Semantics-based Visualization

Martín Larrea[1,2], Silvia Castro[2]

[1]Consejo Nacional de Investigaciones
Científicas y Técnicas (CONICET)
[2]Laboratorio de Investigación y Desarrollo en Visualización y
Computación Gráfica, (VyGLab)
Departamento de Ciencias e Ingeniería de la Computación,
Universidad Nacional del Sur,
Av. Alem 1253, Bahía Blanca, Argentina.
{mll,smc}@cs.uns.edu.ar

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. They 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, i.e. to communicate di erent aspect of the data in an e ective way. Even with today's visualization systems that give the user a considerable control over the visualization process, it can be di cult to produce an e ective 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 in the construction of the visualization, by considering the semantic of the data, the semantic of the stages through all the visualization process and the semantic of the external elements a ecting the visualization.

Keywords: Semantic, Visualization, Semantics-based Visualization, Ontology, Reasoning, RDF, OWL

1 Problem

The visualization challenge is to find a visual metaphor that the user can understand and perceive e ectively ([1], [2], [3]) combined with the interaction methods ([4]) that make it possible for the user to work with and probe the data as e ectively and e ortlessly 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 linearly. 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 e ort to identify and comprehend relationships relevant to a particular analytic task. Nowadays, a wide diversity of users access, extract, and display information that is distributed on various sources, which di er in type, form and content. In many cases the users have an active control over the visualization process but even then it is di cult to achieve an e ective visualization. For example, since the goal of visualization is to provide a representation, along with its associated interactions, 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 di erent parameters involved in the visualization. The Visualization field has matured substantially during the last decades; new techniques have appeared for di erent data types in many domains. With the use of visualization becoming more generalized, a formal understanding of the visualization process is needed ([5]).

Our main goal is the development of a visualization model that considers the semantics of both the data and the di erent stages in the visualization process. This model will transform data into information; according to Keller and Tergan ([6]), information is data that has been given meaning through interpretation by way of relational connection and pragmatic context . This *meaning* can be useful. Information may be distinguished according to di erent categories concerning, for instance, its features, origin and relationships. By making these considerations, the visualization process will be able to determine the characteristics of an e ective visualization and guide the user through the di erent stages. The user is an active participant in the visualization process and the goal of a visualization is to present data in a way that helps him to identify trends, features and patterns, generate hypotheses, and assign meaning to the visual information on the screen.

Since 2006 we have been working on the integration of semantic information into the visualization process ([7], [8]) and our main goal is to define an unified semantics for the data model and the process involved.

2 Related Work

In order to understand the challenges of our enterprise and to know the state of the art in the integration of semantic into the visualization process, we began our work by searching for examples of visualizations-related work with semantic information. The first subsection gives a brief overview of several works that used semantic information with visualizations. On the the next subsection we review a ruled-based architecture which used semantic information, in the form of metadata, to create colormaps.

2.1 Semantic in the Visualization

The papers [11], [12], [13], [14] and [15] 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 any of these cases the semantic associated with the data is used to create the visualization. Only in [16] there is a first approach to the use of semantic as an aid to create the visualization. This work 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. This is only a first approach to a system similar to the one we proposed. The main di erences between this work ([16]) and our proposal are the lack of any support for a reasoning process, the use of XML instead of RDF or OWL and the limitations in the application domain, the system is only intended for biological use.

2.2 Rule-Based Architecture

PRAVDA (Perceptual Rule-Based Architecture for Visualizing Data Accurately) ([10]) 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 which are used to constrain operations i.e., selecting a colormap. Rules incorporate information about data, that is metadata, such as minimum, maximum, spatial frequency, among others and also information supplied by the user. This architecture also provides for dependency between rules that control di erent visualization operations. 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 o ering su cient luminance contrasts to be detectable. This network of linked operations help guide 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.

3 Framework

This thesis began with a survey on visualization models, automatic visualization process and visualization taxonomies. While researchers in the area have presented di erent visualization models, they were usually focused on one area of visualization in particular. Examples of these previous work are W. Schroeder's pipeline for scientific visualization ([17]), the pipeline presented by S. Card for information visualization ([18]), the data state model introduced by E. Chi ([19]) and the Unified Visualization Model (UVM) created by Martig et. al ([20]). The UVM is applicable to any particular field and consists of a single model that allows the user to focus on both the processes and the data's states. Our work is based on the UVM, we seek to extend and improve this model by the use of semantic information.

The second survey conducted on this thesis was related to the automatization of the visualization process. The first record of such process was the work done by Mackinlay J. in 1986 ([21]). He established the goal of developing an application for automating the creation of graphic representations. According to him, the user may not have the necessary knowledge to create an e ective visual representation. Mackinlay based his work on the idea that a visual representation is an expression in a graphical language, where this language is formally defined. While Mackinlay set limitations on the scope of his work is the first benchmark for the creation of an automatic visualization. His work set the ground for all researchers in the Visualization area.

Our final survey was conducted on visualization related taxonomies. Every concept evolves through three stages of formality, terminology, taxonomy and ontology, the latter being the more formal. This is why visualization taxonomies are so important for our work. As our goal is the development of ontologies to support semantic and reasoning in the visualization pipeline, we can use predefined taxonomies and take the next step in formality. Without doubt, one of the most relevant taxonomies in the area of Visualization and Graphics is the one presented by Jacques Bertin in Semiology of Graphics ([1]). However this taxonomy requieres a deeper analysis and the inclusion of newest elements such us 3D visualization techniques and interaction classifications.

During the last two years we have been developing several ontologies and prototypes that use, on di erent degrees, semantic information in the visualization process. In all the prototypes we used an application called Brows.AR. We developed Brows.AR as an application for the visualization of file hierarchies in 3D, based on the Spherical Layout ([9]). We are currently working on the architecture to support the integration of semantic information and reasoning process.

4 Challenges

A successful visualization allows the user to gain insight into the data. A successful visualization process takes advantage of the structure and the meaning of the data to create the most e ective visualization. The structure of the data can be obtained from the data itself but not its meaning. Two sets may contain the same data, but if its meaning is di erent then the final visualizations will not necessary be the same. This is why we included the semantic about the data, a way to describe the data about the data.

A visualization is greatly a ected by what the user want to do with it. For the same data set, also with the same meaning, certain visualization may be most suitable for data exploration and another may be better for data comparison. By knowing what the user want to do and its meaning the visualization designer can create a better result. This is our motivation to incorporate the semantic about the tasks.

Additionally, the response time of the interactions is crucial to obtain an e ective visualization. If the user want to explore a 3D visual representation but there is no dedicated GPU on the computer, the user's experience would be negatively a ected. Besides that, a 4 inches screen can not represent a visualization in the same way that a 42 inches screen does. A formal description of the system's hardware could help the visualization designer to enhance the user experience with the visualization. Then, in addition to the data and task semantics, we also included the semantic of the hardware, a description of the actual system's hardware.

All the mentioned semantics should be taken as input to the visualization process. But the visualization process can contain its own semantics as previous knowledge embedded in the system. We are working on several ontologies to describe color coding, color matching, visualization techniques, etc. The goal is to help the user in the decisions that depend on knowledge outside of the user scope. For instance, which colors combine better or which colormap to use to represent a data attribute.

5 Future Work

We have shown that it is possible to assist the user in the creation of a visualization through the use of semantic information. We must still determine the degree of e ectiveness of such visualizations. We have to establish e ectiveness metrics and design the test to evaluate them. The results from these evaluations will be used as feedback for the defined semantics. Many research fields were open through this work. Some of them are, the semantic of the user, the semantic of the tasks, the semantic of the techniques and the semantic of the colores. The semantic of the techniques has been the topic proposed to the CONICET to continue our research through a postdoc grant.

References

1. Bertin, J.: Semiology of Grahics. University of Wisconsin Press (1983).
2. Tufle, E., R.: The Visual Display of Quantitative Information. Graphics Press (1986).

3. Cleveland, S., W.: The Elements of Graphing Data. Wadsworth Publ. Co. (1985).
4. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction. Prentice-Hall, Inc. (1997).
5. Duke, D., J., Brodlie, K., W., Duce, A., D.: Building an Ontology of Visualization. In: Proceedings of the Conference on Visualization '04. pp. 598. IEEE Computer Society (2004)
6. Keller, T., Tergan, S.: Visualizing Knowledge and Information: An Introdution. In: Knowledge and Information Visualization. pp. 1-23, Springer (2005)
7. Larrea, M., Martig, S., Castro, S.: Visualizacin Basada en Semntica. In: Proceedings of the XII Workshop de Investigadores en Ciecnias de la Computacón. pp. 270-274. Red UNCI (2010).
8. Larrea, M., Martig, S., Castro, S.: Semantics-based Color Assignment in Visualization. In: Journal of Computer Science and Technology. pp. 14-18, ISTEC (2010)
9. Larrea, M., Martig S., Castro S.: Spherical Layout - Layout for 3D Tree Visualization. In: Proceedings of the IADIS 2007, Multi Conference on Computer Science and Information Systems. pp. 91-98 (2007).
10. Bergman, L., D., Rogowitz, B., E., Treinish, L., A.: A rule-based tool for assisting colormap selection. VIS '95: Proceedings of the 6th conference on Visualization '95 (1995).
11. Zhao, X., Huajun, C., Zhaohui, W.: Applying Semantic Web Technologies for Geodata Integration and Visualization. Lecture Notes in Computer Science (2005).
12. Kashyap, V., et. al: An Ontology-based Approach for Data Integration - An Application in Biomedical Research. Real-world Applications of Semantic Web Technology and Ontologies (2008).
13. Kalogerakis, E., Christodoulakis, S., Moumoutzis, N.: Coupling Ontologies with Graphics Content for Knowledge Driven Visualization. VR '06: Proceedings of the IEEE conference on Virtual Reality (2006).
14. Nguyen, T., D., and Ho, T., B., Nguyen, D.: Data and Knowledge Visualization in Knowledge Discovery Process. VISUAL '02: Proceedings of the 5th International Conference on Recent Advances in Visual Information Systems (2002).
15. Weng, Z., M., Bell, D.: Integrating Visual Ontologies and Wavelets for Image Content Retrieval. DEXA '98: Proceedings of the 9th International Workshop on Database and Expert Systems Applications (1998).
16. Bouzeghoub, M., Goble, C., Kashyap, V., Spaccapietra, S.: Semantics for Grid Databases. First International IFIP Conference on Semantics of a Networked World (2004).
17. Kitware Inc.: The Visualization Toolkit, Third Edition (2003).
18. Card, S., K., Mackinlay, J.: The structure of the information visualization design space. Proceedings of the 1997 IEEE Symposium on Information Visualization (1997).
19. Chi, E., Riedl, J.: An Operator Interaction Framework for Visualization Systems. Proceedings of the 1998 IEEE Symposium on Information Visualization (1998).
20. Martig, S., Castro, S., Fillottrani, P., Estevez, E.: Un Modelo Unificado de Visualizacin. Proceedings 9 Congreso Argentino de Ciencias de la Computacin (2003).
21. Mackinlay, J.: Automating the design of graphical presentations of relational information. ACM Trans. Graph. (1986).