Visualización en la Web Semántica

Sergio Martig Silvia Castro

Departamento de Ciencias e Ingeniería de la Computación
Laboratorio de Investigación en Visualización y Computación Gráfica
Instituto de Investigación en Ciencia y Tecnología Informática (IICyTI)
Universidad Nacional del Sur
{srm,smc}@cs.uns.edu.ar

Introducción

Estamos en la puerta de una nueva generación de la Web. Si bien la Web fue diseñada originalmente para que la información en la misma sea para consumo humano, se pretende que la nueva generación de la Web, denominada Web Semántica, provea datos que sean procesables por las máquinas. Esto permitirá un gran rango de servicios inteligentes, agentes de búsqueda, filtros de información, etc; éste es un proceso al que Berners-Lee describe como "Bringing the Web to its full potential".

La Web Semántica es una extensión de la Web actual basada en la idea de intercambiar información con descripciones explícitas, formales y accesibles por las máquinas. Si bien es sumamente difícil automatizar toda la información en la Web, la solución propuesta es proveer información adicional para describir los datos. El proveer información con semántica permitirá la construcción de aplicaciones que tengan conocimiento creciente de qué está contenido en la información que procesan y qué pueden operar, por lo tanto, más exactamente. Esto tiene el potencial de mejorar la forma en la que tratamos con la información en el sentido más amplio posible, es decir, mejores motores de búsqueda, agentes móviles para distintas tareas o aplicaciones que aún no se conocen. Para poder mejorar el intercambio de semántica de información, uno primero necesita coincidir en cómo modelarla explícitamente. Las ontologías son un mecanismo clave para representar descripciones de dominios formales y compartidos. Pueden usarse para anotar datos con etiquetas (metadatos) indicando su significado, es decir haciendo explícita su semántica y accesible por la máquina.

Las ontologías son metadatos que proveen un vocabulario controlado, con cada término teniendo una semántica definida explícitamente y procesable por la máquina. De este modo las ontologías ayudarán tanto a las personas como a las máquinas a comunicarse más efectivamente. Tendrán, por lo tanto, un rol crucial en permitir que el acceso a la Web sea basado en su contenido, interoperabilidad y comunicación. Las ontologías sobre la Web Semántica serán cruciales para el desarrollo de aplicaciones Web tales como el e-commerce, el soporte de agentes inteligentes, etc. Las ontologías pueden usarse para diferentes propósitos como el intercambio de datos, la consulta de datos, la visualización de datos, etc.

El subproyecto de *Visualización en la Web Semántica* está embebido en un proyecto general cuyo objetivo es el desarrollo de Herramientas Inteligentes para la Web Semántica. El Grupo Responsable del Proyecto está compuesto por cuatro Laboratorios de Investigación y Desarrollo del Departamento de Ciencias e Ingeniería de la Computación de la UNS: *Inteligencia Artificial, Visualización y Computación Gráfica, Sistemas Distribuidos e Ingeniería de Software*.

Las capacidades avanzadas que se esperan desarrollar en las aplicaciones para la Web Semántica requieren, sin duda, el desarrollo de herramientas de propósito general para la creación y

mantenimiento de ontologías que utilicen técnicas avanzadas de Visualización. Éste es el objetivo de este subproyecto, del cual son responsables los integrantes del *Laboratorio de Visualización y Computación Gráfica* .

Contexto de la Investigación

El acceso a la información diseminada en el mundo aumenta día a día ya que la Web es el más amplio repositorio de información existente. Sin embargo encontrar, clasificar y organizar la información de interés son tareas sumamente duras. En lo que se refiere a la Web Semántica, tanto para los diseñadores como para los desarrolladores es importante la visualización de la estructura de determinada información formalmente definida. Esto motiva la necesidad urgente de técnicas y herramientas de visualización en distintas comunidades, tanto científicas como no científicas.

La Visualización de Información es una forma muy poderosa para comunicar efectivamente información al usuario. Reconociendo que la WWW es un grafo sumamente gigante de nodos y arcos, cada uno de los cuales puede tener además atributos adicionales, podemos imaginarnos varias formas de mostrar y navegar las distintas subestructuras de la Web para asistir al usuario en su descubrimiento exploratorio. Un grafo que muestra la relación de una home page a todas sus página subsidiarias, por ejemplo, le dará al usuario información útil del mapa a todas estas páginas; sin embargo, la insuficiencia de contenido semántico explícito (en HTML, por ejemplo) hace bastante dificultoso crear mapas con determinado significado.

La Visualización de Información pretende producir representaciones gráficas de estructuras abstractas de información, en tanto la Web Semántica pretende contar con un marco universal de recursos que pueda ser utilizado por los agentes de software. Por otro lado, la Web Semántica y la Visualización de Información pueden complementarse entre sí en varias cuestiones fundamentales concernientes a la organización, a la clasificación y al acceso a recursos de información en gran escala. La Web Semántica enfatiza que los datos deben ser entendidos por la máquina en tanto que la Visualización de Información pretende maximizar nuestras habilidades perceptuales y cognitivas para comprender la información a partir de los insights obtenidos de las representaciones visuales de estas estructuras de información abstractas.

Nosotros estamos básicamente interesados en la visualización de ontologías para luego aplicar esto al proceso de generar visualizaciones de recursos Web (Figura 1). Las páginas Web serán nuestros datos crudos y luego de clasificarse en una ontología predefinida, que juega el rol de estructura semántica, ésta será la información a partir de la cual se deberá generar la presentación visual. Deben encontrarse representaciones visuales adecuadas para este dataset que ha sido generado a partir de una dada ontología.

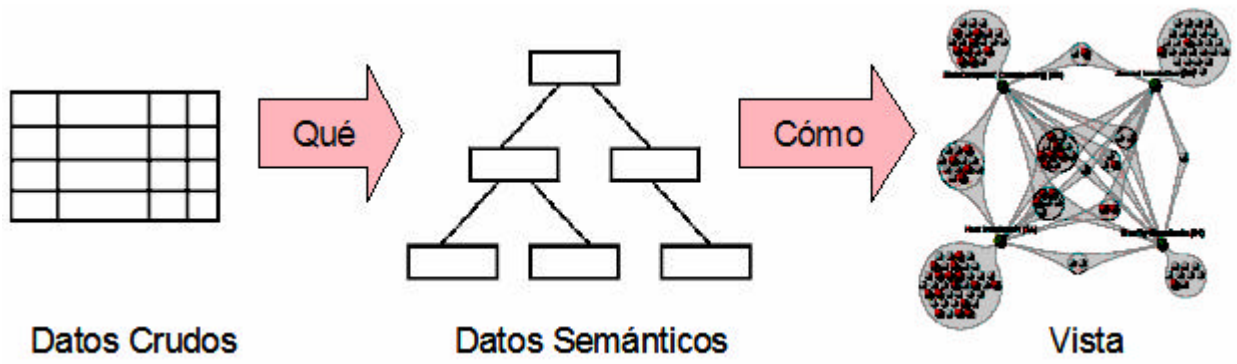

Figura 1

¿Qué debe hacerse para visualizar estos grandes espacios de información? Debido a que los espacios de información son tan grandes, hay distintos elementos a tener en cuenta en la creación de abstracciones visuales que comuniquen la información y las relaciones apropiadas. Una forma razonable para llevar a cabo esta tarea es comenzar con diagramas basados en grafos; pero los grafos no son la única metáfora visual para la representación y la navegación de grandes espacios de información: hay otros conceptos explorados en este sentido como mapas, diseño de libros, etc. En este sentido consideramos que se debe desarrollar una amplia variedad de formas de presentar y manipular esta información para poder experimentar con ella.

En el caso de resultados de búsquedas, por ejemplo, el presentar al usuario una lista ordenada de documentos basados en alguna noción de relevancia no lo ayuda mucho a entender mejor el espacio altamente dimensional dentro del que se define su relevancia. En este caso resulta evidente por qué deben estudiarse las ventajas y desventajas de las visualizaciones existentes integrándolas a los motores de búsqueda de la Web para mejorar la imaginación de los usuarios y ayudarlos a crear nuevas herramientas. También se ve la necesidad de tratar con metáforas visuales para presentar, entender, organizar y navegar ricos espacios de información.

Objetivos del Proyecto

El objetivo del proyecto general en el que este subproyecto está inmerso, es el desarrollo de Herramientas Inteligentes para la Web Semántica. Como es natural, estas herramientas respetarán los únicos estándares generales vigentes, representados por las recomendaciones del W3C (*World Wide Web Consortium*).

Las capacidades avanzadas que se esperan desarrollar en las aplicaciones para la Web Semántica requieren, sin duda, el desarrollo de herramientas de propósito general para la creación y mantenimiento de ontologías que utilicen técnicas avanzadas de Visualización. El enfoque que se adoptará integra teorías y técnicas relacionadas para poder lograr las transformaciones de las ontologías que preserven las estructuras semánticas más salientes. Cuando se debe capturar estructuras implícitas en una representación hipermedial, es clara la necesidad de contar con un enfoque visual inteligente e integrado. Se debe lograr un marco de diseño interactivo e integrado que, a partir de una representación de las estructuras semánticas latentes extraídas, las transforme en una representación visual adecuada. Estas representaciones visuales, no son un fin en sí mismas, sino que representan un medio de exploración. Las distintas representaciones visuales generadas deben constituirse en una herramienta netamente interactiva que asista y habilite al usuario a explorar el espacio semántico, detectando sus características, permitiendo modificar su contenido y generando las nuevas estructuras semánticas que la soporten.

En resumen, el objetivo de este subproyecto es el desarrollo de herramientas para la creación, mantenimiento y visualización de ontologías.

En el marco de investigación propuesto, se proyectan realizar las siguientes tareas:

1. Exploración y análisis de distintas metáforas de visualización que puedan ser apropiadas para la representación de ontologias en función de su aplicabilidad en el contexto de la Web Semántica. Se pretenden diseñar las interacciones necesarias en las representaciones visuales para la visualización de ontologías.
2. Análisis de las representaciones visuales alternativas y determinación de las interacciones requeridas por cada técnica para lograr una interacción efectiva con la Web Semántica.
3. Construcción de prototipos que brinden herramientas para la creación, mantenimiento y visualización de ontologías. Se pretenden obtener prototipos de herramientas que permitan

la creación, mantenimiento, visualización y explotación (a partir de la inferencia) de la infraestructura de soporte cognitivo de la Web Semántica.

Muchas de las tareas de investigación y desarrollo descriptas más arriba interactúan entre sí. Por lo tanto, la precedencia en la secuencia descripta es aproximada. La elaboración de los prototipos, y su ejercitación, permitirá realizar modificaciones y refinar los resultados previos.

Bibliografía

[BYRN99] Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Addison Wesley, 1999.

[BCK98] Bass, L., Clements, P., Kazman, R., *Software Architecture in Practice*, Addison Wesley, 1998.

[BS03] Bederson, B., Shneiderman, B., *The Craft of Information Visualization: Readings and Reflections*, Morgan Kaufmann Pub, 2003.

[BLHL01] Berners-Lee, T., Hendler, J., Lassila, O., *The Semantic Web.* Scientific American, volume 284, number 5, mayo 2001, págs.34-43.

[BL85] Brachman, R.J., Levesque, H. *Readings in Knowledge Representation.* Morgan Kaufmann, 1985.

[CAR80] M. J. Carpano, *Automatic Display of Hierarchized Graphs for Computer Aided Decision Analysis*, IEEE Trans. Syst. Man Cybern., SMC-10, no. 11, 1980.

[CKK02] Clements, P., Kazman, R., Klein, M., *Evaluating Software Architectures, Methods and Case Studies*, Addison Wesley, 2002.

[CMS99] Card, S., Mackinlay, J., Shneiderman, B., *Readings in Information Visualization – Using Vision to Think*, Morgan Kaufmann, 1999.

[DAML01a] DAML-S, http://www.daml.org/services, 2001.

[DCvH02] Dean, M., Connolly, D., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L., *OWL, Web Ontology Language 1.0 Reference W3C Working Draft*, 29 Julio de 2002. Disponible en http://www.w3.org/TR/owl-ref.

[DKW01] Dikel, D.; Kane, D., Wilson, J., *Software Architecture. Organizational Principles and Patterns*, Prentice Hall Inc., 2001.

[DFAB98] Dix, A., Finlay, J., Abowd, G., Beale, R., *Human-Computer Interaction*, Prentice Hall Europe, Second Edition, 1998.

[DIB99] G. Di Battista, P. Eades, R. Tamassia y I. Tollis, *Graph Drawing, Algorithms for the visualization of graphs*, Prentice Hall, 1999.

[FGW02] Fayyad, U., Grinstein, G., Wierse, A., *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Pub., 2002.

[GCh03] Fayyad, U., Grinstein, G., Wierse, A., *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Pub., 2002.

[GF02] Geroimenko, V.,, Chen, Ch., (Eds.) *Visualizing the Semantic WEB*, Springer Verlag, 2003, Second Print.

[KAM89] Kamada, T., *Visualizing Abstract Objects and Relations*. World Scientific Series in Computer Science, 1989.

[PUR97] Purchase, H., *Which aesthetics has greatest effects on human understanding,* Graph Drawing ´97, vol. 1353 of Lectures Notes in Computer Science, Springer-Verlag, 1997.

[Shn98] Shneiderman, B., *Designing the User Interface,* Addison-Wesley Pub. Company, 1998.

[Spe01] Spence, R., *Information Visualization,* Addison-Wesley, ACM Press, 2001.

[SUG81] K. Sugiyama, S. Tagawa y M. *Toda, Methods for Visual Understanding of Hierarchical Systems,* IEEE Trans. Syst. Man Cyber., SMC-11, no. 2, 1981.

[Tuf90] Tufte, E.R., *Envisioning Information,* Cheshire, CT Graphics Press, 1990.

[Tuf97] Tufte, E.R., *Visual Explanations: Images and Quantities, Evidence and Narrative,* Cheshire, CT Graphics Press, 1997.

[vHBFH01] van Harmelen, F., Broekstra, J., Fluit, Ch., et al. *Ontology-based Information* VSW'01, 2001.

[W3C] Documentos y recomendaciones del sitio http://www.w3.org (*World Wide Web Consortium).*