Semantic Based Visualization: A first approach

Martín Larrea, Sergio Martig, Silvia Castro
Departamento de Ciencias e Ingeniería de la Computación
Laboratorio de Investigación en Visualización y Computación Gráfica - Universidad Nacional del Sur
{mll, srm, smc}@cs.uns.edu.ar
Tel. 54-291-4595135 Fax. 54-291-4595136
Bahía Blanca, CP 8000, Argentina

ABSTRACT

Visualization is the process of mapping data into visual dimensions to create a visual representation to amplify cognition. Visual representations are essential aids to human cognitive tasks and are valued to the extent that they provide stable and external reference points upon which dynamic activities and thought processes may be calibrated and upon which models and theories can be tested and confirmed. The active use and manipulation of visual representations makes many complex and intensive cognitive tasks feasible. A visual representation is able to convey relationships among many elements in parallel and provides an individual with directly observable memory. A successful visualization allows the user to gain insight into the data, in other words to communicate different aspect of the data in an effective way. Even with today's visualization systems that give the user a considerable control over the visualization process, it can be difficult to produce an effective visualization. To obtain useful results, a user had to know which questions to pose. Problems had to be framed in very precise terms. A strategy to improve this situation is to guide the user in the selection of the parameters involved in the visualization. Our research goal is the design of a visualization system that assist the user to do the work, by considering the semantic of the data together with the semantic of the stages through all the visualization process.
Keywords: Semantic, Visualization pipeline.

1. INTRODUCTON

The visualization challenge is to find a visual metaphor that the user can understand and perceive effectively [1] [2] [3], and to provide interaction methods [4] that make it possible for the user to work with and probe the data as effectively and effortlessly as possible. Computer technology allows the exploration of big information resources. Huge amount of data are becoming available on networked information systems, ranging from unstructured and multimedia documents to structured data stored in databases. On one side, this is extremely useful and exciting. On the other side, the ever growing amount of available information generates cognitive overload and even anxiety, especially in novice or occasional users. While computational power has increased exponentially, the ability to interact with useful information has only increased incrementally. In recent decades, the exponential increase in computing power has allowed many more questions to be posed and more complex problems to be addressed. Information is now massive, disparate, and disorganized. The dimensionality of data has also increased, requiring greater effort to identify and comprehend relationships relevant to a particular analytic task.

Nowadays, a wide diversity of user access, extract, and display information that is distributed on various sources, which differ in type, form and content. In many cases the users have an active control over the visualization process but even then it is difficult to achieve an effective visualization. For example, since the goal of visualization is to provide a representation which helps them to interpret their data or to communicate meaning, it is important that the mapping from physical to perceptual dimensions be under control. A strategy to improve this situation is to guide the user in the selection of the different parameters involved in the visualization. The Visualization field has matured substantially during the last decade; new techniques have appeared for different data types in many domains. With the use of visualization becoming more generalized, a formal understanding of the visualization process is needed [15] [16].

2. PREVIOUS WORK

2.1 RULE-BASED ARCHICTURE EXAMPLE

PRAVDA (Perceptual Rule-Based Architecture for Visualizing Data Accurately) [5] is a rule based architecture for assisting the user in making choices of visualization color parameters. This architecture provides sets of appropriates choices for visualization based on a set of underlying rules [6] [7] which are used to constrain operations *i.e.*, selecting a colormap. Rules incorporate information about data, which is call metadata, such as minimum, maximum, or spatial frequency, and also information supplied by the user. This architecture also provides for linkages between rules that control different visualization operations, with a choice of parameters for one operation constraining choices that are available for others. For example, if the user selects a colormap, that information is fed back to the operation for selecting contour lines, where rules constrain the parameters of the contour lines depending on which colormap has been selected. Hence, if the contour lines are superimposed over a dark region, as defined by the colormap, legibility rules would constrain the set of color choices to those offering sufficient luminance contrasts to be detectable. This network of linked operations guides the user through the complex design space of visualization operations. The key element in this rule based architecture is the use of metadata; system provided metadata, as data type, data range, metadata computed by algorithm, as spatial frequency, and metadata provided by the user. These metadata would, for example, represent the dynamic range of the data or the geometric relationships between objects in the scene.

2.2 SEMANTIC IN THE VISUALIZATON

The papers [17] [18] [19] [20] and [21] are good examples of how semantic information is integrated into visualization tasks. However in all these examples the role of the semantic is to improve the integration, querying and description of the data in the visualization; in none of these cases the semantic associated with the data is use to create the visualization. Only in [14] we can find a first approach to the use of semantic as an aid to create the visualization. The worked done define a customizable representation model which allows the biologist to change the graphical semantics associated to the data semantics. The representation models are base on an XML implementation; such models are based on an XML Schema definition that prescribes the correctness of the model and provides validation features. Unfortunately this work is only intended for biological use; does not take advantage of the RDF or OWL representation and doesn't include any reasoning process with the semantic information.

3. VISUALIZATON PROCESS

The different visualization models presented in the last years cover partially the aspects of the exploration process; Upson [8] and Card [9] models give an overview of the visualization process but do not offer enough details for the user exploration. Chi model [10] does not describe properly the interactions and Chuah and Roth model [11], presents a detailed definition of the interactions, but does not seem to be enough to cover all the possible applications. In order to overcame these problem we have developed a model that represents all the visualization process stages and the interactions between them and the user. The "Unified Visualization Model" [12] was developed to create an unified conceptual framework, independent from the data domain. This model takes under consideration the characteristics of all visualization areas. The unified model focuses on the visualization processes as well as in the data stages. In this model, the user's interactions play a central point, because is the user who interacts with the visualization and, based on his/her interpretations of the representation, modifies the image to steer the calculation, remap the data representation in order to better understand its structure, or create a visualization which highlights a particular feature.

This model is represented by stages along a flow, the flow represent the transformations of data. Each stage is a data stage and the edges are the transformations to move from one stage to the next. The unified model considers five stages and four transformations. The transformations and the stages along this flow reflect the user interaction on the visualization process. We present now a brief description of the stages and transformation in the Unified Visualization Model.

The "Unified Visualization Model"	
Stages	**Transformations**
Stage "Raw data" Data from the application domain.	**Transformation "Raw data to Abstract Data"** This transformation allows the user to select the data he/she wants to visualize. After the selection, the data moves from the data domain representation to an inner and manageable structure.
Stage "Abstract Data" Data to be potentially visualize by the user. Besides this data the user also has the metadata created in the previous transformation.	**Transformation "Abstract data to Data to be Visualize"** From the "Abstract data" stage the user will select all the data that will be visualized.
Stage "Data to be Visualize" Data that will be visualized. It can be a subset of the "Abstract data"	**Transformation "Visual Mapping"** This transformation allows the user to specify how he/she wants to visualize all the data in the previous stage. All the necessary structures to support the spatial substrate, the visual elements and their attributes are created from this transformation.
Stage "Visual Mapped Data" Data to be visualized along with all the necessary information for its visual representation.	**Transformation "Visualization Transformation"** This transformation allows the creation on screen of all the data in the "Visual Mapped Data". This will usually include the application of some visualization technique that supports all the restrictions imposed in the "Visual Mapping" transformation.
Stage "Visualize data" This is the result from the visualization process. This is the starting point for the user to begin his/her visual exploration and navigation process.	

4. OUR GOAL

The user is an active participant in the visualization process, and the goal of visualization is to present data in a way which helps him/her identify trends, features and patterns, generate hypotheses, and assign meaning to visual information on screen. Our goal is to develop a visualization model that considers the semantic of the data and of the different stages in the visualization process. This model will transform data into information; according to Keller and Tergan [13], "information is data that has been given meaning through interpretation by way of relational connection and pragmatic context". The information is the same only for those people who attribute to it the same meaning. This 'meaning' can be useful, but does not have to be. Information may be distinguished according to different categories concerning, for instance, its features, origin and relations. By making these considerations, the visualization process will be able to determinate the characteristics of an effective visualization guiding the user through the different stages. The metadata will define a higher level characterization of the data which provides a higher level interface to the user, and a higher level input to visualization rules. All the data from the different application domains will be categorized according to [9].

At present, we are surveying the visualization techniques and the different data models and interactions involved. For each technique we will study its interactions under representative application domains. All these techniques will be analyzed in the context of the "Unified Visualization Model" [12]. Taking all this into account we will begin to define the semantic of the stages involved in the visualization process. Our goal is to define an unified semantic for the data model and the process involved. We have concluded that the first stage of the “Unified Visualization Model”, the raw data, will include an XML representation of the input data and with this the associated semantic; both RDF and OWL are being consider for the XML representation. All the final and intermediate results will be published.

This work is in progress at the “Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica”, Computers Sciences and Engineering Department, Universidad Nacional del Sur. This work is close related with the next research projects:

- “Representaciones Visuales e Interacciones para el Análisis de Grandes Conjuntos de Datos (24/N02015)”. Directora: Dra. Silvia Castro.
-
- “Herramientas de Visualización para la exploración de Datos (24/ZN12)”. Director: Sergio Martig.

- “Desarrollo de Herramientas Inteligentes para la Web Semántica (PICT año 2003 Nro 15043)”. Director: Guillermo Simari.

- “Sistemas Inteligentes para apoyo a los Procesos Productivos”, Subproyecto Servicios de WEB e Inteligencia en la WEB, (PAV año 2003 Nro. 00076).

In conclusion we consider that a visualization process model with its proper interactions is not enough to assure an effective visualization. To achieve this, a meta-data model for the visualization process, visualization stages, data and interactions also need to be developed.

4. ONTOLOGY OF COLORMAP

Our first step into the creation of a semantic based visualization is the definition of an ontology for colormap selection. This ontology will include the concepts of color, transparency, colormap, and internationalization. We will migrate the taxonomy presented in [5] into an ontology and enhanced with color and internationalization information. The work done in [5] gave us color properties base on the data type and its spatial frequency, i.e. luminance, hue, saturation; we will introduce the concept of color and how it is related to the data and its semantic into the ontology as well.

For example, let’s take a data set that represent test scores; the semantic may stated that this information is between 1 and 10, it is a ordinal data type, below 4 is bad and above 5 is good, and we know that this visualization is taking place in Argentina. In this case, the system, through the ontology, could establish that variations of green could be use for the “good” values and variations of red for the “bad” ones; because is an ordinal data type the color map will not be continuous. The concepts of “good” and “bad” will be part of the ontology and will be associated with the colors green and red respectively under a specific internationalization i.e. “West Culture”.

5. ACKNOWLEDGMENTS

This work was partially supported by the PGI 24/N020 and 24/ZN12, Secretaría de Ciencia y Tecnología of the Universidad Nacional del Sur (SECyT, UNS) and the PICT 2003 Nº 15043, Agencia Nacional de Promoción Científica y Tecnológica.

6. REFERENCES

[1] Bertin, J. “Semiology of Graphics”, 1967, reprinted by University of Wisconsin Press, Madison, WI, 1983.

[2] Tufte, E. “The Visual Display of Quantitative Information”. Connecticut: Graphics Press, 1983.

[3] Cleveland, W.S. “The Elements of Graphing Data”, Monterey, California: Wadsworth, Inc., 1991.

[4] Baecker, R. M. and Buxton, W. A. S., “Readings in Human-Computer Interaction”, San Mateo CA., Morgan Kaufmann Publishers, 1995.
[5] Bergman, L. D., Rogowitz, E. B., Treinish, A. L., “A Rule-based Tool for Assisting Colormap Selection”. IBM Thomas J. Watson Research Center, Yorktown Heights, NY.
[6] Rogowitz, E. B., Treinish, A. L., “Data Structures and Perceptual Structures”. SPIE – The International Society for Optical Engineering. February 1993. Volume 1913.
[7] Rogowitz, E. B., Treinish, A. L., “Using Perceptual Rules in Interactive Visualization”. SPIE – The International Society for Optical Engineering. February 1993. Volume 2179.
[8] Upson, C., Faulhaber, T., Kamins, D., Laidlaw, D., Schlegel, D., Vroom, J., Gurwitz, R., van Dam, A., “The application visualization system: A computational environment for scientific visualization”. CG&A, IEEE, 9(4):30–42, 1989.
[9] Card, S., Mackinlay, J., Shneiderman, B., “Readings in Information Visualization – Using Vision to Think”, Morgan Kaufmann, 1999.
[10] Chi,E.; “A Taxonomy of Visualizations Techniques using the data state reference model”, InfoVis 2000, pp. 69-75. IEEE Publishers, 2000.
[11] Chuah, M.C., Roth, S.F., “On the Semantics of Interactive Visualizations”, Proceedings of the INFOVis 96, pp. 29-36. IEEE Publishers, 1996.
[12] Martig, S., Castro,S., Fillottrani, P. & Estévez, E., “Un Modelo Unificado de Visualización”. Proceedings, pp. 881-892, 9º Congreso Argentino de Ciencias de la Computación. Oct.2003. Argentina.
[13] Keller T, Tergan S-O. Visualizing knowledge and information: an introduction. In: Tergan S-O and Keller T (Eds). Knowledge and Information Visualization: Searching for Synergies. Lecture Notes in Computer Science LNCS 3426. Springer: Berlin, Heidelberg, New York, 2005; 1-23.
[14] Semantics of a networked world: Semantic Visualization of Biochemical Databases (Paris, 17-19 June 2004, revised selected papers) ICSNW 2004 : international IFIP conference N°1, Paris , FRANCE (17/06/2004) 20041973, vol.3226, pp.199-214. ISBN 3-540-23609-0.
[15] Brodlie, K. W., Duce, D. A., Duke, D. J. and al., e. Visualization Ontologies: Report of a Workshop held at the National e-Science Centre, Report e-Science Institute, 2004.
[16] D. J. Duke , K. W. Brodlie , D. A. Duce, Building an Ontology of Visualization, Proceedings of the conference on Visualization '04, p.598.7, October 10-15, 2004
[17] Zhao Xu, Huajun Chen, Zhaohui Wu: Applying Semantic Web Technologies for Geodata Integration and Visualization. ER (Workshops) 2005: 320-329
[18] Richard Potter, Helen Wright: An Ontological Approach to Visualization Resource Management. DSV-IS 2006: 151-156
[19] Kalogerakis, E., Christodoulakis, S., and Moumoutzis, N. 2006. Coupling Ontologies with Graphics Content for Knowledge Driven Visualization. In *Proceedings of the IEEE Conference on Virtual Reality* (March 25 - 29, 2006). VR. IEEE Computer Society, Washington, DC, 6
[20] Nguyen, T. D., Ho, T. B., and Nguyen, D. 2002. Data and Knowledge Visualization in Knowledge Discovery Process. In *Proceedings of the 5th international Conference on Recent Advances in Visual information Systems* (March 11 - 13, 2002). S. Chang, Z. Chen, and S. Lee, Eds. Lecture Notes In Computer Science, vol. 2314. Springer-Verlag, London, 311-322
[21] Weng, Z. M. and Bell, D. 1998. Integrating Visual Ontologies and Wavelets for Image Content Retrieval. In *Proceedings of the 9th international Workshop on Database and Expert Systems Applications* (August 26 - 28, 1998). DEXA. IEEE Computer Society, Washington, DC, 379.