

Avances hacia la Aplicación de las Tecnologías de la Web Semántica en las Organizaciones.

Alonso Pérez Soltero

Rene Francisco Navarro Hernández

Gerardo Sánchez Schmitz

Mario Barceló Valenzuela

Departamento de Ingeniería Industrial, Universidad de Sonora.

Hermosillo, Sonora. México

{aperez, navarro, gsanchez, mbarcelo}@industrial.uson.mx

Resumen

Actualmente las organizaciones utilizan tecnologías que se utilizan en la Web actual y las han incorporado para la gestión de su información, sin embargo, tienen dos carencias cruciales. En primer lugar, no incorporan mecanismos que permitan el procesado automático de la información; esto es, que hagan que la información pueda ser procesada por máquinas; en segundo lugar, la Web actual no incluye mecanismos para la interoperabilidad completa de los sistemas de información basados en la Web.

El objetivo del presente trabajo es presentar algunos proyectos y prototipos en los que se utilizan las tecnologías de la Web semántica y que potencialmente tienen aplicación dentro de las empresas para que les facilite compartir sus datos, lograr el intercambio de información interorganizacional, facilitar la búsqueda, recuperación y manipulación de información de manera automatizada; todo esto con la finalidad de que sean más eficientes y competitivas con ayuda de las tecnologías de la información.

En el presente artículo se describen distintos proyectos y prototipos recientes en los que se aplican tecnologías de la Web semántica y que pueden ser susceptibles de implementarse en las empresas.

Palabras Clave

Web semántica, proyectos, prototipos, organización, ontologías, XML, RDF, OWL.

Introducción

Las tendencias en el mundo como la globalización, desarrollo tecnológico y la competencia llevan a las organizaciones a redefinir sus estrategias para alcanzar sus objetivos. En un mundo de impresionantes avances tecnológicos, especialmente en el área computacional y en telecomunicaciones, obliga a las organizaciones a estar a la vanguardia en sus sistemas de información. Para estas organizaciones la tecnología de información

puede ser una herramienta muy poderosa para ayudarlas a definir sus estrategias, optimizar su organización y manejo de su información y así incrementar su productividad.

El objetivo del presente trabajo es presentar algunos proyectos y prototipos en los que se utilizan las tecnologías de la Web semántica y que potencialmente tienen aplicación dentro de las empresas para que les facilite compartir sus datos, lograr el intercambio de información interorganizacional y facilitar la búsqueda de información a usuarios internos y externos; todo esto con la finalidad de que sean más eficientes y competitivas con ayuda de las tecnologías de la información.

La estructura del artículo inicia con conceptos generales sobre la Web actual y sus problemas, Web semántica y las tecnologías que utiliza; después se describen distintos proyectos y prototipos recientes de aplicación de las tecnologías de la Web semántica y que pueden ser susceptibles de aplicarse en las empresas; finalmente se muestran las conclusiones del presente trabajo.

Conceptos Generales

A continuación se describen los principales conceptos que permiten entender la importancia de las tecnologías de la Web semántica para resolver algunos de los principales problemas que se tienen con la Web actual.

Web actual

La característica más llamativa de la Web es su estructura hipertextual: la World Wide Web contiene una gigantesca cantidad de documentos de hipertexto (llamados páginas Web). Los documentos de hipertexto contienen enlaces que conectan con otros documentos; un enlace puede ser una palabra, una frase o una imagen.

El uso actual de la Web se puede resumir según lo define el grupo de trabajo del W3C (W3C XML Protocol Working Group): Hoy, el principal uso de la Web es para acceso interactivo a documentos y aplicaciones. En casi todos los casos, este acceso es hecho por usuarios humanos, que típicamente trabajan con navegadores Web, reproductores de audio u otros sistemas interactivos en el lado del usuario.

Problemas principales de la Web Actual

Como lo menciona (Abian, 2005), si bien la Web actual constituye un avance tecnológico impresionante, tiene dos carencias cruciales. En primer lugar, no incorpora mecanismos que permitan el procesamiento automático de la información; esto es, que hagan que la información pueda ser procesada por máquinas.

En segundo lugar, la Web actual no incluye mecanismos para la interoperabilidad completa de los sistemas de información (SI) basados en la Web. Dicho en otras palabras: no facilita la creación de una comprensión común y compartida de un dominio, de forma que ésta pueda ser usada por personas, organizaciones y máquinas. Para que haya una comprensión común de un dominio, los SI de las partes interesadas en él deben cumplir tres tipos de interoperabilidad: *Interoperabilidad técnica*, se refiere a la capacidad de los SI para

intercambiar señales. Esta interoperabilidad exige una conexión física (cable, fibra óptica, ondas electromagnéticas) entre los sistemas y un conjunto de protocolos de comunicaciones (como la pila TCP/IP); *Interoperabilidad sintáctica*, se refiere a la capacidad de los SI para leer datos procedentes de otros SI y obtener una representación que puedan usar (por ejemplo, objetos). En definitiva, significa que, en todos los SI involucrados, tanto la codificación de los datos como los protocolos de acceso sean compatibles; *Interoperabilidad semántica*, es la capacidad de los SI para intercambiar información basándose en un común significado de los términos y expresiones que usan. En otras palabras: denota la capacidad de los SI y para intercambiar información consistente con el significado que se le supone.

De la falta de interoperabilidad de la Web se derivan muchos problemas, demasiados para enumerar los aquí. Se considerarán dos de los más relevantes desde el punto de vista económico, esto es, los dos problemas que más caros resultan a los usuarios y a las organizaciones que usan el Web (Abian, 2005):

1.- *Dificultad para encontrar información*. Los buscadores Web actuales buscan las palabras en los documentos de la Web, sin atender a la sinonimia o polisemia de las palabras, esto provoca que a veces sea necesario consultar docenas o cientos de páginas hasta encontrar la información deseada. Hasta el momento, la intervención humana resulta imprescindible para seleccionar la información que uno desea. No son los buscadores Web los responsables últimos de la dificultad para encontrar información en la Web, sino el lenguaje con que se construyen las páginas Web: HTML. Este lenguaje de etiquetado sólo permite especificar cómo se va a mostrar una página HTML, pero resulta inútil para dar "contexto" a una página. En pocas palabras: HTML no proporciona semántica a los datos.

2.- *Dificultad para implantar comercio electrónico B2B*. En el presente, el principal reto al que se enfrentan las empresas es la integración con otras (clientes, proveedores, fabricantes y subcontratistas). Para lograr esto, los SI de las empresas deben permitir el intercambio de información interempresarial, ya sea comercial (albaranes, facturas, pedidos, etc.) o de carácter logístico o de planificación. Compartir información no resulta fácil: cada SI usa distintos formatos, definiciones, estructuras de datos, modelos de negocio, entre otros.

La solución a los problemas anteriores vendrá de lo que se conoce como la Web semántica.

Web Semántica

La Web representa un enorme repositorio de información formado de un conjunto de fragmentos que de alguna manera están integrados e interrelacionados, asociados a dominios particulares y a las empresas. Sin embargo, la información como se almacena actualmente, generalmente no tiene un claro significado que facilite su recuperación y manipulación de manera automatizada o manual. La idea principal de la Web semántica es tratar de resolver esta deficiencia (Barcelo-Valenzuela et al., 2006).

La Web semántica no es una Web separada, sino una extensión de la actual, en la cual la información se relaciona a un significado bien definido que permite trabajar en cooperación a las personas con computadoras (Berners-Lee et al., 2001).

La WS es la nueva generación de la Web mundial, basada en un formalismo de representación semántica del conocimiento de la red, que permite que la información esté empaquetada en forma de declaraciones del objeto-atributo-valor, denominadas triadas. Se

asume que los términos usados en estas declaraciones están basados en ontologías formalmente especificadas (para la comunidad de interés), estas tríadas se pueden procesar semánticamente por los agentes de manera automática (Stojanovic & Handschuh, 2002).

El propósito de la iniciativa de la Web semántica es tan amplio como el de la Web: crear un medio universal para el intercambio de datos. Se considera para interconectar eficazmente la gestión de la información personal, la integración de las aplicaciones empresariales y compartir globalmente datos comerciales, científicos y culturales. Los servicios para poner datos comprensibles por las máquinas se están convirtiendo rápidamente en una prioridad para las organizaciones, individuos y comunidades.

La Web sólo puede alcanzar todo su potencial si llega a ser un sitio donde se puedan compartir datos y sean procesados por herramientas automáticas, así como por personas. Para adaptar la Web, los programas del mañana deben ser capaces de compartir y procesar datos incluso cuando estos programas se hayan diseñado de forma completamente independiente (W3C, 2001).

La Web semántica permitirá a los usuarios buscar información de una manera imposible hoy. Los buscadores del futuro permitirán consultas como: "Busco todos los mecánicos que tengan su taller a menos de 1 km de la calle Manuel Candela y que trabajen para la compañía de seguros Siempre Seguro. Muéstrame primero los más baratos". La Web semántica también permitirá consultas basadas en otras consultas. Asimismo, la Web semántica también permitirá el uso de agentes personales encargados de extraer información de múltiples fuentes, cada una con su propia manera de representar los datos, y de informar a los usuarios de los sucesos que son de interés (conciertos, charlas, estrenos de películas, etc.). Por último, los buscadores Web del mañana no sólo encontrarán las páginas donde aparezcan los términos de la búsqueda, sino también todas aquellas páginas donde haya sinónimos de esas palabras (Abian, 2005).

Tecnologías de la Web Semántica

A continuación se mencionan algunas tecnologías que son necesarias para lograr la funcionalidad previamente descrita.

Lo primero es tener una representación de los datos apropiada dentro de las páginas Web para que puedan ser procesados por las máquinas. El término *metadato* se refiere a este tipo de representación: datos sobre los datos. Los metadatos capturan la parte del significado de datos, esto es el término *semántico* en la Web semántica (Antoniou & Harmelen, 2004).

Los principales componentes de la Web semántica son los metalenguajes y estándares de representación XML, XML Schema, RDF, RDF Schema y OWL. La OWL Web Ontology Language Overview (W3C, 2001), describe la función y relación de cada uno de estos componentes de la Web Semántica: XML aporta la sintaxis superficial para los documentos estructurados, pero sin dotarles de ninguna restricción sobre el significado. XML Schema es un lenguaje para definir la estructura de los documentos XML. RDF es un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos. Aporta una semántica básica para este modelo de datos que puede representarse mediante XML. RDF Schema es un vocabulario para describir las propiedades y las clases de los recursos RDF con una semántica para establecer jerarquías de generalización entre dichas

propiedades y clases. OWL añade más vocabulario para describir propiedades y clases: tales como relaciones entre clases (p.ej. disyunción), cardinalidad (por ejemplo "únicamente uno"), igualdad, tipologías de propiedades más complejas, caracterización de propiedades (por ejemplo simetría) o clases enumeradas (Wikipedia, 2006).

Uno de los elementos fundamentales en la Web semántica es el uso de ontologías para la integración semántica. El concepto de ontología ha recibido múltiples definiciones a lo largo de la historia. Para nuestros propósitos definiremos a una ontología como una especificación explícita de una conceptualización (Gruber 1995). Lo "explícito" va relacionado a los conceptos, las propiedades, funciones y axiomas, que están definidos explícitamente. La "conceptualización", se refiere a un modelo abstracto del mundo real. Posteriormente, la definición fue expandida: Una ontología es una especificación explícita y formal de una conceptualización compartida (Borst 1997), las palabras claves en la segunda son "formal" relacionada a que sea interpretable de manera automática, y "compartida" que implica, el consenso que se da entre los agentes sobre la conceptualización y para ello se debe de disponer de un vocabulario consistente y coherente, para que sea compartido entre los diferentes agentes involucrados. Es decir, esta última definición enfatiza que el conocimiento contenido en una ontología debe ser un conocimiento "público" y consensuado.

La utilización de ontologías para la integración semántica representa ciertas ventajas (Noy, 2004; Uschold, 2004):

- La posibilidad de representar diferentes estructuras o modelos de datos.
- La combinación de deducción y sistemas de bases de datos relacionales que extienden las capacidades de los mapeos dentro del dominio de la lógica del negocio.
- Un mayor nivel de abstracción, ya que el modelo se separa del almacenamiento físico de los datos.
- Mayor flexibilidad y mayor reusabilidad.

Adicionalmente, podemos señalar cuatro posibles escenarios de aplicación de ontologías (Jasper, 1999):

- *Desarrollo Neutral*: Dada la gran cantidad de herramientas y formatos, que funcionan de manera aislada, una determinada compañía u organización se puede beneficiar mediante el desarrollo de su propia ontología neutral y después desarrollar traductores de esta ontología a la terminología de los diferentes sistemas. El modelado de los procesos de negocios de una organización es un buen ejemplo de la aplicación de ontologías neutrales, particularmente en organizaciones que deben integrar diferentes sistemas empleando un núcleo central semánticamente uniforme.
- *Acceso común a la información*: En cualquier área en la cual se requiera interoperar con sistemas de software siempre será necesario traducir diferentes formatos y representaciones que evolucionaron independientemente. Este es un escenario similar al anterior, ya que es necesario crear una ontología neutral del dominio de los sistemas que debe interoperar. La diferencia principal es que la traducción es bidireccional.

- *Especificaciones basadas en ontologías*: La idea es crear una ontología que represente y especifique los aspectos que un sistema de software debe satisfacer y después utilizar esta ontología como un conjunto (parcial) de requerimientos para la construcción del software. Por ejemplo, el diseño de todos los artefactos de un producto de software (incluyendo herramientas de visualización y presentación, bases de datos y aun las herramientas de mercadeo y contabilidad para dar seguimiento a las ventas del producto) se pueden basar en una misma ontología. Esto asegura un interoperabilidad más eficiente entre los sistemas de software cuyas relaciones son típicamente implícitas.
- *Búsqueda basada en ontologías*: Para facilitar la búsqueda una ontología puede servir como base para la estructuración de un repositorio de información (e.g. documentos, páginas web, nombres de expertos). Esto permite la organización y clasificación de repositorios de información a un mayor nivel de abstracción al comúnmente usado actualmente. El uso de ontologías para estructurar la información implica el uso de técnicas de indexado semántico o agregar anotaciones semánticas a los documentos. Si diferentes repositorios se indexan en base a diferentes ontologías, entonces un sistema de acceso a la información integrada semánticamente pudiera realizar los mapeos entre diferentes ontologías y generar respuestas de múltiples repositorios.

Proyectos recientes de Web semántica con potencial aplicación en las empresas

A continuación se dará un panorama general y se mencionarán algunos proyectos recientes donde utilizan las tecnologías asociadas a la Web semántica en diferentes ámbitos de aplicación y que tienen potencial de implementarse en las organizaciones. Se describirán el objetivo del proyecto, la estructura general del sistema y las tecnologías utilizadas, posibles áreas de aplicación en la empresa y los beneficios que puede aportar.

Proyecto RETSINA Calendar Agent (RCAL)

Este sistema automatizado para agendar reuniones busca contenidos de la Web semántica recolectando información relativa a eventos y calendarios (codificadas en diferentes ontologías estandarizadas) (Payne, 2002). Para cumplir con este objetivo eficientemente el sistema mantiene información actualizada acerca de las actividades del usuario para evitar solicitar información adicional en el proceso de programar las reuniones convenientemente para el usuario.

RCAL funciona con sistemas de gestión de información personal comerciales (ej. MS Outlook) recuperando detalles de la agenda y contactos deduciendo cuáles son los horarios disponibles para reuniones. Puede también negociar con otros agentes RCAL para determinar disponibilidad mutua de fechas y horarios. Los detalles de agendas y contactos encontrados en la Web semántica se pueden importar mediante un navegador semántico de

agendas web (ej. el usuario introduce un URL del programa de una conferencia y selecciona las ponencias específicas a las cuales el usuario asistirá) u obteniendo e importando automáticamente nuevas agendas compartidas por una comunidad de agentes.

Las ideas desarrolladas en este proyecto se pueden implementar en las empresas en las que de manera automática se requiera la programación de reuniones de trabajo, calendarizar juntas, cursos de capacitación y todas aquellas actividades donde se pretenda la reunión de varios integrantes de la organización que estén presentes presencial o virtualmente durante un lapso de tiempo considerando la disponibilidad de cada uno de los involucrados. También puede servir para compartir datos entre calendarios de las personas que están dentro como fuera de la organización.

Entre los beneficios que puede obtener la organización es una eficiente gestión y aprovechamiento de los recursos humanos de la empresa cuando se requiera su participación en actividades grupales.

Proyecto MAESTRA

Se trata de un proyecto que integra dos áreas del conocimiento a través de la aplicación de conceptos de Ingeniería de Sistemas relacionados con la Arquitectura de la Información y de Ingeniería Civil pertenecientes al sistema constructivo Mampostería Estructural (Vidal-Rojas, 2006). Conformado de un sistema multiagente para la optimización de búsquedas que con base en el modelado del conocimiento en Mampostería Estructural se filtra el conjunto de páginas entregadas por motores de búsqueda convencionales, asegurando un resultado más preciso y acorde con los requerimientos de búsqueda del usuario. MAESTRA entrega resultados concretos que evitan pérdidas de tiempo en costosos procesos de clasificación y análisis de grandes cantidades de información que usualmente generan los buscadores tradicionales, resultados que si bien algunas veces logran satisfacer los requerimientos de búsqueda, en la mayoría de los casos no poseen la información que el usuario requiere.

Los aspectos más importantes del dominio, se representan mediante una ontología, con la cual interactúa el sistema multiagente tanto para la optimización de la consulta inicial del usuario, como para validar cuáles de las páginas devueltas por los motores de búsqueda empleados son de relevancia para él. Es importante resaltar, que por tratarse de un componente independiente del sistema, este modelo del dominio puede ser fácilmente reemplazado para permitir búsquedas en diferentes áreas del conocimiento. En la figura 1 se muestra la arquitectura del sistema MAESTRA.

El modelo elaborado, contiene los conceptos más importantes en el dominio, organizados en forma jerárquica. Esta clasificación involucra tres aspectos básicos de la temática:

- *Tipos de Mampostería*: Representa la clasificación jerárquica de los diferentes tipos de mampostería existentes, según su comportamiento estructural y/o su función arquitectónica.
- *Elementos Componentes*: Agrupa los conceptos que representan los diferentes elementos que hacen parte de una construcción elaborada siguiendo el método constructivo seleccionado.

- *Materiales*: Agrupa los conceptos relacionados con los materiales que se emplean en una construcción de mampostería.

Este sistema esta formado por cuatro agentes desarrollados sobre la plataforma JADE. Los componentes son:

1. *El agente gestor de Usuario (AGU)*: Este agente es quien presenta al usuario la interfaz de búsqueda y la interfaz que despliega los resultados de la consulta.
2. *El agente gestor de índice (AGI)*: se encarga de mantener actualizado el índice, mediante un proceso de actualización periódico para el cual es programado, recibe las páginas que han sido validadas, verifica si las páginas que recibe están en el índice y adiciona aquellas que no se encuentran indexadas.
3. *El Agente de Recursos Web (ARW)*: es quien ejecuta las consultas optimizadas usando el Google Web APIs service y entrega el resultado al agente de validación. Este agente está a la espera de recibir mensajes con consultas optimizadas para ser ejecutadas sobre el motor de búsqueda que, en este caso, es el Google APIs Service.
4. *El Agente Optimizador de Consulta (AOC)*: es el encargado de seleccionar las consultas que serán enviadas al Agente de Recursos Web, de manera que las consultas que se hagan sobre el motor de búsqueda sean adecuadas en el dominio modelado.
5. *El agente de validación (AV)*: es el encargado de recibir el resultado de la consulta y revisar cuáles de los recursos web recuperados son relevantes o no en el dominio seleccionado.

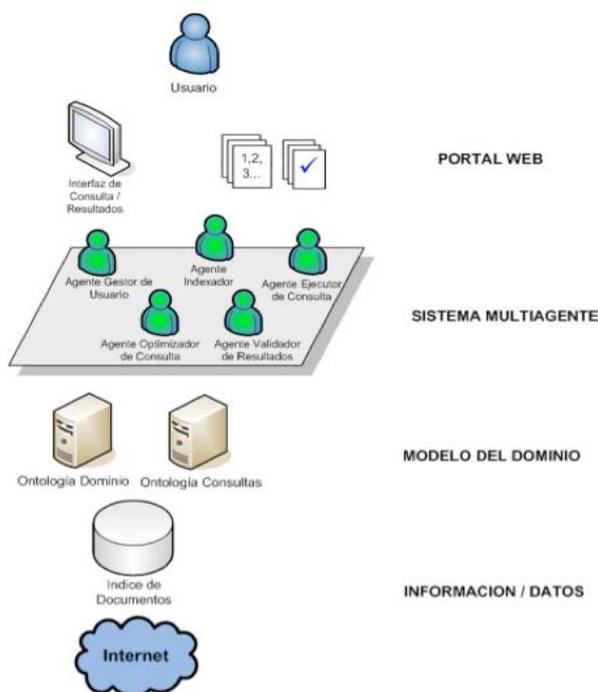


Figura 1. Arquitectura del sistema MAESTRA (Vidal-Rojas, 2006).

Este proyecto tiene un alto potencial de aplicación en aquellas organizaciones donde se tengan grandes repositorios de información en páginas Web. Además de buscar y analizar información específica en sus propios documentos Web en su Intranet, pueden hacer búsquedas más exactas en información que están en Internet. Este proyecto ofrece la ventaja de que el componente del sistema donde se encuentra la ontología es independiente, lo que permite ser utilizada con facilidad en el dominio o dominios de interés que desee la organización.

Entre los beneficios que tendría la empresa si se implementan las ideas de este proyecto estaría la búsqueda eficiente de información, evitando las costosas pérdidas de tiempo en los procesos de análisis de grandes cantidades de información que requieren la intervención de las personas en buscar y seleccionar lo que realmente necesitan.

Proyecto MedIEQ/Web Médica Acreditada

El proyecto europeo MedIEQ (Quality Labeling of Medical Web Content using Multilingual Information Extraction) desarrolla y amplía el trabajo realizado en anteriores proyectos europeos en el campo de la e-Salud y la aplicación de metadatos. Centrándose en temas de calidad de Webs médicas y mostrando el estado actual en la aplicación de tecnologías de rastreo y análisis de contenidos web y extracción multilingüe de la información, y aprovechando la utilización de recursos semánticos y sellos de calidad de dichas webs (Karkaletsis, 2006).

El objetivo es la mejora en la monitorización de las Webs médicas acreditadas así como su identificación y clasificación en áreas temáticas, basándose en ambos casos en metadatos y utilizando siete idiomas diferentes (checo, griego, español, inglés, alemán, finlandés, catalán). Estos metadatos están expresados mediante el estándar RDF/XML lo que permite la integración con herramientas como los motores de búsqueda, que de esta forma serán capaces de “entenderse” con los usuarios al utilizar palabras clave con contenido semántico en el proceso de búsqueda y recuperación de esta información. Además se integran tecnologías de extracción automatizada de contenidos que permitan la simplificación de tareas de revisión y control así como la creación de nuevos recursos de información relacionados.

Desde Web Médica Acreditada se revisan los contenidos de la Web y se estudia su adaptación a las recomendaciones de calidad. Se realiza un informe que se envía al responsable de la Web para que, si es el caso, se realicen las adaptaciones correspondientes y poder así completar el proceso de acreditación. Una vez completado el proceso de acreditación se envía un código HTML para que aparezca el sello de acreditación en la web. A este sello se le asocia un archivo en formato XML/RDF que describe las características básicas de dicha web. Posteriormente se realiza una revisión anual de la web acreditada. Dicha información en RDF queda almacenada en la base de datos de Web Médica Acreditada. En la figura 2 se muestra la arquitectura de este sistema.

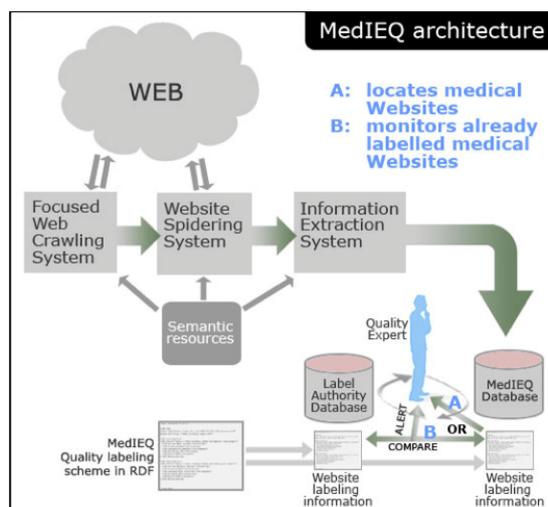


Figura 2. Arquitectura del sistema MedIEQ (Karkaletsis, 2006).

La optimización de resultados y los criterios de búsqueda utilizados se constituyen como los elementos básicos para la obtención de la información acorde con las necesidades de cada búsqueda. En el proyecto MedIEQ los sistemas de extracción de información en diferentes idiomas, se basan en descriptores estandarizados y conceptos semánticos del RDF Schema que se propone. Al compartir definiciones comunes (RDF Schema de metadatos) entre humanos y máquinas (buscadores) se puede garantizar una coincidencia entre los conceptos humanos y los términos de significado semántico con un entendimiento real entre ellos mejorando el resultado de las búsquedas y descripciones de los contenidos Web.

Lo planteado en este proyecto puede tener además otras aplicaciones en las organizaciones como la estandarización de contenidos de las páginas Web de la Intranet en cuanto a la estructura, formatos y descriptores estandarizados, que contengan una serie de elementos mínimos que aseguren la calidad del contenido y la clasificación de las páginas en áreas temáticas basándose en metadatos.

Entre los beneficios que pueden obtenerse de implementar las ideas planteadas en este proyecto está la homogenización en cuanto a la estructura de los contenidos de las páginas Web de la organización, compartir datos y la facilitar la búsqueda de información.

Proyecto EKOSS

El sistema Expert Knowledge Ontology based Semantic Search (EKOSS) permite a los creadores del conocimiento construir descripciones enriquecidas semánticamente de su conocimiento con un mínimo esfuerzo y sin tener conocimiento de las tecnologías de la Web semántica (Ekoss, 2006).

Cuando se registra conocimiento en el sistema EKOSS los datos estructurados en descripciones semánticas basadas en una ontología se pueden registrar al mismo tiempo que el texto en lenguaje natural que expresa el conocimiento que se carga. Cuando se realiza una búsqueda del conocimiento representado de esta manera, un usuario puede crear

consultas semánticamente significativas con la finalidad de recuperar exactamente el conocimiento que se ajusta más a lo que el usuario requiere.

Los sistemas de bases de datos existentes actualmente pueden manejar como máximo 100 conceptos categóricos. Mientras que EKOSS se basa en una ontología completamente funcional que puede fácilmente clasificar miles de conceptos a través de la inherente estructuración lógica de las relaciones que emplea para definir una ontología. De esta manera EKOSS puede expresar descripciones semánticas de un mayor rango de conceptos de conocimiento de manera más flexible que otros sistemas. Esto debido a que el usuario puede utilizar la estructura de la ontología existente para navegar un gran número de conceptos representados para encontrar el concepto en la ontología más cercano al concepto particular que el usuario desea expresar. Mientras que otros sistemas de bases de datos generalmente no tienen un esquema claramente definido o una estructura semántica de las categorías de los datos contenidos en la base de datos (en algunos casos existe un esquema implícito en el diseño de la base de datos, pero en algunos casos no hay ninguna estructura semántica en la base de datos) EKOSS tiene una estructura semántica claramente definida mediante los conceptos y relaciones semánticas usados para expresar los contenidos del sistema. Esta es la razón por la cual EKOSS puede proporcionar resultados más exactos En la figura 3 se muestra la página Web del portal de EKOSS.

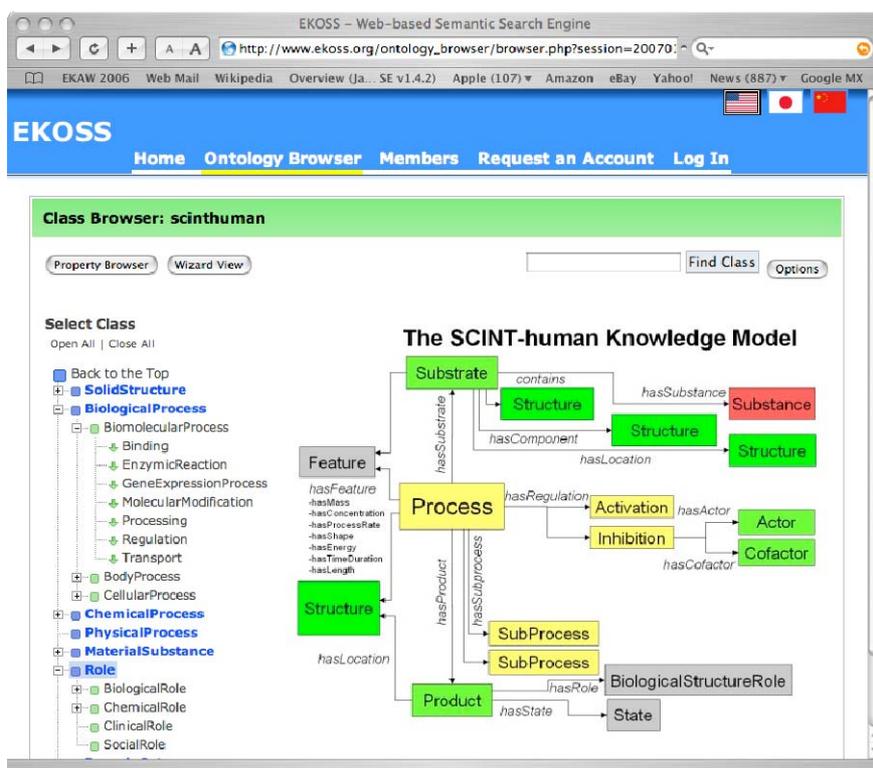


Figura 3. Página web del portal EKOSS (Ekoss, 2006).

Con lo planteado en este proyecto EKOSS se pueden desarrollar en las organizaciones sitios web que proporcionen un ambiente colaborativo para compartir

conocimiento en las diversas áreas de la empresa en la que los expertos pueden documentar, almacenar y compartir la experiencia que ellos han desarrollado en diversos recursos de conocimiento como artículos de investigación, documentación de casos prácticos, aplicación de mejores prácticas, entre otros. Toda esta información estaría organizada por medio de la ontología.

Entre los beneficios estaría que los usuarios podrían buscar de manera más eficiente y con mayor exactitud la información solicitada dentro de los recursos de conocimiento usando la inferencia lógica basada en la ontología de dominio guardada en el sistema.

Proyecto Gnowsis/Semantic Desktop

La naturaleza recolectora de los seres humanos en la era digital se refleja en las colecciones de páginas web, documentos, mensajes de correo o notas que se van acumulando en su computadora. Esta acumulación de información no debería representar inconveniente alguno, nada más alejado de la realidad. Gnowsis es el prototipo de un *escritorio semántico* enfocado principalmente a la integración de herramientas del escritorio del sistema operativo y la información que estas manejan a través de tecnologías de la Web semántica. De tal forma que principalmente se desean integrar aplicaciones para los sistemas de gestión personal de información (Gnowsis, 2006).

Los principales componentes de este sistema son: 1) *el servidor gnowsis* que efectúa el procesamiento/almacenamiento de la información y la interacción con las aplicaciones nativas, 2) una aplicación cliente que funciona como interfaz gráfica implementada en Java o como interfaz web. En la figura 4 podemos apreciar la arquitectura general de esta aplicación.

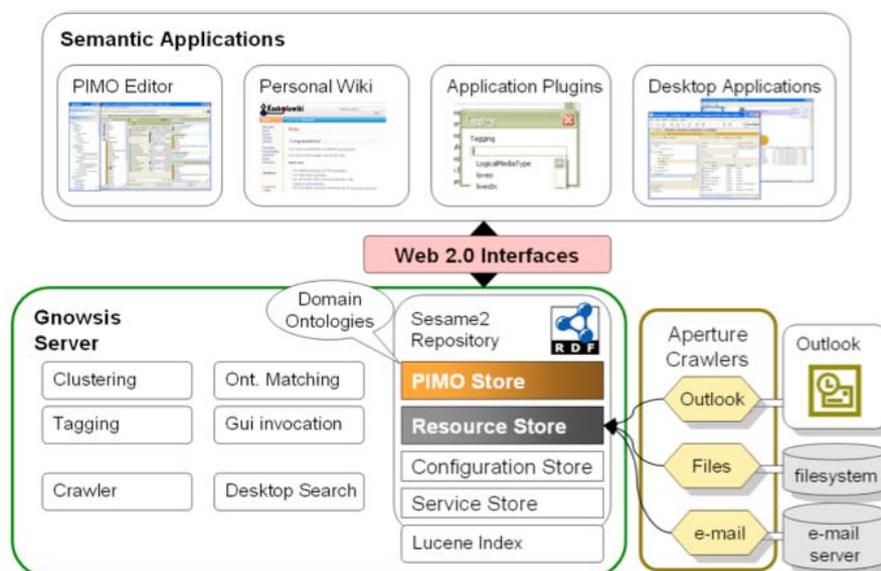


Figura 4. Arquitectura del sistema Gnowsis (Gnowsis, 2006).

En particular, este proyecto es para el mejoramiento de los sistemas informáticos que utilizan las personas en las organizaciones, en especial, ayudaría en la gestión de la

información de los usuarios almacenada en la computadora.

Los beneficios que puede aportar este proyecto es que se podría mover fácilmente a través de los archivos, correos electrónicos, contactos y fotos como si se estuviera en un World Wide Web diminuto usando navegadores y motores de búsqueda. Se puede incluir cualquier cosa en el área de favoritos, unir todo y disfrutar navegando sobre los datos.

Proyecto ANICETO

Los expertos del grupo de Analistas Financieros Internacionales AFI, empresa española, generan notas de investigación, análisis del mercados económicos y financieros, información sobre sectores, recomendación sobre oportunidades de inversión. Para darle soporte en el área de tecnologías de la información, AFI crea la empresa Tecnología, Información y Finanzas TIF, la cual ha desarrollado varios portales web afines a los perfiles de los distintos clientes. Para mantener el interés de los clientes, cada uno de ellos puede elegir los contenidos que le sean de utilidad a partir de una carta de categorías (Bravo, 2006).

Algunos de los clientes de AFI sirven como redistribuidores de la información a través de diferentes medios que dan seguimiento a la economía española. Es por esto que es de vital importancia que la información sea proporcionada de manera flexible para permitir su integración, procesamiento y distribución para lograr mayor satisfacción en sus necesidades particulares.

La Universidad Autónoma de Madrid en colaboración con la empresa TIF y otras instituciones académicas desarrollaron el sistema Aniceto. Los objetivos de este proyecto fueron:

- Definir una ontología para el dominio financiero económico.
- Desarrollar gestores de generación de contenidos para mostrar los datos de la ontología definida.
- Desarrollar una aplicación web que permita una navegación semántica en los portales generados “a medida” para los diversos clientes.
- Soporte de búsqueda semántica en los contenidos que aumente considerablemente la precisión de la búsqueda en comparación con las búsquedas tradicionales.
- Inclusión de perfiles de usuario, que permitan una personalización de los contenidos.
- Que no interfiera con el sistema en explotación aunque las fuentes de información sean comunes.

En la figura 5 se muestra la arquitectura y funcionamiento del sistema Aniceto.

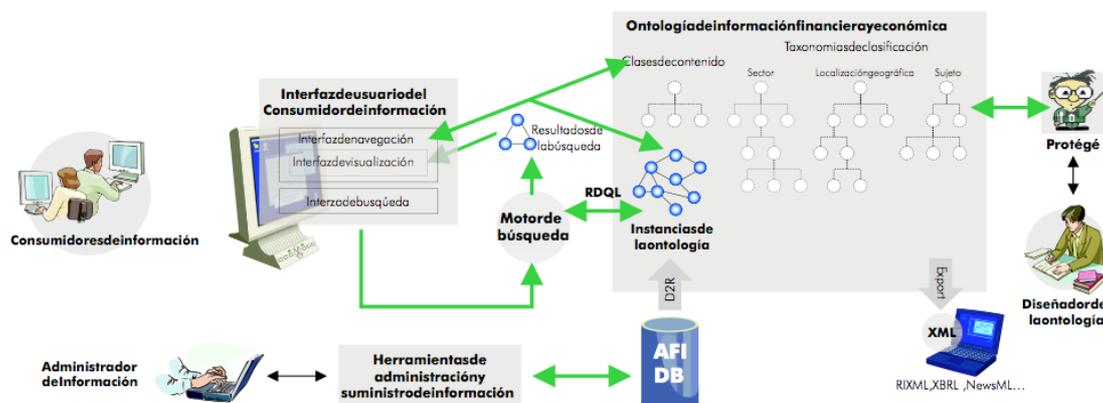


Figura 5. Arquitectura del sistema ANICETO (Bravo, 2006).

Como es evidente, este proyecto es de aplicación práctica en el sector financiero y entre sus principales beneficios están que proporciona información de manera flexible para permitir su integración, procesamiento y distribución.

Conclusiones

En la actualidad, la capacidad que tenga una organización para generar y disseminar conocimiento forma parte de una de las principales ventajas competitivas. Es por esto que es prioritario contar con una infraestructura que apoye la creación y compartición de conocimiento. La existencia de islas de información dificulta su aprovechamiento y disponibilidad para distribuirla. Esta fragmentación de la información como se ha visto se puede reducir mediante el enriquecimiento semántico de la misma a través de la utilización de ontologías. En primer lugar, las ontologías proveen una descripción común y estandarizada acerca de dominio de interés. En segundo lugar, capturan y formalizan el conocimiento ligando el significado que dan las personas a cada concepto permitiendo el procesamiento automatizado de los mismos. De tal manera que las ontologías juegan un papel esencial en la concepción de Web semántica.

En los últimos años, los gobiernos de Estados Unidos y de países de la Unión Europea han reconocido la importancia de la Web semántica y han establecido programas dedicados a financiar investigaciones encaminadas al desarrollo de su tecnología base. Los ejemplos presentados en este artículo demuestran los frutos de este esfuerzo y que con la madurez tecnológica que se tiene actualmente es posible concretizar la visión de la Web semántica. Aunque es evidente que existe un interés de las empresas líderes en tecnología de información (Microsoft, Oracle, IBM, Sun y otras) donde además participan activamente en su desarrollo tecnológico, la gran mayoría de las organizaciones aún no las han aprovechado y explotado.

Por otro lado, la Web semántica no ha tenido una evolución y aceptación en las organizaciones de la misma forma que lo ha tenido la Web actual, esto se debe probablemente a que se requieren conocimientos más especializados para su implementación. Comparadas con las tecnologías tradicionales, que son relativamente sencillas, la complejidad inherente de la Web semántica (a nivel conceptual) hace que la curva de

aprendizaje sea más pronunciada. Adicionalmente, debemos considerar que además de los métodos tradicionales de desarrollo y técnicas de ingeniería de software, se requieren aplicar técnicas de la ingeniería del conocimiento para modelar el conocimiento del dominio de aplicación, para que este pueda ser procesado por el sistema y los usuarios.

Finalmente, si se toman en cuenta todos estos elementos como: las ventajas que ofrecen las tecnologías asociadas a la web semántica de integración de las aplicaciones empresariales, compartir globalmente datos y su facilidad de búsqueda, recuperación y manipulación de manera automatizada; aprovechando los conceptos e ideas vistas en algunos de los proyectos mostrados previamente y que pueden ser la base para el desarrollo de aplicaciones comerciales; e incrementando la capacitación para eliminar algunas de las barreras para la rápida aceptación en las organizaciones, se darían las condiciones para que las tecnologías de la Web semántica se implementen en las empresas que les permita además ser más eficientes y competitivas en la operación y gestión de su información.

Bibliografía

- Abian M.A. (2005). "El futuro de la Web: XML, RDF/RDFS, Ontologías y la Web Semántica", <http://www.javahispano.org>, (12/mar/2006)
- Antoniou G. & Harmelen F.V. (2004). A Semantic Web Primer. MIT press. USA.
- Barcelo-Valenzuela, M.; Sánchez-Schmitz, G.G.A.; Perez-Soltero A. (2006). "La Web Semántica como apoyo a la Gestión del Conocimiento y al Modelado Organizacional", *Revista Ingeniería Informática*, Edición No. 12, Chile.
- Berners-Lee T.; Hendler J.; Lassila O. (2001). "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", *Scientific American Magazine*, May 2001.
- Borst, W. N. (1997). Construction of engineering ontologies for knowledge sharing and reuse. *PhD Thesis*. University of Twente, The Netherlands, Enschede
- Bravo J. Et al. "Aplicación de Tecnologías de la Web Semántica a la Gestión de Información Financiera y Económica". *V Congreso en Interacción Persona-Ordenador (Interacción 2004)*. Lleida, mayo 2004, pp. 326-329.
- Gnowsis (2006). Gnowsis Semantic Desktop Environment. <http://www.gnowsis.org/>
- Gruber, T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human- Computer Studies*, 43(5-6): 907-928.
- Ekoss (2006). "EKOSS - A Web-based Semantic Search Engine for Engineering Knowledge". <http://www.ekoss.org>.
- Jasper, R. y M. Uschold. (1999). "A Framework for Understanding and Classifying Ontology Applications". 12th Workshop on Knowledge Acquisition Modeling and Management KAW'99.
- Jhingran, A. D., N. Mattos y H. Pirahesh. (2002). "Information Integration: A Research Agenda". *IBM Systems Journal*, Vol. 41. No. 4. pp. 555-560.
- Karkaletsis V. y Mayer M.A. (2006). "Quality Labelling of Medical Web Content: the MedIEQ proposal", *eHealth 2006 High Level Conference and Exhibition*. Málaga, España.
- Leyman, F. y D. Roller. (2002), "Using Flows in Information Integration". *IBM Systems Journal*, Vol. 41. No. 4. pp. 732-742.
- Noy, N. (2004). "Semantic Integration: A Survey of Ontology-Based Approaches", *ACM SIGMOD Record*, Vol. 33 No. 4, pp. 65-70.
- Payne T. R., Singh R. y Sycara K. (2002). "Calendar Agents on the Semantic Web". *IEEE Intelligent Systems*, Vol. 17 No. 3, pp84-86.
- Pollock, J. (2002a). "Integration's Dirty Little Secret: It's a Matter of Semantics", Whitepaper, *Modulant: The Interoperability Company*.

- Pollock, J. (2002b). "The Web Services Scandal: How Data Semantics Have Been Overlooked in Integration Solutions", *eAI Journal*, Agosto, 2002. pp. 20-23.
- Stojanovic N. & Handschuh S. (2002). "A Framework for Knowledge Management on the Semantic Web", <http://www2002.org/CDROM/poster/130.pdf>
- Uschold, M. y M. Gruninger. (2004) "Ontologies and Semantics for Seamless Connectivity", *ACM SIGMOD Record*, Vol. 33 No. 4, pp. 58-64.
- Vidal-Rojas et al. (2006). "MAESTRA - Sistema Multiagente basado en Ontologías para Optimizar la Recuperación de Información". *IV Congreso Iberoamericano de Telemática (CITA 2006)*. Monterrey, NL. México
- W3C: World Wide Web Consortium (2001), "Semantic Web", <http://www.w3.org/2001/sw/>, (12/mar/2006).
- Wikipedia (2006). Web Semántica, http://es.wikipedia.org/wiki/Web_sem%C3%A1ntica, (5/dic/2006).