

Herramientas de Visualización para la Exploración de Datos

Sergio Martig – Silvia Castro
Martín Larrea – Sebastián Escarza
Dana Urribarri – Maximiliano Escudero – Luján Ganuza

Departamento de Ciencias e Ingeniería de la Computación
Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica
Universidad Nacional del Sur

{srm, smc, mll}@cs.uns.edu.ar

Resumen

Los recientes avances en la tecnología disponibles para la sociedad en general están generando inmensos conjuntos de datos. Estos conjuntos desafían la habilidad de las comunidades científica y no científica en lo referente a medir, analizar y visualizar datos. Al mismo tiempo, las computadoras se están convirtiendo en una herramienta cada vez más habitual en la vida de cada vez más personas. Esto sugiere una necesidad crítica de herramientas y ambientes que habiliten un análisis visual de los conjuntos de datos que sea efectivo y eficiente y que pueda adaptarse tanto a las necesidades de los usuarios como a las de los diseñadores de las técnicas en los distintos dominios de Visualización

1. Introducción

Una visualización exitosa puede reducir considerablemente el tiempo que se tarda en entender los datos subyacentes, en encontrar relaciones y en obtener la información que se busca. Para generar una visualización, es necesario un mapeo de los datos en el espacio cartesiano de dos o tres dimensiones, que represente las relaciones contenidas en los mismos de manera tan intuitiva como sea posible. Esto debe permitir al usuario de la visualización usar sus habilidades innatas para entender las relaciones espaciales. Encontrar una buena representación espacial de los datos es una de las tareas más difíciles en la visualización de información abstracta.

Las técnicas de Visualización no tratan de sustituir la capacidad humana en esta tarea de búsqueda de información sino que pretenden ser usadas para amplificar su efectividad. En este contexto, los datos a ser visualizados deben presentarse de modo tal que el sistema de percepción visual del usuario de la visualización sea estimulado para entender relaciones y reconocer patrones. Dicho usuario responde con una cierta técnica de interacción especificando, por ejemplo, cierta área de interés o cambiando los parámetros de vista. La visualización provee entonces realimentación al usuario y se establece entonces un ciclo interactivo. Esto se repite hasta que se encuentre la información buscada o hasta que, por ejemplo, el usuario elija una visualización alternativa.

La diversidad de aplicaciones que se benefician con Visualización de Información es grande y crece constantemente siendo actualmente un área de activo desarrollo. Así como los usuarios creativos empujan los límites de las herramientas actuales, los diseñadores serán presionados para proveer aún mayor funcionalidad.

El desarrollo del área de investigación es un gran aporte al desarrollo teórico básico, sumamente necesario, en el campo emergente de Visualización. Esto redundará sin duda en el beneficio de los diferentes dominios de aplicación.

En este sentido, podemos afirmar que los resultados obtenidos en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) serán de gran importancia para el desarrollo del sector de Software y Servicios Informáticos y de TICs.

2. Contexto de la Investigación

Debido a la gran variedad de dominios de Visualización, el desafío es diseñar un ambiente que permita a los usuarios llevar a cabo, de manera intuitiva, una variedad de tareas de visualización. Aunque diferentes dominios de aplicación requieren representaciones visuales distintas, muchos de éstos comparten operaciones de transformación de los datos y manipulaciones de los mismos que son similares a lo largo de todo el proceso de visualización.

En este contexto ya hemos desarrollado un Modelo Unificado de Visualización (MUV), que constituye un modelo de referencia para la Visualización; hemos analizado y categorizado las similitudes entre distintos dominios de aplicación y hemos caracterizado las interacciones que puedan tener lugar en las distintas etapas del proceso. Precisamente se tomará este modelo como punto de partida para el diseño de un ambiente de visualización, que:

- Brinde a los usuarios la posibilidad de utilizar técnicas de visualización que le permitan adquirir nuevos *insights* en sus datos mediante una completa interacción.
- Brinde a los diseñadores de Sistemas de Visualización un marco de referencia para generar, adquirir o adecuar datos, incorporar nuevas técnicas de Visualización y elementos de interacción, permitir incorporar elementos de historia, etc..
- Permita su descomposición con el objetivo de maximizar el aprovechamiento de recursos.

El desarrollo de un prototipo de una herramienta de Visualización como la planteada, teniendo como base el MUV propuesto previamente, puede ser potencialmente transferible a distintas áreas del sector productivo. Algunas de las posibles transferencias que se podrían realizar son a:

- Visualización aplicada a la WEB, a Bases de Datos, a Sistemas Distribuidos, a Conjuntos de Documentos Complejos, Ontologías, etc.
- Visualización en la WEB. Web Services para visualización.
- Interfaces WEB, multimodales, juegos y sistemas en general.

3. Objetivos

El objetivo general de este proyecto es la aplicación de los resultados obtenidos en el MUV que ya hemos elaborado como base para el desarrollo de una plataforma coherente que ofrezca funcionalidad dinámica adecuada para llevar a cabo tareas de análisis y visualización de datos. Las características relevantes de dicha plataforma deben ser la capacidad de conexión a fuentes de datos externas, la consideración de las restricciones temporales y los elementos de historia, la posibilidad de distribución de las tareas de los distintos subsistemas de un sistema de visualización, y el aseguramiento de la compatibilidad con los paquetes estándares existentes tanto gráficos como de visualización. La intención primaria es poder contar con un ambiente de Visualización que:

- Brinde a los *usuarios de Sistemas de Visualización* la posibilidad de utilizar técnicas de visualización que le permitan adquirir nuevos *insights* en sus datos mediante una completa interacción.

- Brinde a los *diseñadores de Sistemas de Visualización* un marco de referencia para generar, adquirir o adecuar datos, incorporar nuevas técnicas de Visualización y elementos de interacción, permitir incorporar elementos de historia, etc..

Así, podemos enunciar que los *subobjetivos generales* son:

- La determinación de los distintos tipos de operadores en el contexto de las distintas técnicas de Visualización, teniendo como base el modelo desarrollado previamente, alrededor de los distintos estados de los datos y sus transformaciones. Esto podrá usarse posteriormente para construir un sistema modular.
- Dado el marco que permite lograr el entendimiento del espacio de diseño de las técnicas de Visualización, la extracción de
 - las *operaciones* cruciales en cada técnica, mediante un análisis de las similitudes y las diferencias entre los operadores en diferentes dominios de datos y
 - las *interacciones* requeridas por cada técnica.
- A partir del MUV se procederá a incorporar la *semántica del proceso y al proceso de visualización*. Consideramos que esta tarea es crucial por un lado para plasmar la semántica del proceso, dirimiendo ambigüedades y definiendo un marco sólido de discusión y evaluación. Por otra parte contar con la semántica del proceso, permitirá asistir al los usuarios en la tarea de visualizar sus datos, minimizando la sobrecarga del proceso en si mismo y potenciando las capacidades de exploración y análisis de la información. En lo atinente a este subobjetivo proponemos
 - Generar una Ontología de Visualización
 - Explotar la semántica para brindar asistencia y guiar el proceso de visualización.
- Otro aspecto crucial es el de la *escalabilidad* no solo a nivel de los datos manipulables por la herramienta, sino de la escalabilidad visual de las representaciones generadas.
- Permitir que la herramienta o el ambiente de visualización opere de manera distribuida en la red. Para abordar este objetivo se ha comenzado a estudiar y diseñar *Servicios Webs de Visualización*.

Somos concientes que los subobjetivos planteados son ambiciosos, pero tenemos como meta la obtención de un prototipo operacional, teniendo en cuenta los resultados de las etapas previas y del análisis de usabilidad. Esto ayudará en el proceso de desarrollo permitiendo identificar los problemas potenciales de performance en forma temprana y permitiendo tener un prototipo en las primeras etapas del desarrollo. Para el prototipo se seleccionará un conjunto de técnicas generales de visualización que serán elegidas teniendo en cuenta que deben permitir visualizar grandes volúmenes de datos multidimensionales, considerando la escalabilidad de los mismos como un factor esencial. La interafaz de este prototipo tendrá en cuenta los distintos tipos de usuarios para los que fue diseñada y los asistirá en todas las etapas del proceso de Visualización mediante una interacción apropiada

Siguiendo los lineamientos definidos para los Sistemas de Visualización, nos centraremos en un dominio que, si bien es general, constituye un conjunto restringido del área de Visualización. El enfoque sobre un dominio particular ayudará a cubrir la distancia entre las propuestas teóricas y las aplicaciones de las mismas, constituyendo un modo de establecer el valor de las propuestas realizadas. Se trabajará en el área particular de Visualización de Grafos. Esta área de aplicación si bien es particular, es de gran potencialidad por ser el modelo subyacente en distintos dominios de aplicación.

Bibliografía

- 1 Card, S., Mackinlay, J., Shneiderman, B., *Readings in Information Visualization – Using Vision to Think*, Morgan Kaufmann, 1999.
- 2 Chi, E., *A Taxonomy of Visualization Techniques using the Data State Reference Model*, IEEE Visualization 2000, Actas en CD-Rom.
- 3 Connelly, Ch.S., *Towards An Understanding Of DCs Control Operator Workload*, Beville Engineering, Inc.
- 4 Hauser, H., Maktovik, K., Sianitzer, R., Gröller, E., *Process Visualization with Level of Details*, Proceedings of the Conference on Visualization 2002. IEEE Computer Society Press.
- 5 Charles D. Hansen and Christopher R. Johnson, editors. *The Visualization Handbook*. Elsevier, Academic Press, 2005
- 6 Keim D. A., Kriegel H.-P., Seidl T., *Supporting Data Mining of Large Databases by Visual Feedback Queries*, Proc. 10th Int. Conf. on Data Engineering, Houston, TX, 1994, pp. 302-313.
- 7 Kelly, P., Keller, M., *Visual Cues: Practical Data Visualization*, IEEE Computer Society Press, 1992.
- 8 Lamping, J., Rao, R., *The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies*, Journal of Visual Languages and Computing, 7(1), pp. 33-55.
- 9 Martin, A., Ward, A., *High dimensional brushing for interactive exploration of multivariate data*, Proceedings of the Conference on Visualization 1995, pages 271–278. IEEE Computer Society Press.
- 10 Robertson, G., Card, S., Mackinlay, J., *Information Visualization Using 3D Interactive Animation*, Communications of the ACM, 36(4), pp. 56-71, 1993.
- 11 Strohbar, D.A., *Human Factors in Distributed Process Control*, Belville Engineering, Inc.
- 12 Thomas, James J. and Cook, Kristin A., editors. *Illuminating the Path: The Research and Development*. Agenda for Visual Analytics. National Visualization and Analytics Center, 2005.
- 13 Tufte, E.R., *The Visual Display of Quantitative Information*, Cheshire, CT Graphics Press, 1983.
- 14 Tufte, E.R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, CT Graphics Press, 1997.
- 15 Wong P.C., Bergeron R.D., *30 Years of Multidimensional Multivariate Visualization*, Proc. Workshop on Scientific Visualization, IEEE Computer Society Press, 1995.