

Semántica e interoperabilidad de procesos

Félix O. Fernández Peña

Correo electrónico: felix@ceis.cujae.edu.cu

Yariel Ramos Moreno

Correo electrónico: yramosm@ceis.cujae.edu.cu

Alejandro Rosete Suárez

Correo electrónico: rosete@ceis.cujae.edu.cu

Instituto Superior Politécnico José Antonio Echeverría, Cujae, Ciudad de La Habana, Cuba

Resumen

Diversos procesos de gestión de conocimiento apoyan la labor educativa en la universidad. Este elemento es de particular importancia en el momento actual de la educación superior en Cuba, involucrada en transformaciones estructurales de la enseñanza, con el objetivo de apoyar el desarrollo de la sociedad de la información en el país. En este trabajo se presentan detalles técnicos de la concepción de un esquema para la integración de múltiples herramientas de gestión de información que apoyan la enseñanza y más allá, la labor educativa integral universitaria en el contexto cubano. La propuesta se basa en la definición de una ontología de dominio que describe formal y explícitamente la semántica de los intereses de estudiantes y profesores en pos de facilitar los procesos de consolidación de información distribuida en el campus universitario.

Palabras clave: gestión de la calidad, análisis de riesgos, AMFE, laboratorios biológicos, salud ocupacional, sistemas integrados de gestión

Recibido: octubre 2009

Aprobado: diciembre 2009

Publicado: Ingeniería Industrial, Vol. XVIII, No. 3, 2007*

INTRODUCCIÓN

La sociedad de la información se desarrolla de manera sostenida a nivel mundial cubriendo todas las áreas institucionales de la sociedad, entre las cuales es prioritaria la educación en general y la educación superior en particular. De hecho, las NTIC están influyendo de manera significativa en la Educación Superior en todas sus funciones y áreas de gestión, y han permitido además el desarrollo de programas académicos que incorporan en algún grado procesos virtuales. [1]

Se hace indispensable lograr una preparación cada vez más integral en los egresados universitarios; profesionales creativos, generadores de ideas ante cambios potenciales,

capaces de lograr resultados prácticos en las condiciones cambiantes del mercado. La Conferencia Mundial sobre Educación Superior -CMES-, celebrada en París en 1998, bajo los auspicios de la UNESCO, reconoció ampliamente la contribución fundamental de las NTIC a la calidad y pertinencia de la gestión de la información y el conocimiento que se realiza en el ámbito de la educación superior, además de que la administración haciendo uso de las nuevas tecnologías aumenta la eficiencia de los recursos humanos disponibles.

El presente trabajo se ha desarrollado teniendo como hipótesis que la descripción semántica explícita de los recur-

* Este artículo ha sido seleccionado de la base de publicaciones del Instituto Superior Politécnico José Antonio Echeverría, sometido a revisión técnica y a una edición mejorada para su publicación en Revista Cubana de Ingeniería.

Los sistemas informacionales en sí mismos facilita: El funcionamiento de los procesos de integración entre aplicaciones de gestión de información distribuidas y la generación de conocimiento a favor de los usuarios de tales sistemas, mediante esquemas más naturales para la comunicación hombre-máquina.

Sobre la base de los lenguajes ontológicos surgidos a partir de la propuesta de desarrollar una web semántica, definida por Tim Berners-Lee como la web descrita en sí misma de una forma que facilite a los programas trabajar con ella, [2] se ha propuesto como objetivo definir un modelo de componentes lo más sencillo posible que permita la gestión semántica de recursos informacionales en pos de la interoperabilidad de procesos de gestión.

El presente trabajo describe los sistemas de gestión de información que sirvieron como punto de partida, y sus características. Así mismo fundamenta el uso de ontologías para la descripción semántica de los recursos informacionales en el entorno de la interoperabilidad de procesos. También se describe la solución propuesta en el entorno de aplicación en que se desarrolla como caso de estudio, para finalmente exponer los principales resultados alcanzados y arribar a conclusiones.

ANTECEDENTES

Los sistemas de bases de datos gestionan grandes volúmenes de datos en pos de obtener información valiosa para los usuarios. Estos sistemas se basan en un repositorio de datos con una estructura formalizada. El modelo relacional de gestión de bases de datos establece, formalizado matemáticamente, el universo del discurso en función de categorías de tipos de datos establecidas, y relaciones existentes entre los elementos de datos definidos. Las interacciones entre las relaciones son descritas a su vez de forma explícita en cuanto a cardinalidad y grado de dependencia entre ellas cuando se disparan eventos de actualización de los datos.

La gestión de la información en pos de los usuarios depende del funcionamiento del gestor de la base de datos y está en función de las reglas que restringen el dominio de los datos y la política de actualización de los datos, y un conjunto de reglas del negocio que son las que determinan la utilidad de un dato en específico. Sin embargo, estas reglas están embebidas en el código de los sistemas de manipulación de los datos, resulta muy complejo separarlas y por lo general, darles mantenimiento también se convierte en una tarea engorrosa (figura 1).

Como puede apreciarse en dicha figura, el significado de los datos está implícitamente formalizado en la estructuración del sistema. Su uso en la generación de información depende de procesos programáticos internos y la estructuración de la validación, verificación, y monitorización del comportamiento de los datos depende de la política que se fije para el desarrollo de la aplicación.

En el ámbito de la intranet del Centro de Estudios de Ingeniería de Sistemas (CEIS) del Instituto Superior Politécnico José A. Echeverría, se han implementado múltiples sistemas de apoyo a la educación presencial y semipresencial. Existen aplicaciones para el intercambio de materiales complementarios a las clases, [3] entre estudiantes y profesores (figura 2), servicios de consulta en línea donde también se publican recursos de interés para la labor académica incluyendo servicios de noticias, [4] entre otros. La característica común de estos sistemas radica en que todos, de una forma u otra, gestionan recursos informacionales. *

De hecho, la estructura de las bases de datos de muchos de estos sistemas tienen múltiples puntos en común; sin embargo, la integración de los servicios de información que brindan resulta una tarea harto complicada puesto que cada aplicación ha sido definida de forma independiente sin considerar los puntos comunes existentes.

En el año 2004, el CEIS comenzó a incursionar en el campo del uso de los lenguajes ontológicos desarrollados por las comunidades de web semántica, dígame lenguaje RDF, RDFS, y OWL, cada uno de los cuales ofrece diferentes primitivas de modelación, y por tanto, diferentes niveles de complejidad. [5] Definidas por Gruber como una especificación formal y explícita de una conceptualización compartida, [6] las ontologías constituyen una definición formal de relaciones entre conceptos, a partir de taxonomías y un conjunto de reglas de inferencia. [7]

Si bien la definición de OWL como lenguaje para la creación de bases de conocimientos sigue los principios de Frames, es importante reconocer que su capacidad de garantizar la persistencia de las reglas y hechos en documentos XML es un principio básico para la interoperabilidad e integración de sistemas dispares en el contexto de un ámbito de aplicación. [8]

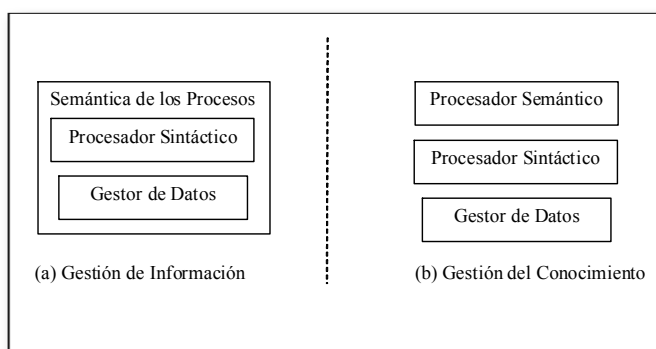


Fig. 1. La gestión de la semántica de la información en sistemas de gestión de información y gestión del conocimiento.

* Se considera recurso informacional a cada unidad básica indivisible de información que conforma la base de información del sistema del que forma parte. [9]



Clases			
Nombre	Fecha	Ver	Desc
Conferencia 1	13-Dic-2001, 1:57 PM	ppt	rar

Ejemplos			
Nombre	Fecha	Ver	Desc
Recordset y Navegador en PHP	7-Dic-2001, 4:40 PM		rar

Bibliografía			
Nombre	Fecha	Ver	Desc
DHTML de Microsoft	14-Dic-2001, 1:50 AM	chm	rar
JavaScript (Netscape)	14-Dic-2001, 1:48 AM	html	rar
JScript (Microsoft)	14-Dic-2001, 1:46 AM	chm	rar
MySQL Reference Manual 3.23.33	12-Dic-2001, 5:43 PM	html	gz
DynDUO	12-Dic-2001, 3:45 PM	html	

Install			
Nombre	Fecha	Ver	Desc
MySQLFront	8-Dic-2001, 1:40 PM		

Fig. 2. Sistema para el intercambio de materiales complementarios .[3]

Entre los trabajos que demuestran la aplicabilidad de los lenguajes ontológicos para la catalogación de recursos informacionales en dominios específicos se encuentran los de Hyvönen, [10] que propone el uso de una ontología de dominio para la categorización semántica de los piezas de un museo, Hendler [11] que propone una estrategia de gestión de la semántica en sistemas distribuidos basado en agentes, y otros como Willmott [12] que diferencian ontologías de lenguaje como subconjunto propio dentro de las ontologías de dominio para describir explícitamente la semántica de la capa de comunicación en aplicaciones distribuidas basadas en servicios.

El impedimento fundamental encontrado en las propuestas analizadas está asociado a la dificultad del mantenimiento de las ontologías que constituyen la descripción semántica formal y explícita de las fuentes documentales. Una mala categorización de los recursos informacionales disponibles, lejos de hacer más natural el proceso educativo, puede traer consigo la interpretación errónea de la información, y la actualización de dicha categorización para un volumen grande de recursos informacionales lo que ocasiona una cantidad de trabajo significativo.

Para la gestión de información en el orden académico, tal y como es el interés en la investigación actual, no se conoce de la existencia de ontologías que describan el entorno de aplicación y permitan la catalogación de la información de que se dispone por la explotación del sistema; por lo que se propuso: La creación de una ontología que responda a los intereses de gestión de información de los usuarios del sistema; la catalogación de la documentación existente, y la adaptación de los mecanismos de recuperación, organización, filtrado y relación en función de garantizar una gestión más eficiente de la información.

SOLUCIÓN PROPUESTA

Para considerar un mayor grado de flexibilidad de los sistemas de gestión, dígame que a partir de fuentes informacionales se genere o infiera nueva información que determine cambios en el comportamiento del sistema de

que se trate, se requiere un mayor nivel de abstracción de la aplicación (figura 1), tal y como se definen los sistemas de bases de conocimiento.

La ventaja del uso de ontologías en este contexto está dada por las implicaciones que tiene el uso de una especificación XML en vez de una estructura propietaria cerrada sobre la capacidad de interoperabilidad que se puede alcanzar entre los componentes y por tanto sobre la interoperabilidad de los procesos.

Creación de la ontología

Para la creación de la ontología se partió de la identificación de intereses de usuarios y la categorización de las fuentes documentales existentes. Cada herramienta de gestión existente se considera un módulo en el nuevo sistema integrado que permite la interoperabilidad basada en la descripción semántica del contexto de aplicación como una única ontología de dominio, según las diferentes áreas de impacto que permiten manipular la información acorde con el tipo de usuario de que se trata. Se consideró la posible integración de tres sistemas de recursos informacionales, dígame:

1. Sistema de gestión del proceso docente-educativo, que incluye los servicios del sistema cuya interfaz se muestra en la figura 1, dedicado a servir como repositorio de materiales complementarios a la educación presencial, y otros sistemas que sirven para divulgar la labor de procesos de enseñanza extracurricular.

2. Sistema de noticias, que mantiene actualizado a estudiantes y profesores en las novedades tecnológicas de actualidad.

3. Servicio de control de investigaciones encargado de controlar los resultados de investigación del CEIS.

Esta selección se debe a que la concepción inicial de la ontología en función de estos sistemas determina su aplicabilidad en áreas fundamentales del trabajo del centro: Investigación y docencia, y porque dichos sistemas contienen un conjunto apreciable de recursos informacionales, haciendo mayor el valor de la integración de los mismos.

A partir de la descripción de los intereses de los usuarios potenciales del sistema (estudiantes y profesores de la carrera de ingeniería informática) y del conocimiento de los especialistas en gestión del conocimiento consultados se diseñó una propuesta de ontología, que luego fue evaluada en rondas de interacción con expertos. La figura 3 ilustra los conceptos básicos fundamentales que describen el contexto de aplicación.

La gestión de la semántica según los niveles de abstracción propuestos tiene como beneficios que permite la categorización de recursos informacionales, tipos de mensajes y contenido de mensajes de un sistema distribuido en el contexto de los intereses de la aplicación, permitiendo así la definición de métricas que determinan el grado de conocimiento del sistema en los diferentes subdominios de conocimiento y la relevancia de los recursos y temáticas en general, para un momento de ejecución determinado, tal y como se ha definido previamente, [9, 12] y se ilustra en la figura 4.

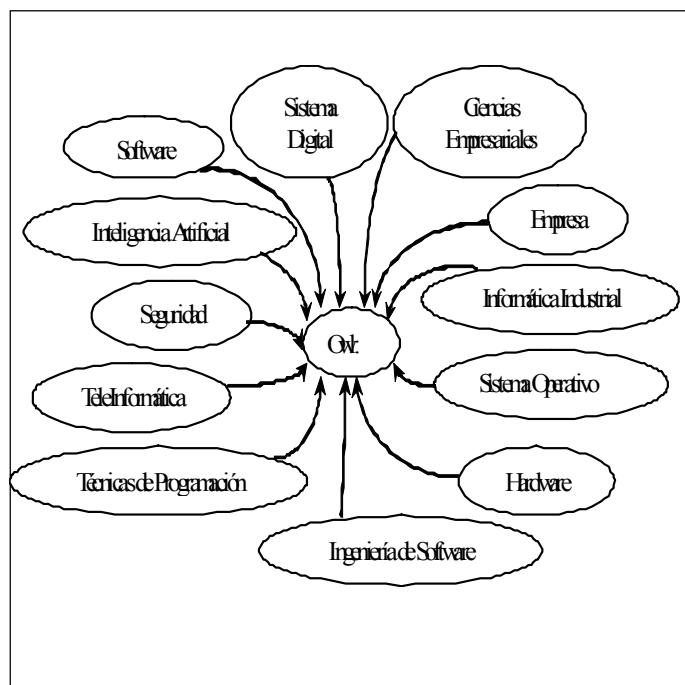


Fig. 3. Conceptos principales a partir de los cuales se creó la ontología.

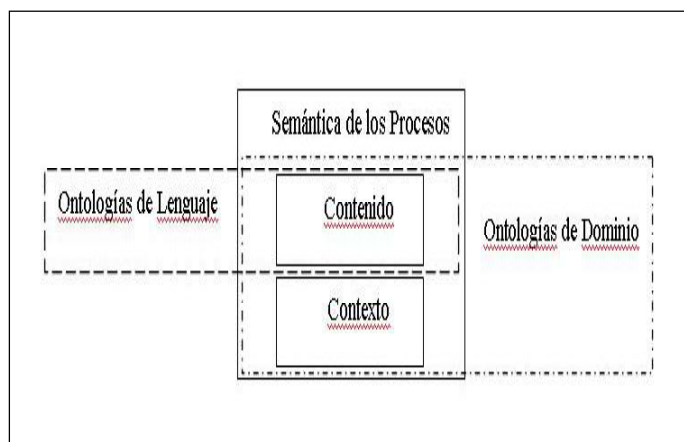


Fig. 4 Niveles de abstracción semántica.

Catalogación de los recursos informacionales

Teniendo como punto de partida la experiencia previa en la descripción semántica de recursos informacionales para un sistema de recuperación de información distribuido [4,9,12], se ha diseñado un esquema único de catalogación de los recursos informacionales existentes en las bases de datos de los tres sistemas utilizados como casos de estudio. Dicho sistema establece las relaciones de pertenencia que existen entre los diferentes recursos informacionales y los términos definidos en la ontología de dominio.

Este proceso es realmente una labor continua durante la explotación del sistema, y constituye un factor decisivo en la efectividad del proceso de interoperabilidad de los sistemas en pos de satisfacer las necesidades globales de información de los usuarios potenciales.

Adaptación de los mecanismos de procesamiento

El incluir un nivel de abstracción superior en la concepción del sistema de gestión determina la aparición de mecanismos para el procesamiento de los recursos informacionales que exploten las nuevas capacidades de descripción de las fuentes informacionales en los diferentes ámbitos en que esta tiene un impacto, dígame: Recuperación, organización, filtrado y relación en función de garantizar una gestión más eficiente de la información.

Esta filosofía de organización y estructuración de la información es válida para los tres tipos de aplicaciones incluidos en el caso de estudio. Es por ello que la capa de abstracción de aplicación de la propuesta incluye:

1. Un componente Java a través del cual se gestiona la descripción de recursos informacionales haciendo uso de la ontología de dominio definida.

2. Un componente Java a través del cual se generan las vistas semánticas asociadas a la navegación por los conceptos cuya semántica se hace explícita con la ontología de dominio utilizada.

3. El uso de Jena como biblioteca de clases que permite la interacción con una maquinaria de inferencia OWL específica para el razonamiento.

4. El encapsulamiento de las funcionalidades en servicios cuya interfaz se hace pública según los requerimientos de donde se vaya a utilizar la propuesta.

De esta forma, la integración semántica de las diferentes aplicaciones se realiza en forma homogénea (figura 5). Estos procesos están orientados a las capacidades de predicción y razonamiento basados en las reglas incluidas en la ontología de dominio utilizada, y que complementan mecanismos de búsqueda y recuperación de información tradicionales haciendo más natural la comunicación hombre - máquina. Se trata de que haya una interacción más natural entre el proceso automatizado de organización de la información y el ser humano que interactúa con este.

La figura 5 muestra cómo sería la navegación por la aplicación que permite la interoperabilidad de los sistemas basados en la ontología de dominio definida.

En la parte izquierda de la figura 6 se visualiza la estructura jerárquica que describe el contexto de aplicación -vista semántica-, y que constituye un menú arbóreo para la navegación por los recursos informacionales disponibles. Para cada elemento que conforma dicha estructura, se controla el conjunto de propiedades que lo caracterizan y las relaciones que guarda con otros recursos del sistema. Dichas propiedades se muestran como descripción al usuario, dándole información adicional sobre el área de conocimiento activa en el sistema para la recuperación de recursos informacionales. Semánticamente asociados a los recursos informacionales que se muestran como noticias en la interfaz cliente, se encuentra un conjunto de recursos que se recomienda debido a su relación con el contenido principal de la aplicación en el contexto activo de la misma.

Así mismo, el proceso de actualización de las relaciones entre los recursos informacionales y las entidades que componen la ontología de dominio es otro aspecto cuyo mantenimiento encarece el ciclo de vida del sistema sustancialmente. Se requiere la evaluación de mecanismos que permitan desarrollar dicho proceso de forma semiautomática.

La gestión de la descripción semántica explícita constituye un nivel de abstracción superior en la concepción de los sistemas de gestión de información, que determina un acercamiento hombre-máquina importante a la vez que implica un incremento en la complejidad de los algoritmos de actualización de la información. Con el desarrollo de este trabajo se ha podido comprobar que reviste vital importancia el desarrollo de mecanismos que favorezcan la actualización dinámica de los sistemas que permitan el mantenimiento por módulos de los mismos con el mínimo de afectaciones sobre los esquemas de interacción con los usuarios finales.

dad de desarrollar herramientas que faciliten el proceso de la creación y mantenimiento de ontologías, así como de las bases de conocimiento que se crean a partir de los esquemas de producción definidos en las ontologías. La definición de estas estructuras es tan solo un primer paso en la concepción de sistemas que logren una verdadera interoperabilidad en un nivel de abstracción mayor, en que los componentes de software colaboran con un mayor grado de libertad que los sistemas tradicionales sintácticos.

1. "Declaración de Quito sobre el Rol de las Universidades en la Sociedad de la Información". En *Universia: Seminario sobre Universidades Virtuales en América Latina y el Caribe*. Iesalc-

- The image shows a screenshot of a website's 'EMPRESA' (Company) section. The page layout includes a top navigation bar with a search box and a sidebar on the left with various menu items. The main content area displays several news articles. Red annotations are used to highlight specific elements:

 - Vistas:** A red box highlights the 'Vista Actual' dropdown menu in the sidebar, with an arrow pointing to it from the label 'Vistas'.
 - Descripción:** A red box highlights the text of an article titled 'Symantec cierra año fiscal 2006', with an arrow pointing to it from the label 'Descripción'.
 - Recomendadas:** A red box highlights the 'Recomendadas' (Recommended) section in the sidebar, with an arrow pointing to it from the label 'Recomendadas'.
 - Noticias:** A red box highlights the main list of news articles, with an arrow pointing to it from the label 'Noticias'.

The sidebar menu items include: Vista Actual, Noticias, Tecnología, Economía, Medio Ambiente, Salud, Deportes, Cultura, Opinión, and Más. The main article list includes headlines such as 'Symantec cierra año fiscal 2006', 'Verbatim presenta sus nuevos USB HD Drives', and 'Siemens Networks propone conectividad avanzada'.

Principales dificultades encontradas

Las dificultades fundamentales encontradas están asociadas a la dificultad del mantenimiento de las ontologías que constituyen la descripción semántica formal y explícita de las fuentes documentales. Una mala categorización de los recursos informacionales disponibles lejos de hacer más natural el proceso educativo puede traer consigo la interpretación errónea de la información, y la actualización de dicha categorización para un volumen considerable de recursos informacionales trae consigo un volumen significativo de trabajo. Próximos trabajos investigativos han de estar enfocados a tales dificultades.

6. GRUBER, T. R.: "A Translation Approach to Portable Ontology Specifications". *Knowledge Acquisition*, vol. 5, nº2, p.199-220, 1993.
7. FENSEL, D.: *Spinning the Semantic Web*, 2003. Consultada: 10/01/2005:
<<http://mitpress.mit.edu/books/chapters/0262062321intro1.pdf>>.
8. FERNÁNDEZ PEÑA, F. y FERNÁNDEZ SAÚCO, I. "Sistema automatizado para la entrada de datos offline sobre la base de la tecnología XML (EXDES)". En *II Congreso Internacional de la Sociedad de la Información y el Conocimiento*, Madrid, España, mayo, 2003.
9. FERNÁNDEZ PEÑA, F.; WILLMOTT, S. y ROSETE SUÁREZ, A. "Gestión de Recursos de Información basada en la Semántica de la Intención de los Mensajes". En *III Jornadas para el Desarrollo de Grandes Aplicaciones de Red*, Alicante, España, mayo, 2006.
10. HYVONEN, E.; SAARELA, S. and VILJANEN, K. "Application of Ontology Techniques to View-Based Semantic Search and Browsing". En *First European Semantic Web Symposium*, 2004.
11. HENDLER, J.: "Agents on the Web". *IEEE Intelligent Systems. Special Issue on the Semantic Web*, March/April, 2001 vol.16, nº 2.
12. WILLMOTT, S. et al. *Adapting Agent Communication Languages for Semantic Web Service Inter-Communication*. En *Proceedings of the IEEE/WIC/ACM WI-2005*, 2005.

AUTORES

Félix O. Fernández Peña

Ingeniero Informático, Doctor en Ciencias Técnicas, Profesor Auxiliar, Centro de Estudios de Ingeniería de Sistemas (CEIS), Instituto Superior Politécnico José Antonio Echeverría, Cujae, Ciudad de La Habana, Cuba

Yariel Ramos Moreno

Ingeniero Informático, Máster en Informática Aplicada, Instructor, CEIS, Instituto Superior Politécnico José Antonio Echeverría, Cujae, Ciudad de La Habana, Cuba

Alejandro Rosete Suárez

Ingeniero Informático, Doctor en Ciencias Técnicas, Profesor Titular, CEIS, Instituto Superior Politécnico José Antonio Echeverría, Cujae, Ciudad de La Habana, Cuba

Semantic and Process Interoperability

Abstract

Knowledge management systems support education at different levels of the education. This is very important for the process in which the higher education of Cuba is involved. Structural transformations of teaching are focused on supporting the foundation of the information society in the country. This paper describes technical aspects of the designing of a model for the integration of multiple knowledge management tools supporting teaching. The proposal is based on the definition of an ontology for the explicit formal description of the semantic of motivations of students and teachers in the learning process. Its target is to facilitate knowledge spreading.

Key words: quality management, risk analysis, FMEA, biological laboratories, occupational health, integrated management systems