enabling users to coordinate visualization components that encapsulate data from different sources. A user coordinates two visualization components by choosing the actions to tightly couple. Users can then organize and interact with the instantiated visualization components in the visualization workspace. Directly manipulating visualization schemas enables users to rapidly construct and modify visualization workspaces.

In the example in Figure 1, the biologist creates a scatter plot of experimental results and a line graph of related experimental data. She coordinates these views so that selecting genes of interest in the plot displays additional details about them in the line graph. The biologist then creates a tree-view of the hierarchical gene categories extracted from a public database. These views are coordinated based on associations defined in the data schema. Now corresponding categories are highlighted when genes are selected in the plot, enabling her to analyze functional relationships of genes.

DATA MINING

Discovery tools that combine information visualization and data mining can enable more effective data analysis while preserving user control [5]. Fusion integrates data mining components in the data schema and visualization schema. Mining components can be specified in the data schema to establish dynamic associations between relations. In the visualization schema, mined associations extend the functionality of coordinations between visualizations. User interactions on coordinated visualizations identify inference rules that are augmented with statistical calculations of significance, and provide input for mining parameters. Mining algorithms, such as ILP Redescription [3], can dynamically compute additional interesting inference rules based on the data, that users can then interactively view in the coordinated visualizations. Hence, in an interactive loop, users' exploratory interactions provide guidance for mining algorithms, which in turn guide users' exploration.

In the example in Figure 1, the biologist integrates a mining component in the data and visualization schemas. The mining algorithm suggests selections for brushing-and-linking between the plot and the categories. When suggested gene clusters are selected in the plot, the corresponding functional categories are highlighted based on significance levels, enabling her to discover new unexpected functional relationships.

CONCLUSION

Fusion integrates diverse data, visualizations, and mining algorithms to enable user-customized visualization workspaces. It supports the use of diverse visualizations produced both in-house and by third party developers, and can publish workspaces on the web. Data schemas enable users to associate related data from different sources. Visualization schemas provide a visual overview and user interface to the coordination structure. The integration of data mining into the model enables Fusion to act as a user-customized discovery tool.

REFERENCES

1. Baldonado, M., Woodruff, A., Kuchinsky, A., Guidelines for Using Multiple Views in Information Visualization, *Proc. ACM Advanced Visual Interfaces*, pg. 110-119, (2000).
2. Derthick, M., Roth, S., Kolojejchick, J., Coordinating declarative queries with a direct manipulation data exploration environment, *Proc. IEEE Information Visualization Symposium*, pg. 65-72, (1997).
3. Heath, L., Ramakrishnan, N., Sederoff, R., Whetten, R., Chevone, B., Struble, C., Jouenne, V., Chen, D., van Zyl, L., Grene, R., Studying the Functional Genomics of Stress Responses in Loblolly Pine using the Expresso Microarray Management System, *Comparative and Functional Genomics*, 3, pg. 226-243, (2002).
4. North, C., Conklin, N., and Saini, V., Visualization Schemas for Flexible Information Visualization, *Proc. IEEE Information Visualization Symposium*, pg. 15-22, (2002).
5. Shneiderman, B., Inventing discovery tools: combining information visualization with data mining, *Information Visualization*, 1(1), pg. 5-12, (2002).