



## **Sistema Semántico Basado en Agentes Inteligentes Para Recuperación de Recursos En Un Ambiente Distribuido**

### **Semantic System Based on Intelligent Agents for Resource Recovery in a distributed environment**

*Carlos Nelson Henríquez Miranda*<sup>1</sup>

#### **RESUMEN**

En la actualidad se cuenta con un exceso de información inherente a nuestro entorno que no es manejada apropiadamente. Es común encontrar, por ejemplo en la Web hasta un 90% de información inútil, producto del mismo exceso de información y del mal manejo y estructuración de la misma.

Para abordar esta problemática existen tecnologías emergentes que se han dedicado al manejo preciso, eficiente y estructurado de la información. Una de estas tecnologías es la Web semántica, un área nueva que combina la inteligencia artificial y tecnologías Web donde se representa la información adecuadamente, para facilitar la localización y acceso a los datos no solo por humanos sino sistemas inteligentes.

El proyecto consiste en la creación de un sistema semántico elaborado a partir de una arquitectura propuesta compuesta por varias capas. La primera capa encargada de la descripción de los datos (metadatos). La siguiente encargada de las relaciones de los datos y su significado (ontologías). En la capa posterior se construye un sistema de agentes inteligentes que se encarga de obtener la información sobre la capa ontológica de manera ágil y útil. El sistema semántico construido localiza por intermedio de agentes inteligentes recursos ubicado en diferentes equipos conectados en una red de confianza.

Este proyecto es de gran interés no solo para la comunidad académica, sino para el sector corporativo y público, en cuanto a la estructuración de los datos utilizando una organización semántica lo que implica una mejor organización y localización de la información.

*Palabras Clave: Agentes Inteligentes, Web semántica, meta datos, ontología.*

#### **ABSTRACT**

At present days we have overload information inherent in our environment that is not managed properly. It is very common to find, on the Web up, for example, still 90% of useless information, as a product of the excess of information and the poor management and structuring of it.

To argue about this issue there are emerging technologies that have been oriented to precise handling, efficient and structured of the information. One of them, it is the Semantic Web, a new area which combines artificial intelligence and Web technology where information is represented properly, to facilitate the location and access to information not only for humans but intelligent systems as well.

The project involves of the creation of a semantic system developed from a proposed architecture consists of several layers. The first layer is the responsible for the data description (metadata). The next, it is the charged of the relations of data and its meaning (ontologies). In the subsequent layer it is built a system of intelligent agents responsible for obtaining information on the ontological layer in an agile and useful way. The semantic system built is able to locate through intelligent resource agents located on different computers connected in a trusted network.

This project is of great interest not only to the academic community, but for the corporate and public sector

<sup>1</sup> Magíster en Ingeniería de Sistemas y computación – Profesor Universidad Autónoma del Caribe. chenriquez@uac.edu.co,



as well, related to the structuring of the data using a semantic organization which means better organization and location of the information.

*Keywords: Intelligent Agents, Web semantics, metadata, ontology.*

## 1. INTRODUCCIÓN

El problema de la recuperación de documentos siempre ha sido un inconveniente para las empresas, industrias, universidades y en general para todas las organizaciones. Por ejemplo el boom de la Internet ha creado una infinidad de información pública para todo el mundo. Actualmente la Web es un medio que ha cambiado la forma en que vivimos, la forma como nos relacionamos, comunicamos y adquirimos conocimiento. La Web es hoy extremadamente flexible y económica para la comunicación, el comercio y los negocios, ocio y entretenimiento, acceso de información y servicio, cultura entre otras. Pero sin embargo, al mismo tiempo, los factores que han propiciado el éxito de la Web, han originado sus principales problemas: heterogeneidad de fuentes de información, interoperabilidad, poca fiabilidad, infoxicación y acceso restringido [1].

Para abordar esta problemática existen tecnologías emergentes que se han dedicado al manejo preciso, eficiente y estructurado de la información. Una de estas tecnologías es la Web semántica, una nueva área que combina la inteligencia artificial y tecnologías Web donde se propone una representación de la información para facilitar la localización, distribución y acceso a los datos no solo por humanos sino sistemas inteligentes. [2] Esta tecnología ha tenido un gran desarrollo en los últimos tiempos y se espera que sea la tendencia a nivel mundial tanto en la Internet, así como, en el manejo de información en las organizaciones.

El siguiente artículo está estructurado de la siguiente manera: primero se muestra la metodología que se utilizó, seguidamente se presenta la arquitectura y luego su implementación en un caso de estudio. Finalmente se muestra las conclusiones del proyecto.

## 2. ASPECTOS METODOLÓGICOS

El tipo de investigación utilizado en este proyecto es la “Investigación aplicada”, porque el objeto de esta investigación es comprobar la efectividad de las búsquedas de recursos mediante agentes inteligentes

con el desarrollo de un prototipo, este último utilizado para comprobar la arquitectura propuesta.

Para consecución de los objetivos del proyecto, este se dividió en varias fases bien definidas. En la primera fase se hizo una exploración documental sobre las teorías y conceptos redundantes en la investigación. Luego de este paso se examinó el estado actual de las tecnologías sobre Web semántica para el planteamiento de la arquitectura y por último se construyó una aplicación que la implementa.

Para la recolección de datos primarios se utilizó la observación documental, así mismo, se hicieron consultas a expertos sobre el tema que pudieron representar una fuente de confianza para el desarrollo de la investigación. Se revisaron algunas investigaciones locales, como es el caso de “*Modelo Para La Representación Y Visualización Del Conocimiento Basados En Ontologías Web*”, el cual trata de la formulación y diseño de una arquitectura de conocimiento soportada en los conceptos de ontologías Web, integración de tecnologías recomendadas por estándares internacionales como lo es el consorcio de la W3C, para tecnologías como RDFS, OWL-DL, razonadores basados en lógica de descripción y Web Services [3]. Para la recolección de datos secundarios se utilizó los recursos de la Web. Se efectuaron búsquedas relacionadas con el tema, Web semántica y agentes inteligentes, y se consultaron páginas de organizaciones especializadas donde se pudieron encontrar textos, referencias a autores, libros e investigaciones, entre otros. Una de las aplicaciones más interesantes en la actualidad es google Alerts (<http://www.google.com/alerts>), una herramienta que permite hacer automáticamente monitoreo de información para posteriormente mandarla al correo electrónico del usuario solicitante.

## 3. ARQUITECTURA SISTEMA MULTIAGENTE “ARSMA”

La arquitectura de la Web semántica definida por [2] define una serie de capas para el manejo y recuperación eficiente de la información en la Web, pero no puntualiza en como se debería recuperar los datos. Así que según [4] que define un sistema multiagente como parte de la Inteligencia artificial

distribuida (IAD) que estudia el comportamiento de agentes inteligentes para resolver un problema de manera cooperativa, y complementando con otros autores, en los que se destacan: Coppin [5] que dice que “Un agente es una entidad que puede llevar a cabo algunas tareas que son, usualmente, para ayudarnos a los seres humanos” y Mancilla [6] que define una clasificación de agentes se propone la siguiente arquitectura Multiagente.

La arquitectura que da vida al sistema esta compuesta por el desarrollo de cinco (5) capas (ver figura 1) para la realización de un sistema semántico basado en Multiagentes para la descripción, localización, recuperación y acceso a documentos producidos en un ambiente distribuido.

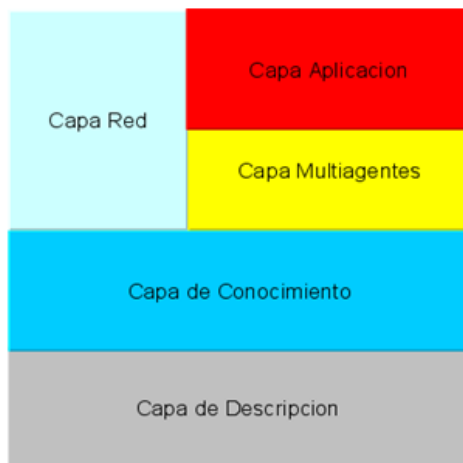


Figura 1. Arquitectura Basada en Agentes

La capa de descripción se encarga de definir el dominio de información del sistema, es decir, que tipo de documento se desea recuperar. Dentro de estos documentos pueden estar libros digitales, artículos, tesis de grado, revistas y en general todo material digital que se requiera manejar. Además debe proporcionar la forma más idónea para describirlo. Para esta descripción lo más recomendado es el uso de metadatos [7], que se deben definir en algún sistema descriptor (dublincore, LOM, Mets, entre otros). Estos esquemas para descripción permiten utilizar lenguajes basados en XML que permite aportar semántica a los contenidos. En aras de buscar una descripción estándar, ha surgido RDF, un lenguaje basado en XML que busca una máxima interoperabilidad entre sistemas inteligentes [8].

La capa de conocimiento define un marco común o estructura conceptual sistematizada, no sólo para

almacenar la información, sino para poder buscarla y recuperarla. Esta capa hace uso de cualquier tipo de representación de información desde archivos planos, base de datos tesauros y ontologías [9]. Como recurso más importante para la Web semántica se recomienda el uso de ontologías. Esta capa también se encarga de definir un lenguaje idóneo para la construcción de la representación escogida, así como también de la selección de las mejores herramientas para su desarrollo. Para hacer una buena elección de recursos es conveniente tener en cuenta las bases recomendadas por la Web semántica respaldadas por estándares internacionales.

La capa de red es la responsable del envío de la información sobre el sistema utilizado en cada caso, brinda los recursos que se deben implementar para transmitir datos a través del sistema. Permite definir la distribución de estructuras de conocimiento y precisa el nivel de confianza entre estos, es decir, para la distribución se necesitan varios nodos ubicados y conectados bajo una red de confianza para poder compartir la información. Esta capa se basa en los protocolos actuales de red (TCP/IP, FTP, HTTP) e involucra aspectos como direccionamiento, configuración de servidores, socket, servicios, clientes entre otros.

La capa de Multiagentes se encarga de definir los agentes adecuados para la localización y recuperación de los documentos. Aquí se establece la forma de comunicación entre los agentes, la relación con la capa de conocimiento y aplicación, y el ambiente de trabajo en la cual se desarrollarán los agentes. Se proponen cinco tipos de agentes para esta operación. El agente *Mediador*, agente *Móvil*, agente *Respuesta*, agente *Espía* y agente *Directorio*.

El agente mediador captura la información a buscar por el usuario proveniente de la capa de aplicación. Su existencia se fundamenta en la necesidad de tener un agente especializado en comunicar al usuario con los agentes encargados de moverse por el sistema en busca de la solicitud. Este agente se comunica con el agente móvil, el cual le pasa un mensaje en un formato determinado.

El agente móvil viaja por todo el sistema distribuido en busca de coincidencias semánticas, luego de la comunicación efectiva con el agente mediador. Este será el encargado de moverse por los nodos de la red en busca de la información que ha

sido solicitada por el usuario. El agente Móvil se encargará de preguntar a los distintos agentes ubicados en los diferentes nodos del sistema distribuido si poseen información relacionada con la búsqueda del usuario. Para moverse, este agente se clonará a sí mismo y enviará a sus clones a buscar. Una vez los clones hayan o no, encontrado la información, regresarán al servidor principal, entregarán su información y se destruirán.

El agente de respuesta es el único agente que se ubica por cada nodo, donde se encuentra el conocimiento almacenado. Estos agentes son los encargados de llevar a cabo el proceso de consulta en el repositorio y buscar información relacionada a la petición del usuario, la cual le ha sido entregada por el agente móvil. Si ha encontrado o no información afín a la solicitada, notificará al agente móvil y este se regresará al servidor principal con los resultados.

Cuando la base del conocimiento se cambia, se hace necesario que el agente de respuestas cargue nuevamente la ontología en memoria. El agente espía, un agente de tipo móvil, se mueve constantemente en los repositorios en busca de cambios. Al haber cambios, avisa al agente de respuestas para que recargue y al agente buscador sobre el cambio en las bases de conocimiento.

Por último un agente directorio que registra los nodos distribuidos en todo el sistema, que son lo que permitirán tener la suficiente confianza para compartir recursos, basados en la capa de red. Este será el que se encargue de tener una lista de los servicios que pueden ser prestados en el servidor principal por los diferentes agentes localizados en él.

Esta arquitectura funciona con un entorno que la hace posible e incluye los recursos para implementar cada una de las capas, las herramientas computacionales para su operacionalización y todo esto soportado o avalado por estándares internacionales. [10]

#### 4. CASO DE STUDIO (IMPLEMENTACIÓN ARQUITECTURA)

El caso de estudio en donde se operacionaliza la arquitectura es un sistema para la recuperación de artículos científicos que se encuentran apostados en revistas pertenecientes a varias facultades de la universidad Autónoma del Caribe. Cada facultad tiene

en su dominio sus propias revistas, que son publicadas para poder acceder a ellas. Los usuarios podrán hacer búsquedas precisas de artículos sin enterarse en donde se encuentran. (Ver figura 2)

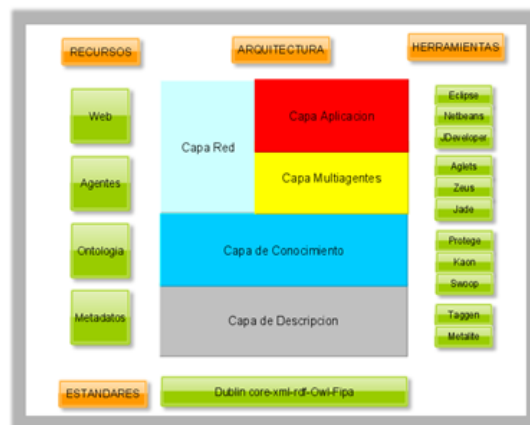


Figura 2. Operacionalización de Arquitectura

A continuación se muestra el proceso de implementación de cada una de las capas.

#### 4.1 DESCRIPCIÓN DE METADATOS.

El dominio de conocimiento como se prevé son los artículos científicos pertenecientes a las revistas de las facultades de la universidad. Se obtuvo las características más importantes de los artículos, y se plasmaron en propiedades representadas en metadatos (ver tabla 1). Se analizaron varios sistemas para describir recursos electrónicos mediante metadatos y se tomó el estándar internacional dublicore (DC). La escogencia del mismo se hizo en un análisis detallado en donde se tuvieron criterios bien definidos (ver figura 3) [11]

Etiqueta del estándar Dublin Core	DESCRIPCIÓN
DC. Title	Título dado al artículo.
DC. Creator	Autor o autores del artículo.
DC. Subject	Palabras claves dentro del artículo.
DC. Description	Breve descripción del artículo.
DC. Type	De igual forma que las revistas los artículos son de tipo texto.
DC. Identifier	Una identificación para los artículos de las revistas.
DC. Source	Fuente de procedencia de los artículos.
DC. Language	Lenguaje en el que se encuentran escritos los artículos.
DC. Relation	Relación del artículo con el recurso general, la revista.

Tabla 1: Etiquetas Dublin core de las Revistas

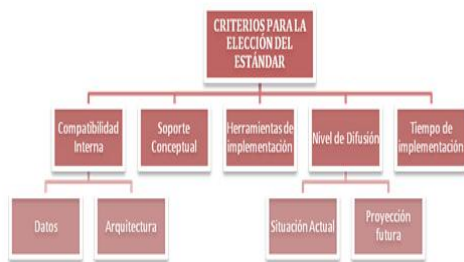


Figura 3. Criterios para elección estándar Metadatos

## 4.2 EXPRESAR CONOCIMIENTO (ONTOLOGÍAS)

Una vez definidos los metadatos se da inicio a la creación de la ontología, esta permite dar significado y definir un conjunto de relaciones y reglas según las propiedades de los documentos. Para poder crear la ontología se debió escoger el mejor lenguaje para trabajarla. En la actualidad existen muchos lenguajes de ontologías. Entre los mas usados podemos encontrar: SHOE, OIL, DAML, DAML+OIL y OWL, como resultado de una evolución constante debido a los avances en la Web semántica (ver figura 4).

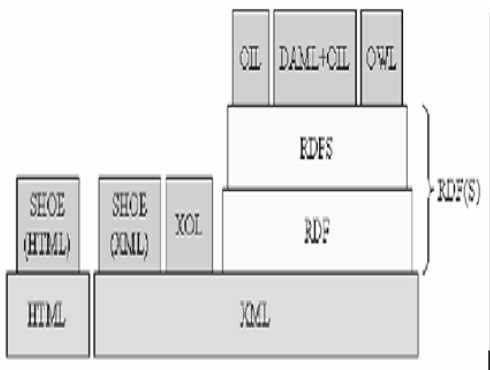


Figura 4. Arquitectura Lenguajes Ontológicos

En este punto fue preciso escoger cual lenguaje aportaba más poder a la hora de elaborar la ontología para nuestro dominio de información, teniendo además en cuenta las tendencias actuales. Se eligió para construir las ontologías a OWL. Esta selección se hace principalmente porque OWL mejora de forma complementaria las deficiencias o desventajas mostradas por los lenguajes anteriores a este, implementado mayor capacidad para expresar las ontologías, mayor facilidad al usarse y reuniendo características compatibles con los sistemas. [12]

Para la elaboración de las ontologías se usa principalmente una herramienta especializada en la construcción de estas, llamada *Protégé*, en su versión 3.3.1. Esta permite crear la ontología a través de una interfaz gráfica especializada; y como producto genera un archivo con extensión OWL (ver figura 5). Además provee la posibilidad de integrar una serie de recursos que le dan robustez a la estructura de esta. Uno de estos recursos son las etiquetas para el tratamiento o clasificación de metadatos basadas en un estándar de recolección de metadatos (Dublin Core) de forma fácil y rápida.

```

- <Revista rdf:ID="Prospectiva">
  <dc:type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PhysicalObject</dc:type>
  <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PROSPECTIVA Una nueva Vision para la ingeniería</dc:title>
- <dc:rights rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
  Los conceptos y opiniones de los artículos son responsabilidad exclusiva de su autor. En ningún momento comprometen las orientaciones de la Facultad de Ingeniería de la Universidad Autónoma del Caribe. Los artículos de esta revista pueden reproducirse citando la fuente.
</dc:rights>
<dc:language rdf:datatype="http://www.w3.org/2001/XMLSchema#string">es</dc:language>
- <dc:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
  El conjunto de artículos que desarrollan nuestros colaboradores abordan diferentes temas de la ingeniería tanto en el campo profesional como en el de la investigación, eficiencia energética y la fatiga como también en el campo de la didáctica de la formación de ingenieros potenciales en investigación condición fundamental para la innovación.
</dc:description>
<dc:publisher rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Editorial Uniautónoma</dc:publisher>
<dc:coverage rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Julio - Diciembre 2006</dc:coverage>
</Revista>
  
```

Figura 5. Apartes de La ontología Construida

## 4.3 RED DE CONFIANZA

En primer lugar, se debe definir una arquitectura adecuada para desarrollar la solución, ésta de acuerdo a la naturaleza de la universidad, la cual es tener agrupados todos los programas por facultad. El sistema deberá estar compuesto por varios nodos, los cuales pueden ser llamados repositorios, donde cada nodo corresponderá a una facultad. (Figura 6)

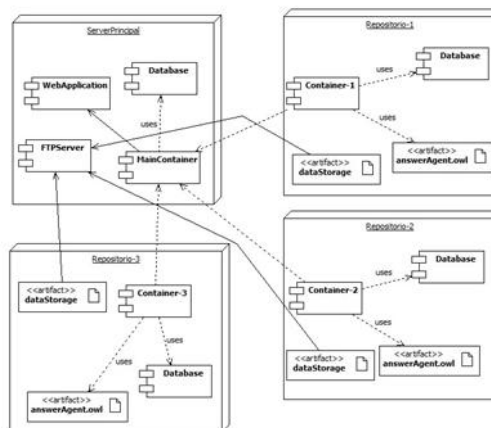


Figura 6. Arquitectura del prototipo

Para la implantación de la arquitectura se creó un prototipo, llamado *SOMBRA*. Este prototipo funciona en una red Lan donde están conectados varios equipos que contienen sus respectivos repositorios, los cuales contienen la información de los artículos. Los nodos representan los repositorios en donde se guardan los documentos. Cada nodo corresponde a cada una de las facultades de la Universidad y tiene una ontología que representa a los artículos científicos, en donde, los agentes inteligentes son capaces de consultar para poder obtener información precisa, relacionada a la petición del usuario. Estos agentes de búsquedas están registrados en el contenedor de agentes del servidor principal, en donde, el ambiente de trabajo (Framework para agentes) los coordina.

#### 4.4 CREACIÓN DE AGENTES Y CAPA DE APLICACIÓN

Para la creación del entorno y ambiente de agentes, se eligió *JADE*. Este tiene muchos años de desarrollo y es muy popular por lo que se encuentra bastante información relacionada con él. Cumple con todos los estándares *FIPA*, lo que permite que se integre fácilmente con otros sistemas multiagentes. Para la escogencia y la implementación de los agentes en *JADE* se hizo un estudio detallado [13].

Para llevar a cabo el proceso de construcción del prototipo (ver figura 7), el sistema se ha dividido en diferentes secciones. Cada sección contendrá diferentes tipos de agentes:

- Agente mediador

El agente mediador será el que reciba la búsqueda que haga el usuario. Su existencia se fundamenta en la necesidad de tener un agente especializado en comunicar al usuario con los agentes encargados de moverse por el sistema en busca de la solicitud.

- Agente Directorio

Este agente será el que se encargue de tener una lista de los servicios que pueden ser prestados en el servidor principal por los diferentes agentes localizados en él.

- Agente Móvil

El agente móvil, como su nombre lo indica, será el encargado de moverse por los nodos de la red en busca de la información que ha sido solicitada por el usuario. Este agente se encargará de preguntar a los distintos agentes ubicados en los diferentes nodos del sistema distribuido si

poseen información relacionada con la búsqueda del usuario. Para moverse, este agente se clonará a sí mismo y enviará a sus clones a buscar.

- Agente De respuesta

El único agente que se ubica en los nodos es el agente de respuestas. Estos agentes son los encargados de llevar a cabo el proceso de consulta de las ontologías y buscar en ellas información relacionada a la petición del usuario, la cual le ha sido entregada por el agente.

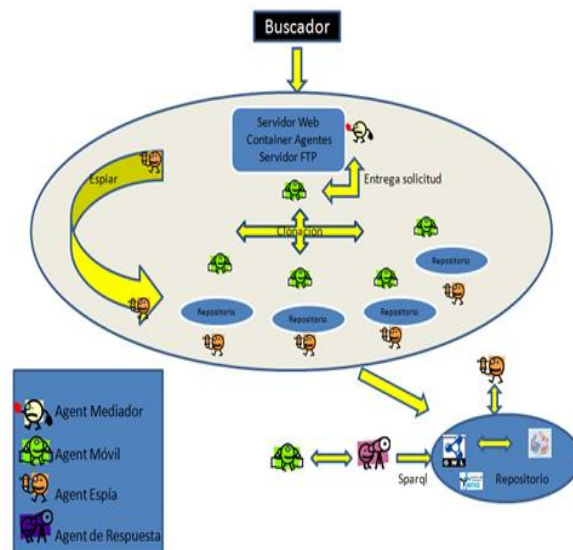


Figura 7. Descripción Prototipo

## 5. CONCLUSIÓN

El modelamiento, diseño e implementación de la arquitectura se hizo en una forma coherente, acertada y efectiva teniendo en cuenta la realización de tres fases completamente identificadas. La primera fase que analizo y construyo los metadatos a partir de las revistas escogidas. La segunda, a partir de estos metadatos, construyo una ontología donde creo una serie de conceptos, relaciones y reglas. En la ultima fase se realizo un sistema WEB de búsqueda basado en agentes sobre la ontología y utilizando consultas semánticas.

La arquitectura propuesta es ideal para la construcción de sistemas de recuperación de documentos en un entorno distribuido independiente del formato del recurso electrónico.

En esta investigación se nota como las ontologías y metadatos ofrecen una gran estructura para la interoperabilidad e intercambio de información, dándole además significado a los recursos, y como los sistemas multiagentes se encargan de las tareas



específicas como búsqueda y la recuperación de Documentos.

La escogencia de un Framework para el desarrollo de soluciones multiagentes es un proceso delicado, y se deben estudiar en profundidad.

El planteamiento de una arquitectura base como la de Tim Berners - Lee en, cuanto a la Web Semántica se refiere, y las recomendaciones hechas por el consorcio W3C, nos hacen crear productos tecnológicos con gran horizonte y calidad, pero que son únicos en su especie y utilizan diferentes enfoques y tecnologías. Este sistema de multiagentes se hace bajo una arquitectura propuesta que define todas las capas a implementar en su desarrollo, y que fácilmente se puede emplear en crear sistemas basados en agentes cooperantes bajo entornos de similares características.

## REFERENCIAS

[1] Gómez L. Las nuevas tecnologías en los procesos de cooperación documental: Aumento de la visibilidad para REDINED. 2007

[2] Lee T. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. 2001

[3] Piñeres M. Modelo Para La Representación Y Visualización Del Conocimiento Basados En Ontologías Web. Tesis presentada a la Universidad del Norte para optar al grado de Magíster en Ingeniería de Sistemas y computación 2008.

[4] Quintero A, Rueda S y Ucros M. Agentes y Sistemas Multiagente: Integración de Conceptos Básicos. Disponible en <http://agamenon.uniandes.edu.co/yubarta/agentes/agentes.htm>

[5] Coppin, B. Artificial Intelligence Illuminated. Estados Unidos: Jones and Bartlett Publishers. 2004

[6] Mancilla L. (2008) ¿Qué son los agentes inteligentes de software? Disponible en [http://energia.guanajuato.gob.mx/gaceta/GacetaIdeas/Archivos/31072008\\_QUE\\_SON\\_AGENTES\\_INTELIGENTES\\_SOFTWARE.pdf](http://energia.guanajuato.gob.mx/gaceta/GacetaIdeas/Archivos/31072008_QUE_SON_AGENTES_INTELIGENTES_SOFTWARE.pdf)

[7] Delcor J. Descripción, Indexación, búsqueda Y Adquisición De Secuencias De Vídeo Mediante Descriptores Mpeg. 2006. Universidad UPC.

[8] W3C Consortium. RDF Primer W3C Recommendation 10 February 2004. Disponible en <http://www.w3.org/TR/REC-rdf-syntax/>

[9] W3C Consortium. Recomendaciones del W3C 2004 [Disponible en <http://www.w3.org/2007/09/OWL-Overview-es>]

[10] Henríquez C. Modelamiento, Diseño e Implementación de una arquitectura basada en MultiAgentes para recuperación de artículos científicos en las facultades de la Universidad Autónoma del Caribe. Tesis presentada a la Universidad del Norte para optar al grado de Magíster en Ingeniería de Sistemas y computación 2009.

[11] Castaño A y Loaiza A.. Descripción De Revistas Y Artículos Científicos Situados En Un Entorno Distribuido, Empleando Un Estándar De Metadatos. Tesis presentada a la Universidad Autónoma del Caribe para optar al grado de Ingeniero de Sistemas 2008.

[12] Ucros d, y Vargas I. Construcción De Una Ontología Web Para Un Sistema Basado En Agentes Inteligentes Para La Descripción, Localización Y Acceso A Artículos Científicos En Un Sistema Distribuido. Tesis presentada a la Universidad Autónoma del Caribe para optar al grado de Ingeniero de Sistemas 2008.

[13]. Florez J y Messino A. Desarrollo De Un Sistema Basado En Agentes Inteligentes Para La Descripción, Localización Y Acceso A Artículos Científicos En Un Sistema Distribuido. Tesis presentada a la Universidad Autónoma del Caribe para optar al grado de Ingeniero de Sistemas 2008.