

4.5. ANOTACIONES SEMÁNTICAS.

Como ya hemos comentado varias veces a lo largo del presente trabajo, estamos acostumbrados a usar el Web dado que es una gran fuente de información accesible mediante nuestros ordenadores pero, hasta ahora, sólo comprensible para los seres humanos. Al principio, los documentos Web se hacían a mano y nacían orientados al intercambio de información entre las personas. Todos estos documentos constituyen una enorme cantidad de texto, imágenes e incluso sonido, sin significado para una computadora. El usuario era el encargado de extraer e interpretar la información relevante. Actualmente, dado el asombroso crecimiento de la información contenida en Internet resulta imposible que un único usuario realice esas tareas en un tiempo aceptable. Al mismo tiempo, han surgido nuevas tecnologías que facilitan la gestión y recuperación de la información y, con ellas, aparece también la generación semi-automática de documentos Web.

En la actualidad, la presentación en la WWW de este tipo de documentos tiende a tratarse independientemente de su contenido, principalmente mediante la utilización de XML, RDF, RDF Schema o OWL. Pero aunque con ello se facilite el procesamiento automático de la información, un ordenador no puede llevar a cabo por sí solo las tareas de acceso, extracción e interpretación de la información relevante. Conseguir que las máquinas entiendan el significado, la semántica, de los textos escritos y de las páginas Web es uno de los dos pilares principales que sustenta el desarrollo del Web Semántico. En este contexto, la "anotación semántica" de páginas Web, que hace explícito el significado para un ordenador, se ha convertido en un punto clave.

Los investigadores de Inteligencia Artificial han encontrado en las ontologías el modelo del conocimiento ideal para describir formalmente los recursos Web y su vocabulario y, por tanto, para hacer explícito de algún modo el significado subyacente de los términos incluidos en las páginas Web. La Semántica Ontológica [Niremburg,01] es una teoría que estudia el significado del lenguaje humano o lenguaje natural, así como una aproximación al Procesamiento del Lenguaje Natural que utiliza un modelo abstracto del mundo (la ontología) como recurso central para extraer y

representar el significado de textos en lenguaje natural, al razonar con el conocimiento que se deriva a partir de estos textos. Asimismo, la ontología es también el eje central a la hora de generar textos en lenguaje natural basados en las representaciones de su significado.

Con la Semántica Ontológica como punto de partida para la anotación de recursos Web con información ontológica, se pretende conseguir que esta anotación permita el acceso inteligente a dichos recursos, facilite la búsqueda y navegación en la Web y explote nuevos enfoques de inferencia a partir de estos recursos [Gea,02]. La anotación de páginas Web constituye uno de los campos más activos dentro de la actividad del Web Semántico¹.

4.5.1. ¿Qué se entiende por anotación semántica?

Una anotación se puede considerar como una información sobre las entidades o conceptos de una ontología (*Figura 12*), que aparecen en un texto y su situación en el mismo, o también las referencias que hay en un texto sobre un repositorio semántico en el que hay más conocimiento [Atanas,03]

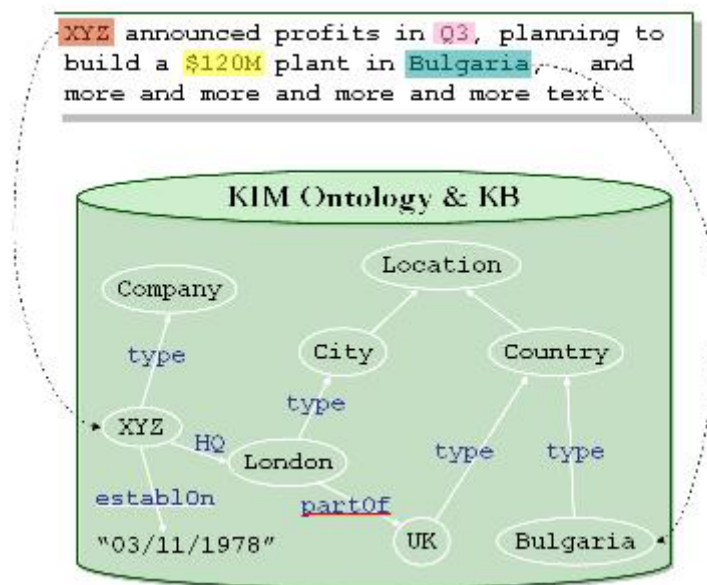


Figura 12. Anotación semántica a una ontología.

¹Semantic Web, Annotation & Authoring. Disponible en: <http://annotation.semanticWeb.org/tools/> (2/07/04).

Según el diccionario de la RAE, anotación “constituye la acción y efecto de anotar”, y anotar, en su principal acepción, “poner notas en un escrito, una cuenta o un libro”.

En lingüística computacional una anotación es una nota añadida a una parte específica de un texto. Existen dos alternativas para realizar anotaciones: anotaciones incrustadas en el texto (en las páginas Web) creadas mediante los lenguajes de marcado (XML) o anotaciones externas, no incrustadas en las páginas Web.

4.5.2. Herramientas de anotación.

Las herramientas de anotación permiten añadir contenido semántico a las páginas Web de Internet. Es decir, estas herramientas permiten estructurar la información publicada mediante su clasificación en base a conceptos semánticos. Este es el primer paso para permitir el procesamiento automático de la información de la WWW por parte de las máquinas, es decir, para la creación de la Web semántica [Murua,04].

Las herramientas de anotación podemos englobarlas en dos grupos:

- Herramientas de anotación externa: permiten asociar metainformación a páginas Web que ya existen en el WWW. La metainformación no se almacena dentro de la misma página, sino que se almacena de forma externa en un repositorio destinado específicamente a mantener las anotaciones. Estos repositorios suelen ser bases de datos RDF.
- Herramientas de anotación de autor: permiten incluir la información estructurada dentro de las propias páginas, mediante los lenguajes de marcado XML o RDF.

Las anotaciones se guardan clasificadas como objetos o propiedades correspondientes a un concepto de una ontología. Por tanto, como herramientas complementarias a las de anotación disponemos de los editores de ontologías, que facilitan la tarea de definición de ontologías, la unión de diferentes ontologías, y el desarrollo distribuido de ontologías.

En el proyecto RODA (Red de Conocimiento Descentralizado a través de Anotaciones)² se han evaluado algunas utilidades de anotación, por ejemplo:

COHSE (The Conceptual Open Hypermedia Project)³: es un proyecto de investigación sobre métodos que mejoren significativamente la calidad, consistencia y la amplitud de documentos Web enlazados mientras se recuperan (cuando los lectores navegan sobre los documentos) y al mismo tiempo que se crean (cuando los autores crean los documentos). COHSE utiliza tres tecnologías: un servicio de razonamiento de ontologías que se utiliza para representar un sofisticado modelo conceptual de términos documentales y sus relaciones, un servicio abierto de enlaces hipermedia basado en Web, que sea escalable, y la integración de estos dos para poder enlazar documentos vía metadatos describiendo sus contenidos. Para su empleo, el usuario final sólo necesita tener instalado el navegador Web Mozilla, junto con cierto software de COHSE para este navegador. Con el navegador, el usuario puede realizar anotaciones y ver las notaciones realizadas por el resto de usuarios.

Annotea⁴: Proyecto que pretende mejorar el ambiente colaborativo de la W3C (el "Consortio de la World Wide Web") a través del uso de anotaciones compartidas; se entiende que una anotación puede ser un comentario, una nota, una explicación o cualquier texto que se pueda adjuntar a un documento Web externamente, es decir, sin necesidad de tocar el documento mismo. El usuario recibe las notas de un servidor o varios, adjuntas a un documento junto con éste, y puede ver qué piensan sus colegas del documento. En este proyecto se ha desarrollado un navegador Web propio, llamado **Amaya**, para poder ver y crear las anotaciones. Para su empleo, el usuario final sólo necesita tener instalado dicho navegador o el de Mozilla. En el caso de emplear Mozilla, necesita instalar el software de Annozilla para este navegador. Con estos navegadores, el usuario puede realizar anotaciones y ver las notaciones realizadas por el resto de usuarios.

²Proyecto RODA. Red de conocimiento descentralizado a través de anotaciones. Disponible en: <http://roda.ibit.org/index.cfm> (6/07/04).

³COHSE, The Conceptual Open Hypermedia Project. Disponible en: <http://cohse.semanticWeb.org/> (3/07/04).

⁴Annotea Project. Disponible en: <http://www.w3.org/2001/Annotea/> (3/07/04).

Annozilla⁵: Proyecto diseñado para ver y crear anotaciones asociadas a una página Web mediante el navegador Mozilla, como define el proyecto **Annotea** del W3C.

Trellis Web⁶: Entorno interactivo que permite a los usuarios añadir sus observaciones, puntos de vista y conclusiones sobre la información que analizan, realizando anotaciones semánticas a los documentos y a otros recursos on-line. En realidad es una forma de adquirir conocimiento donde el usuario puede añadir nuevo conocimiento al sistema basándose en su propia experiencia a medida que analiza la información. Para su empleo, el usuario final sólo necesita tener instalado un navegador Web capaz de soportar "frames". Con el navegador, el usuario puede realizar anotaciones y ver las notaciones realizadas por el resto de usuarios.

Las anteriores son herramientas de anotación externa. Como herramientas de anotación de autor se mencionan a modo de ejemplo:

MnM⁷: MnM es una herramienta de anotación basada en ontologías que permite anotar páginas Web con contenidos semánticos de forma automática y semi-automática. MnM integra un navegador Web con un editor de ontologías y proporciona unas APIs abiertas para enlazar MnM con servidores de ontologías y para integrar MnM con herramientas de extracción de información. MnM trabaja con diferentes lenguajes de ontologías como RDF, DAML+OIL y OCML.

OntoMat-Annotizer⁸: Herramienta de anotación de páginas Web interactiva y amigable. Ayuda al usuario en la tarea de crear y mantener ontologías basadas en DAML+OIL, por ejemplo, para crear instancias, atributos y relaciones. Incluye un navegador de ontologías para explorar la ontología y las instancias y un navegador HTML que visualiza las partes anotadas del texto.

SHOE Knowledge Annotator⁹: Herramienta que permite realizar anotaciones en documentos Web sin tener que preocuparse de los códigos

⁵ Annozilla (Annotea on Mozilla) . Disponible en: <http://annozilla.mozdev.org/> (3/07/04).

⁶ Trellis Web. Capturing and Exploiting Semantic Relationships for Information and Knowledge Management. Disponible en: <http://trellis.semanticWeb.org/> (3/07/04).

⁷ MnM. Disponible en: <http://kmi.open.ac.uk/projects/akt/MnM/> (3/07/04)

⁸ Ontomat Annotizer. Disponible en: <http://annotation.semanticWeb.org/ontomat/index.html> (3/07/04).

⁹ The SHOE Knowledge Annotator. Disponible en:
<http://www.cs.umd.edu/projects/plus/SHOE/KnowledgeAnnotator.html> (3/07/04).

HTML, pues añade las etiquetas necesarias automáticamente. Las etiquetas están divididas en dos categorías: para la construcción de ontologías y para anotar documentos Web para suscribirlos a una o más ontologías, declarar entidades de datos y realizar afirmaciones sobre las entidades según las normas proscritas por las ontologías.

Algunas de las diferencias entre los dos tipos de anotaciones externas son, que para poder realizar y ver anotaciones con las herramientas de anotación externa, el usuario final sólo necesita tener instalado en su ordenador un navegador Web junto con cierto software adicional para éste. Por tanto, su empleo no requiere formar al usuario en el empleo de ninguna aplicación adicional, a diferencia de lo que ocurre con las herramientas de anotación de autor.

Las herramientas de anotación externa son útiles [Murua,04] para realizar comentarios personales sobre páginas Web, para mantener discusiones en grupo sobre estas páginas, para compartir información, como herramientas de ayuda para el mantenimiento de bookmarks compartidos, etc.

[Niremburg,01] Nirenburg, S. y Raskin, V. (2001) Ontological Semantics (Draft) <http://crl.nmsu.edu/Staff/pages/Technical/sergei/book/index-book.html> (2/07/04).

[Gea,02] Cea, G. et al. Primeras aproximaciones a la anotación lingüístico-ontológica de documentos de la Web Semántica: OntoTag. Revista Iberoamericana de Inteligencia Artificial. No.17, pp. 37-49 ISSN: 1137-3601. (2002). Disponible en: <http://tornado.dia.fi.upm.es/caepia/numeros/18/aguado.pdf> (2/07/04).

[Atanas,03] Atanas, K. et al.. Semantic Annotation, Indexing, and Retrieval. Human Language Technologies Workshop at the 2nd International Semantic Web Conference (ISWC2003), 20 October 2003, Florida, USA. Disponible en: http://www.ontotext.com/publications/SemAIR_ISWC169.pdf (3/07/04).

[Murua,04] Murua, I. La utilidad de las herramientas de anotación. Proyecto RODA. Disponible en: <http://roda.ibit.org/articulo1.cfm> (3/07/04)

[Murua,04] Murua, I. La utilidad de las herramientas de anotación. Proyecto RODA. Disponible en: <http://roda.ibit.org/articulo1.cfm> (3/07/04)