

Aplicación de técnicas de la web semántica

Pablo Castells
Escuela Politécnica Superior
Universidad Autónoma de Madrid
pablo.castells@uam.es

La web semántica es un área pujante en la confluencia de la Inteligencia Artificial y las tecnologías web, que propone nuevas técnicas y paradigmas para la representación del conocimiento que faciliten la localización, compartición e integración de recursos a través de la WWW [Berners-Lee 2001]. Estas nuevas técnicas se basan en la introducción de conocimiento semántico explícito que describa y/o estructure la información y servicios disponibles.

La web semántica mantiene los principios que han hecho un éxito de la web actual, como son los principios de descentralización, compartición, compatibilidad, o la apertura al crecimiento y uso no previstos de antemano. En este contexto un problema clave es alcanzar un entendimiento entre las partes: usuarios, desarrolladores y programas de muy diverso perfil. La web semántica rescata la noción de ontología del campo de la Inteligencia Artificial como vehículo para cumplir este objetivo [Gruber 1993].

Una ontología es una taxonomía de conceptos con atributos y relaciones, que proporciona un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas. Durante los últimos años se han desarrollado diversos lenguajes y estándares para la definición de ontologías, entre ellos XML [Bray 2000], RDF [Lasilla 1999], DAML+OIL [Connolly 2001], y más recientemente OWL [Dean 2002], respaldados por el consorcio W3C, uno de los principales promotores de la web semántica.

Entre los campos de aplicación donde las nuevas ideas de la web semántica pueden tener utilidad podemos citar:

- Comercio electrónico
- Gestión del conocimiento corporativo
- Búsqueda de información en la web
- Procesamiento del lenguaje natural
- Enseñanza
- Librerías digitales
- Turismo
- Patrimonio cultural

Durante los últimos años la web semántica ha atraído a investigadores, laboratorios, empresas e instituciones de los cinco continentes, y sigue ganando en popularidad. En muy pocos años se ha consolidado una comunidad internacional con un importante respaldo de los programas de financiación de agencias norteamericanas y europeas. Entre los grupos líderes de esta comunidad podemos citar al consorcio W3C, el Knowledge Systems Laboratory de la Universidad de Stanford, la Universidad Libre de Amsterdam (Dieter Fensel), el Knowledge Management Research Group de la Universidad de Karlsruhe en Alemania, el Information Management Group de la Universidad de Manchester, el Knowledge Media Institute de la Open University en el Reino Unido, el Consejo Nacional de Investigación (Nicola Guarino) en Italia, y el Laboratorio de Inteligencia Artificial de la Universidad Politécnica de Madrid.

Existe un gran interés desde el entorno corporativo, el sector público y el mundo académico por hacer de la web semántica una realidad, ya que se piensa que puede ser una pieza importante para el progreso de la sociedad de la información. Para ello se está invirtiendo un gran esfuerzo en desarrollar a) la infraestructura necesaria para su despliegue, b) aplicaciones que demuestren la viabilidad y el beneficio de la web semántica y a la vez motiven el desarrollo y consumo de infraestructura y c) nuevas soluciones para resolver problemas específicos, e ideas que mejoren, amplíen y/o exploten las posibilidades de la web semántica. Entre las principales líneas de trabajo que están siendo objeto de atención cabe citar:

- Lenguajes de definición de ontologías
- Metodologías de desarrollo de ontologías
- Integración de ontologías
- Aprendizaje de ontologías
- Desarrollo de vocabularios en dominios concretos
- Agentes
- Servicios web

Los resultados alcanzados hasta ahora son muy preliminares si se mira desde la óptica más ambiciosa, la de la adopción universal de la web semántica. En espera de que se alcance esta meta y al margen de ese debate, creemos que se han desarrollado ideas muy aprovechables a niveles más específicos, y se han abierto ya nuevos campos para la innovación.

En nuestro grupo de la Universidad Autónoma de Madrid hemos puesto en marcha una línea de investigación dedicada a la web semántica. En particular, y en el contexto de los temas de interés para AIPO, nos hemos preocupado por la interacción del usuario con la web semántica [Castells 2002b]. La web semántica plantea problemas de construcción, manipulación, consumo, intercambio y aprovechamiento del conocimiento y las estructuras que lo soportan. Surgen nuevas necesidades que resolver a la vez que nuevas posibilidades para mejorar la interacción con el usuario, o facilitar el desarrollo de interfaces. Hemos observado un cierto vacío en cuanto a herramientas para la interfaz de usuario específicamente orientadas a la manipulación de conocimiento basado en ontologías. Se han desarrollado herramientas de autor interactivas para la construcción de ontologías como Protégé [Noy 2001] y OilEd [Bechhofer 2001], pero en nuestro conocimiento no se ha producido aún una tecnología de soporte genérica para dotar de interfaz de usuario a las aplicaciones de la web semántica.

En particular, creemos que las descripciones explícitas de recursos y servicios en la web semántica pueden ser explotadas para desarrollar mejores interfaces de usuario por un coste razonable. En esta línea, hemos desarrollado un sistema para el acceso a bases de conocimiento multimedia distribuidas donde la selección y presentación del conocimiento se hace de forma adaptada al usuario, los dispositivos y otros factores del contexto de uso [Macías 2001a, Castells 2001, 2002a], al tiempo que se permite un alto grado de libertad en la forma de representar internamente el conocimiento. Esta herramienta permite ampliar las ontologías del sistema con otros modelos (usuario, plataforma, otros) para adaptar dinámicamente la presentación. El usuario interactúa con el sistema mediante páginas web generadas automáticamente a partir de un modelo explícito de diseño de página (presentación) estrechamente vinculado a la ontología y el modelo de usuario. El diseñador puede modificar la generación de páginas editando el modelo de la presentación, independientemente de los contenidos.

Nuestro trabajo en este contexto se complementa con el desarrollo de herramientas para la construcción interactiva de ontologías como [Macías 2001b]. Estamos explorando así mismo la utilización de ontologías para identificar fragmentos de datos en una página web, y permitir la autoría de páginas dinámicas en entornos WYSIWYG [Macías 2002a, 2002b, 2003]. Este último trabajo combina técnicas de programación mediante ejemplos [Cypher 1993, ACM 2000] con una representación del conocimiento propia de la web semántica.

Dado que hasta ahora nos hemos ocupado principalmente de la presentación en la interfaz de usuario, una de las perspectivas más inmediatas de ampliación de nuestro trabajo es abordar el diálogo con el usuario. Estamos estudiando la utilización de descripciones explícitas de servicios web para dar soporte al desarrollo del diálogo, mediante 1) la generación de componentes para solicitar información de entrada al usuario, 2) la composición y secuenciación de pasos de interacción para completar procedimientos compuestos 3) la capacidad de recuperación frente a errores del usuario o del sistema y 4) la presentación de resultados al usuario en páginas dinámicas. Para ello nos basaremos en estándares para la descripción de servicios web como WSDL [Christensen 2001], DAML-S [Ankolenkar 2002] y WSMF [Fensel 2002].

En general hemos observado que las técnicas basadas en la modelización ontológica de un dominio son más útiles allí donde la información presenta más riqueza y diversidad (p.e. materias científicas, arte, turismo), se presta a una clasificación profusa (teorema, definición, algoritmo, función; cuadro, escultura, pintor, técnica, escuela, museo; lugar, transporte, hotel, monumento), con gran variedad de relaciones entre las unidades de información. El tipo de técnicas sobre las que estamos trabajando son útiles para sistematizar la introducción del conocimiento y recursos, asegurar la coherencia necesaria para la compartición, flexibilizar la composición dinámica de documentos, multiplicar las posibilidades de navegación (a través de un mayor número de relaciones representadas), afinar la adaptación al usuario (selección de conocimiento, medio, estructura y estilo de presentación), y gestionar repositorios de conocimiento abiertos a la extensión.

Referencias

[ACM 2000] Communications of the ACM. The Intuitive Beauty of Computer Human Interaction. Special issue on Programming by Demonstration, 43, 3, March 2000.

[Ankolenkar 2002] A. Ankolenkar, M. Burstein, J. R. Hobbs, O. Lassila, D. L. Martin, D. McDermott, S. A. McIlraith, S. Narayanan, M. Paolucci, T. R. Payne and K. Sycara. DAML-S: Web Service Description for the Semantic Web. The First International Semantic Web Conference (ISWC), June 2002.

- [Bechhofer 2001] S. Bechhofer, I. Horrocks, C. Goble, R. Stevens. OilEd: a Reason-able Ontology Editor for the Semantic Web. Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, September 19-21, Vienna. Springer-Verlag LNAI Vol. 2174, pp. 396-408, 2001.
- [Berners-Lee 2001] T. Berners-Lee, J. Hendler, O Lassila. The Semantic Web. Scientific American, May 2001.
- [Bray 2000] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler. Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation 6 October 2000. Available at <http://www.w3.org/TR/REC-xml>.
- [Castells 2001] P. Castells, J. A. Macías. An Adaptive Hypermedia Presentation Modeling System for Custom Knowledge Representations. Actas World Conference on the WWW and Internet (WebNet'2001). Orlando (Florida), 2001, pp. 148-153.
- [Castells 2002a] P. Castells y J. A. Macías. Un sistema de presentación dinámica en entornos web para representaciones personalizadas del conocimiento. Revista Iberoamericana de Inteligencia Artificial, No. 16 (2002). AEPIA, 2002, pp. 25-34.
- [Castells 2002b] P. Castells and J. A. Macías. Context-Sensitive User Interface Support for Ontology-Based Web Applications. International Semantic Web Conference (ISWC), Collected Posters. Sardinia (Italy), June 2002.
- [Connolly 2001] D. Connolly, F. van Harmelen, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. DAML+OIL Reference Description. W3C Note 18 December 2001. Available at <http://www.w3.org/TR/daml+oil-reference>.
- [Christensen 2001] E. Christensen, F. Curbera, G. Meredith, S. Weerawarana. Web Services Description Language (WSDL) 1.1. W3C Note 15 March 2001. Available at <http://www.w3.org/TR/wsdl>.
- [Cypher 1993] Cypher A. (ed.). Watch What I Do: Programming by Demonstration. The MIT Press, 1993.
- [Dean 2002] M. Dean, D. Connolly, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. OWL Web Ontology Language 1.0 Reference W3C Working Draft 29 July 2002. Available at <http://www.w3.org/TR/owl-ref>.
- [Fensel 2002] D. Fensel and C. Bussler. The Web Service Modeling Framework WSMF. White paper, available from <http://www.cs.vu.nl/~dieter/wese/publications.html>. 2002.
- [Gruber 1993] T. R. Gruber. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2), pp. 199-220, 1993.
- [Lassila 1999] O. Lassila, R. R. Swick. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999. Available at <http://www.w3.org/TR/REC-rdf-syntax>.
- [Macías 2001a] J. A. Macías and P. Castells. A Generic Presentation Modeling System for Adaptive Web-based Instructional Applications. Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2001), Extended Abstracts. Seattle (Washington), April 2001, pp. 349-350.
- [Macías 2001b] J. A. Macías and P. Castells. An Authoring Tool for Building Adaptive Learning Guidance Systems on the Web. Lecture Notes in Computer Science: Active Media Technology – AMT 2001. Springer-Verlag, Viena, 2001, pp. 268-278.
- [Macías 2002a] J. A. Macías y P. Castells. Personalización de páginas web dinámicas mediante ejemplos. III Congreso Internacional de Interacción Persona-Ordenador (Interacción 2002). Madrid, May 2002.
- [Macías 2002b] J. A. Macías and P. Castells. Tailoring Dynamic Ontology-Driven Web Documents by Demonstration. Proceedings of the 6th International Conference on Information Visualisation - International Symposium of Visualisation of the Semantic Web. IEEE Computer Society, London, July 2002.
- [Macías 2003] J. A. Macías and P. Castells. Dynamic Web Page Authoring By Example Using Ontology-based Domain Knowledge. ACM International Conference on Intelligent User Interfaces (IUI 2003). Miami (Florida), January 2003.
- [Noy 2001] N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Ferguson, & M. A. Musen. Creating Semantic Web Contents with Protege-2000. IEEE Intelligent Systems 16(2), pp. 60-71, 2001.