

Métricas, Técnicas y Semántica para la Visualización de Datos

Silvia Castro^{1,2,3}, Martín Larrea^{1,2,3}, Dana Urribarri^{1,2}, Luján Ganuza^{1,2}, Sebastián Escarza^{1,2}

¹Laboratorio de I+D en Visualización y Computación Gráfica (UNS-CIC Prov. de Buenos Aires)

²Departamento de Ciencias e Ingeniería en Ciencias de la Computación (DCIC)

³Instituto de Ciencias e Ingeniería de la Computación (ICIC),

Universidad Nacional del Sur, Av. Alem 1253, Bahía Blanca

{smc, mll, dku, mlg, se}@cs.uns.edu.ar

RESUMEN

Un desafío importante al iniciar un proceso de visualización es el de encontrar una representación adecuada para un conjunto de datos y brindar las interacciones apropiadas. Esta línea de investigación tiene como objetivo estudiar nuevas alternativas para asegurar que el resultado de un proceso de visualización sea satisfactorio, tanto para aquel que crea la visualización como para el que la consume. Esta línea de investigación incluye el desarrollo y perfeccionamiento de un modelo unificado de visualización, métricas para la evaluación de técnicas de visualización, técnicas para dominios específicos de aplicación y la asistencia al usuario durante la creación de una visualización mediante la incorporación de información semántica y razonadores semánticos.

Palabras clave: *Ontología, Representación Formal, Visualización basada en semántica, Visualización.*

CONTEXTO

Este trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab, <http://vyglab.cs.uns.edu.ar>) del Departamento de Ciencias e Ingeniería de la Computación, de la Universidad Nacional del Sur. Los trabajos realizados bajo esta línea involucran a docentes investigadores, becarios doctorales,

alumnos de grado y investigadores extranjeros. La línea de Investigación presentada está inserta en el proyecto acreditado *Análisis Visual de Grandes Conjuntos de Datos (24/N037)*, dirigido por la Dra. Silvia Castro y en el proyecto *Visualización Basada en Semántica (24/ZN29)* dirigido por el Dr. Martín Larrea; ambos financiados por la Secretaría General de Ciencia y Tecnología de la Universidad Nacional del Sur.

1. INTRODUCCIÓN

El objetivo de una visualización es encontrar una metáfora visual que permita entender y percibir en forma efectiva un conjunto de datos. Una visualización debe proveer también un conjunto de interacciones a partir de las cuales el usuario explorará el conjunto de datos con una mínima carga cognitiva. La tecnología computacional actual permite la exploración de grandes conjuntos de información; dichos conjuntos se hacen disponibles cada día a través de diferentes sistemas de información; estos datos varían entre datos no estructurados, documentos multimediales e información estructurada en base de datos. Por un lado, esta situación es extremadamente útil y excitante, pero la creciente cantidad de información genera una sobrecarga cognitiva llegando a un punto de ansiedad. Actualmente existe un gran número de

modelos de referencia en Visualización ([1, 2, 3]) que, de diversas maneras, han identificado y delineado los principales procesos y transformaciones que sufren los datos para ser visualizados. En particular, en el seno del VyGLab se ha desarrollado el Modelo Unificado de Visualización (MUV) [4], un modelo que constituye un marco conceptual de referencia en términos del cual ubicar los procesos y estados intermedios de los datos y definir las interacciones explícitamente. Si bien estos esfuerzos por compilar, caracterizar y clasificar los aspectos más relevantes del área han sido y siguen siendo fructíferos, aún no existe un consenso definitivo que permita consolidar una teoría de base en Visualización. Sobre este modelo hemos trabajado sobre la incorporación de semántica. La incorporación de una semántica bien definida en el proceso de Visualización ya ha sido identificada como una necesidad ([5, 6]) en el sentido que permitiría formalizar dicho proceso. A partir de esta formalización, se definirá un vocabulario común que permitirá a los usuarios plantear sus requerimientos, y a los diseñadores de herramientas de visualización, expresar las transformaciones de los datos. También sobre el MUV hemos trabajado en la definición de las interacciones sobre el proceso, tanto de alto como bajo nivel. Hemos incorporado al proceso de visualización técnicas específicas a dominios de aplicación, como es el caso de las Ciencias Geológicas y Captura de Movimiento. También, ante la necesidad de contar con métodos que nos permitan evaluar la calidad de una visualización, hemos desarrollado métricas para tal sentido.

2. LÍNEAS DE INVESTIGACIÓN Y

DESARROLLO

Este trabajo presenta tres ejes de investigación las cuales se entrelazan y tienen varios puntos en común. Tales ejes son:

1. “Visualización de Ciencias Geológicas”.
2. “Visualización Basada en Semántica”.
3. “Predicción del desempeño de las técnicas de visualización a partir de métricas sobre los datos”

Visualización de Ciencias Geológicas.

Este trabajo se enfoca en la visualización de un conjunto de datos geológicos, en particular, el conjunto de minerales que integran el grupo de los Espinelos. Este grupo de minerales resulta un excelente candidato a ser explorado y visualizado ya que es considerado un importante indicador petrogenético. Estos minerales proveen información vital en lo referido al ambiente tectónico de las rocas presentes en determinada área en el contexto de la tectónica global. En 2001, Barnes y Roeder ([7]) compilaron en una base de datos más de 26.000 análisis de Espinelos correspondientes a rocas ígneas y metamórficas. En base a estos análisis delinearon contornos para un conjunto de campos composicionales característicos (o patrones). Los geólogos suelen utilizar estos contornos para estimar el ambiente tectónico donde una muestra de Espinelos se podría haber formado. Esta tarea es propensa a errores e involucra una tediosa comparación manual de diagramas superpuestos. El objetivo de este trabajo de investigación es desarrollar una herramienta de Visualización que no solo permita visualizar y explorar conjuntos de datos correspondientes a composiciones minerales pertenecientes al grupo de los Espinelos, sino que también ofrezca técnicas que permitan categorizar

semiautomáticamente e interactivamente un conjunto arbitrario de Espinelos en función de los contornos de Barnes y Roeder.

Visualización Basada en Semántica. Este eje tiene como objetivo crear un modelo de visualización que considere la semántica de los datos, del contexto y de las etapas del proceso de visualización para poder asistir al usuario en el seteo de los parámetros de la visualización para lograr un resultado efectivo. Este nuevo modelo transformará datos en información, de acuerdo a Keller & Tergan [8], *“information is data that has been given meaning through interpretation by way of relational connection and pragmatic context”*. Los mismos datos pueden dar origen a diferente información, de acuerdo a la interpretación que se les de. Considerando esto, el nuevo proceso de visualización será capaz de determinar las características de una representación efectiva y podrá guiar al usuario a través de los diferentes estados del modelo. De esta manera, se logrará desarrollar un meta modelo para el proceso de visualización, los estados, los datos, las interacciones y el contexto.

Predicción del desempeño de las técnicas de visualización a partir de métricas sobre los datos. El objetivo de una visualización es obtener una representación del conjunto de datos que ayude al usuario en la correcta interpretación de los mismos y así lograr un acertado análisis de estos. Dado el constante crecimiento de los conjuntos de datos en diferentes y variados campos de la información, la tarea de elegir la técnica más adecuada para visualizar convenientemente los datos no es sencilla.

Además, el resultado del proceso de visualización depende de todas las decisiones que se hayan tomado a lo largo de dicho proceso: un usuario inexperto es propenso a tomar decisiones equivocadas afectando negativamente la visualización obtenida y, a la larga, frustrando su experiencia con la visualización. Si bien a la hora de visualizar conjuntos de datos pequeños no hay grandes desafíos, la situación cambia al intentar visualizar grandes conjuntos de datos: una mala decisión tomada en cualquier punto del proceso de visualización y el resultado obtenido puede no ser satisfactorio. Una alternativa para solucionar este problema es guiar al usuario en la toma de decisiones a lo largo del proceso. Sin embargo, esta tarea no es sencilla: implica la existencia de herramientas que permitan predecir qué decisión es “más conveniente” tomar. Una forma de elegir la decisión más conveniente es basarse en métricas sobre los datos que describen aspectos claves de la técnica y permitan predecir el resultado final sin necesidad de aplicar la técnica sobre los datos.

3. RESULTADOS Y OBJETIVOS

Publicaciones realizadas bajo el eje de investigación “**Visualización de Ciencias Geológicas**”: [9, 10, 11, 12, 13, 14]

Publicaciones realizadas bajo el eje de investigación “**Visualización Basada en Semántica**”: [15, 16, 17, 18, 19, 20, 21].

Publicaciones realizadas bajo el eje de investigación “**Predicción del desempeño de las técnicas de visualización a partir de métricas sobre los datos**”: [22, 23]

4. FORMACIÓN DE RECURSOS

HUMANOS

En lo concerniente a la formación de recursos humanos se detallan las tesis en desarrollo y concluidas relaciones con las líneas de investigación presentadas, así también como un proyecto en ejecución:

Tesis en Desarrollo

Tesis Doctoral. Sebastián Escarza. Tema: Ontologías de Visualización. Dirección: Dra. Silvia Castro.

Tesis Magister. Georgina Inés Cerúsico. Tema: Taxonomías de Técnicas de Visualización para la Formalización del Proceso de Visualización. Dirección: Dra. Silvia Castro y Dr. Martín Larrea.

Tesis Magister. Alejandra Elizabeth Herrera. Tema: Coordenadas Paralelas. Visualización e Interacciones. Dirección: Dra. Silvia Castro.

Tesis Magister. César Escobal Blanco. Tema: La Reducción de Datos en las Técnicas de Visualización. Dirección: Dra. Silvia Castro.

Tesis Concluidas

Tesis Doctoral. María Luján Ganuza. Tema: Servicios Web en Visualización. Dirección: Dra. Silvia Castro.

5. BIBLIOGRAFÍA

[1] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In Proceedings of the IEEE Symposium on Information Visualization (InfoVis'00), page 6975. IEEE Computer Society Press, 2000.

[2] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. IEEE Symposium on Visual Languages, 0:336, 1996.

[3] K. Brodlie and N. M. Noor. Visualization Notations, Models and Taxonomies. pages 207–212, Bangor, United Kingdom, 2007. Eurographics Association.

[4] Martig S., Castro S., Fillotrani P., Estévez E., Un Modelo Unificado de Visualización, Proceedings, pp. 881-892, 9° Congreso Argentino de Ciencias de la Computación. 6 al 10 de Octubre de 2003. La Plata. Argentina.

[5] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman. Do you see what I mean? IEEE. Computer Graphics and Applications, 25(3):6–9, 2005

[6] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information, and knowledge in visualization. IEEE Comput. Graph. Appl., 29(1):12–19, 2009.

[7] Barnes, S. J.; Roeder, P.L. 2001. The Range of Spinel Compositions in Terrestrial Mafic and Ultramafic Rocks. Journal of Petrology, vol. 42, number 12, pp: 2279-230

[8] Tergan, S. O., & Keller, T. (Eds.). (2005). *Knowledge and information visualization: Searching for synergies* (Vol. 3426). Springer.

[9] Ganuza, M. L., Ferracutti, G., Gargiulo, M. F., Castro, S. M., Bjerg, E., Gröller, E., & Matković, K. (2014). The spinel explorer—interactive visual analysis of spinel group minerals. *IEEE transactions on visualization and computer graphics*, 20(12), 1913-1922..

[10] Ganuza, M. L., Gargiulo, F., Ferracutti, G., Castro, S., Bjerg, E., Gröller, E., & Matković, K. (2015, October). Interactive semi-automatic categorization for spinel group minerals. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on* (pp. 197-198). IEEE.

[11] Ferracutti, G. R., Gargiulo, M. F., Ganuza, M. L., Bjerg, E. A., & Castro, S. M. (2015). Determination of the spinel group end-members based on electron microprobe analyses. *Mineralogy and Petrology*, 109(2),

153-160.

[12] Ganuza, M. L., Ferracutti, G., Gargiulo, M. F., Castro, S. M., Bjerg, E., Gröller, E., & Matković, K. (2014). The spinel explorer—interactive visual analysis of spinel group minerals. *IEEE transactions on visualization and computer graphics*, 20(12), 1913-1922.

[13] Immersive Analytics for Geology: Field Sketch-Like Visualization to assist geological structure analysis during fieldwork. IEEE VIS - Workshop on Immersive Analytics. N. F. Gazcón, J. M. Trippel, J. I. Larregui, M. L. Ganuza, E. A. Bjerg y S. M. Castro. Phoenix, USA.

[14] Interactive Visual Categorization of Spinel-Group Minerals. 33rd Spring Conference on Computer Graphics (SCCG 2017). M. L. Ganuza et al. Mikulov, República Checa.

[15] S. Escarza, S. Castro, S. Martig. Formalización del Proceso de Visualización mediante el uso de Ontologías. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010) pp. 261-265. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[16] M. Larrea, S. Martig and S. Castro. Semantics-based Color Assignment in Visualization. *Journal of Computer Science & Technology*. Vol. 10 - No. 1 – April 2010 - ISSN 1666-6038.

[17] M. Larrea, S. Martig and S. Castro. Formalización del Proceso de Visualización Basada en Semántica. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010) pp. 270-274. Universidad Nacional de la Patagonia San Juan Bosco. 5 y 6 de mayo de 2010. El Calafate. Santa Cruz. Argentina. ISBN: 978-950-34-0652-6.

[18] Escarza, S., Larrea, M. L., Urribarri, D. K., Castro, S. M., & Martig, S. R. (2011).

Integrating semantics into the visualization process. In *Dagstuhl Follow-Ups* (Vol. 2). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[19] Larrea, M. L., Escarza, S., Urribarri, D. K., Castro, S. M., Herrera, A. E., Escobal Blanco, C. A., & Cerúsico, G. (2014, October). Ontologías y Semántica en el Proceso de Visualización. In *XVI Workshop de Investigadores en Ciencias de la Computación*.

[20] Larrea, M. L. (2017). Black-Box Testing Technique for Information Visualization. Sequencing Constraints with Low-Level Interactions. *Journal of Computer Science & Technology*, 17.

[21] Larrea, M., Castro, S., Urribarri, D., & Escarza, S. Visualización Basada en Semántica. *WICC 2017*, 418.

[22] Urribarri, D. K. (2015, August). Predicción del desempeño de las técnicas de visualización a partir de métricas sobre los datos. In *XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015)*

[23] Urribarri, D. K., & Castro, S. M. (2017). Prediction of data visibility in two-dimensional scatterplots. *Information Visualization*, 16(2), 113-125.