



# Sistema Multiagente para recuperación de artículos científicos en un Entorno Educativo Distribuido

Carlos Nelson Henríquez Miranda\*  
chenriquez@uac.edu.co

## Resumen

En este artículo se presenta las generalidades de la construcción de un sistema multiagentes basado en las tecnologías de la Web semántica y como resultado de la investigación titulada "Arquitectura basada en multiagentes para recuperación de artículos científicos". Aquí se abordan los aspectos metodológicos y la recolección de datos que dieron paso a la consecución del proyecto. Después se presenta la arquitectura y su descripción capa por capa, obtenida después de analizar las tecnologías de la Web semántica, estándares internacionales e investigaciones previas. Por último, se muestra el proceso de construcción del sistema que toma como base la arquitectura y el cual consiste de un buscador WEB semántico que trabaja en un entorno distribuido usando agentes inteligentes para localizar y recuperar recursos que se refieren a artículos de una revista científica pertenecientes a una universidad.

## Palabras clave:

Agentes Inteligentes, Web semántica, metadatos, ontologías, sistema distribuido

CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

Autor Ingeniero de Sistemas, Magíster en Ingeniería de Sistemas y computación de la Universidad del Norte. Especialista en estudios pedagógicos. Profesor de tiempo completo de la Universidad autónoma del Caribe en la ciudad de Barranquilla Colombia. Consultor General en tecnologías J2SE y J2EE y arquitectura de Software. Investigador Principal del grupo de Investigación Ingeniería de Software y nuevas tecnologías "SINT" en las áreas de Web semántica, Servicios de persistencia, tarjetas Inteligentes, Agentes Inteligentes y Tecnología Java.

## 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 un 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 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

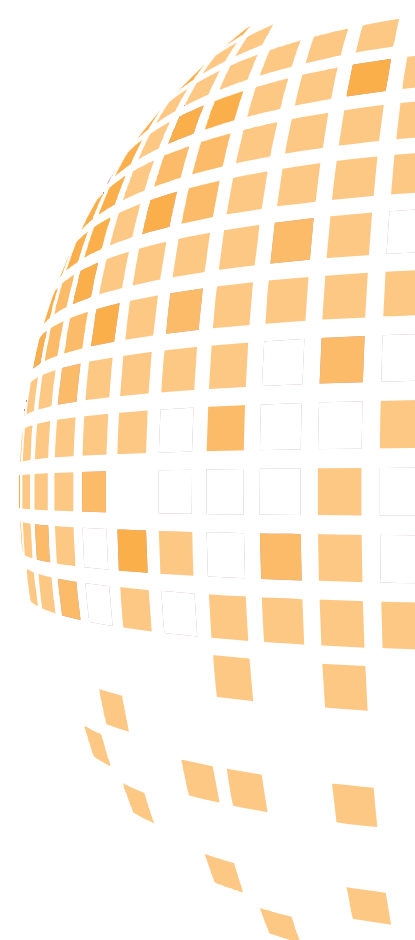
El tipo de investigación utilizado en este proyecto es la “Investigación aplicada”, porque el objeto de esta investigación es comprobar la efectividad de las búsquedas de recursos mediante agentes inteligentes con el desarrollo de un prototipo, este último utilizado para comprobar la arquitectura propuesta.

Para consecución de los objetivos del proyecto, éste 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.

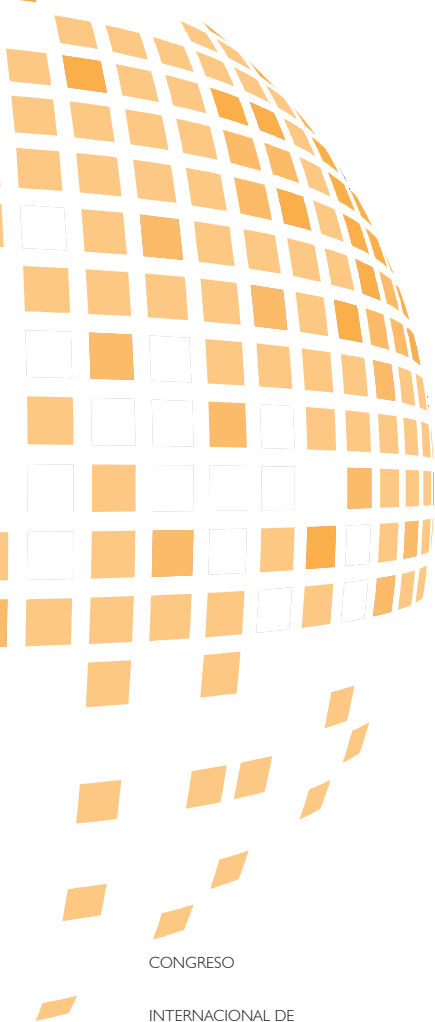
Para la recolección de datos primarios se utilizó la observación documental, así mismo, se hicieron consultas a expertos sobre el tema que pudieron representar una fuente de confianza para el desarrollo de la investigación. Se revisaron algunas investigaciones locales, como es el caso de “Modelo para 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 encontrarse textos, referencias a autores, libros e investigaciones, entre otros.

## 3. Arquitectura Sistema Multiagente “ARSMA”

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



“ESTRATEGIAS DE  
LAS TECNOLOGÍAS DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL”



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

La capa de descripción se encarga de definir el dominio de información del sistema, es decir, qué 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 [4], 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 [5].



Figura 1. Arquitectura Basada en Agentes

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 [6]. 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.

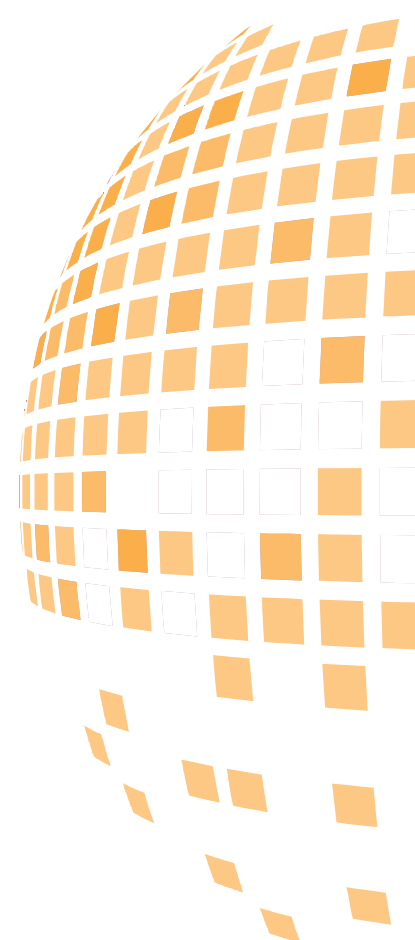
El agente mediador captura la información a buscar por el usuario proveniente de la capa de aplicación. Su existencia se fundamenta en la necesidad de tener un agente especializado en comunicar al usuario con los agentes encargados de moverse por el sistema en busca de la solicitud. Este agente se comunica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

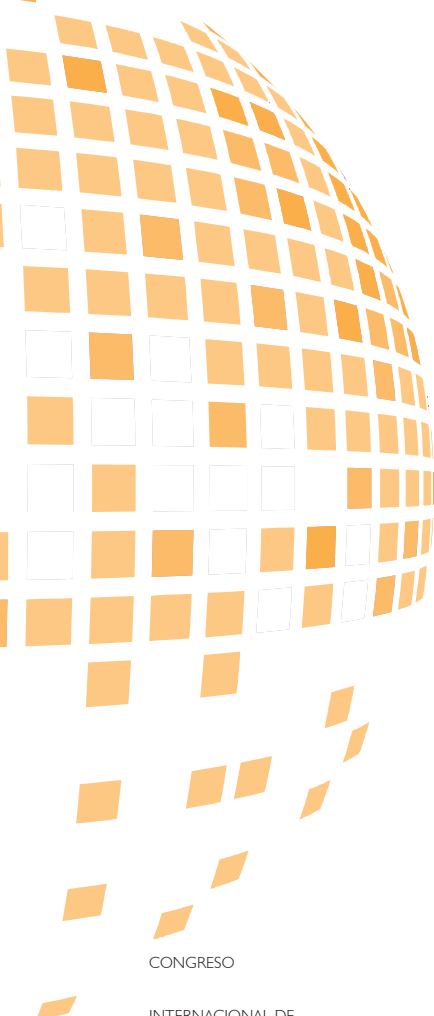
El agente de respuesta es el único agente que se ubica por cada nodo, donde se encuentra el conocimiento almacenado. Estos agentes son los encargados de llevar a cabo el proceso de consulta en el repositorio y buscar información relacionada a la petición del usuario, la cual le ha sido entregada por 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. Éste 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.



"ESTRATEGIAS DE  
LAS TECNOLOGÍAS DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

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. [7]

#### 4. Caso de estudio (implementación arquitectura)



Figura 2. Operacionalización de 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 dónde se encuentran. (Ver figura 2)

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

##### A. 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) [8]



Figura 3. Criterios para elección estándar Metadatos

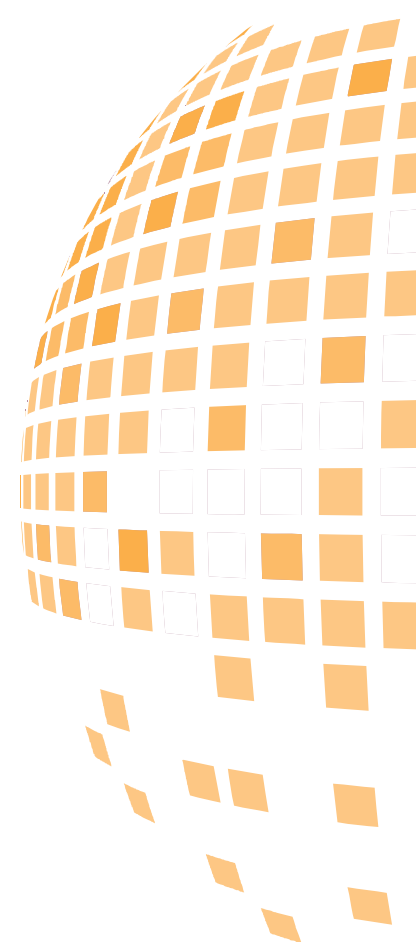
Tabla 1: Etiquetas Dublin core de las Revistas

Dublín Core	
DC. Title	Titulo 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.

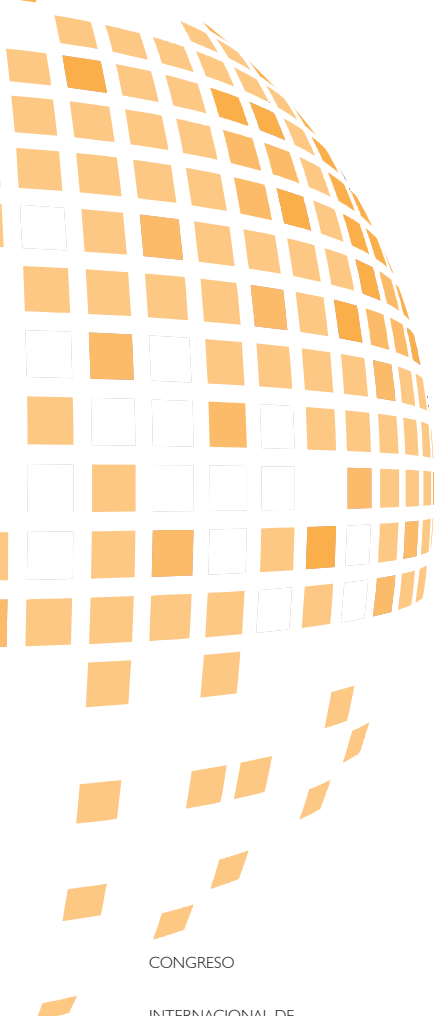
## B. Expresar Conocimiento (Ontologías)

Una vez definidos los metadatos se da inicio a la creación de la ontología. Ésta 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 más 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).

En este punto, fue preciso escoger cuál 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



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"



mejora de forma complementaria las deficiencia 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. [9]

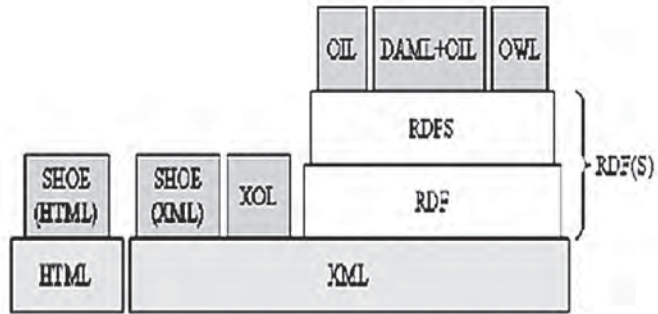


Figura 4. Arquitectura Lenguajes Ontologicos

Para la elaboración de las ontologías se usa principalmente una herramienta especializada en la construcción de éstas, llamada Protégé, en su versión 3.3.1 Protégé permite crear la ontología a través de una interfaz gráfica especializada, y como producto genera un archivo con extensión OWL. Además provee la posibilidad de integrar una de serie 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 ingenieria</dc:title>
  <dc:rights rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Los conceptos y opiniones de los articulos son responsabilidad exclusiva de su autor. En ningun momento comprometen las orientaciones
    Facultad de Ingenieria de la Universidad Autónoma del Caribe. Los articulos 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 articulos que desarrollan nuestros colaboradores abordan diferentes temas de la ingenieria tanto en el campo profesional
    áreas de la automoción, eficiencia energética y la fatiga como también en el campo de la didáctica de la formación de ingenieros potens
    formación en investigación condición fundamental para la innovación
  </dc:description>
  <dc:publisher rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Editorial Unam Autónoma</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

### C. Red De Confianza

En primer lugar, se debe definir una arquitectura adecuada para desarrollar la solución, está 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)

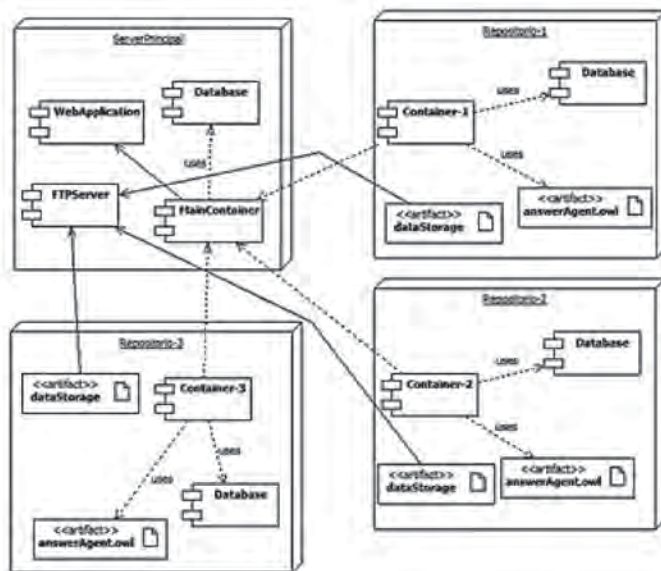


Figura 6. Arquitectura del prototipo

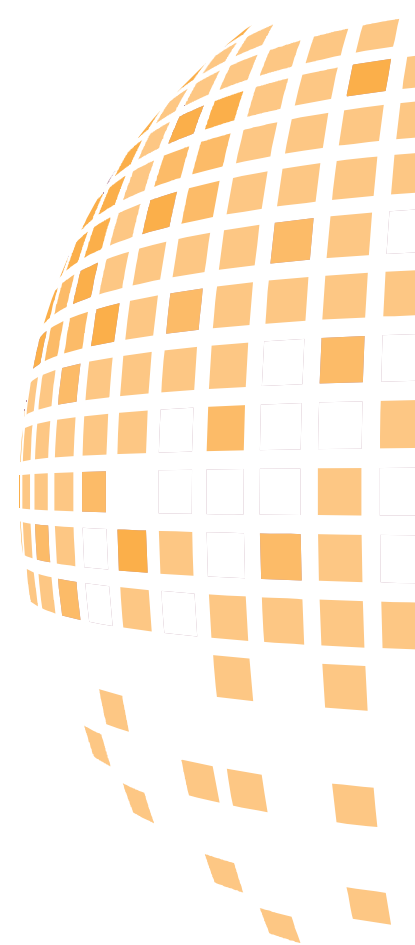
Para la implantación de la arquitectura se creó un prototipo, llamado SOMBRA. Este prototipo funciona en una red LAN donde están conectados varios equipos que contienen sus respectivos repositorios, los cuales contienen la información de los artículos. Los nodos representan los repositorios en donde 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.

#### D. Creación de Agentes y Capa de aplicación

Para la creación del entorno y ambiente de agentes, se eligió JADE. Éste 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 [10].

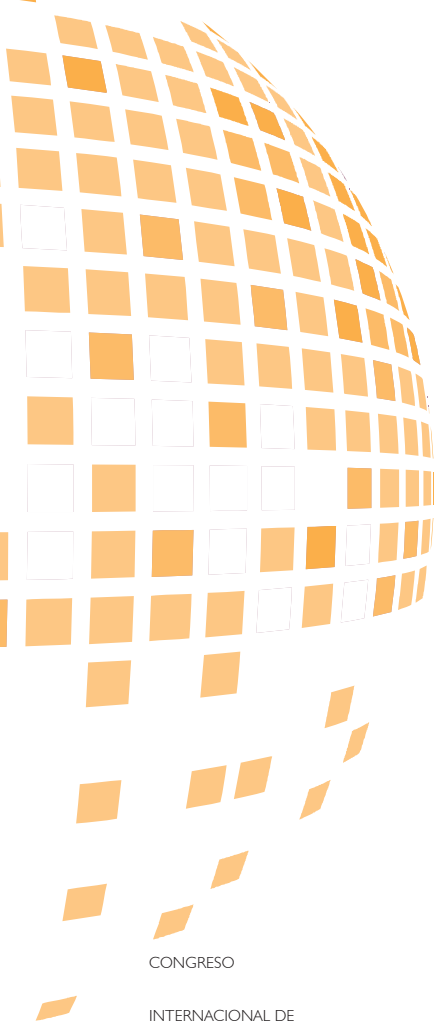
Para llevar a cabo el proceso de construcción del prototipo, el sistema se ha dividido en diferentes secciones. Cada sección contendrá diferentes tipos de agentes:

- 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.



"ESTRATEGIAS DE  
LA TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"





CONGRESO

INTERNACIONAL DE

COMPUTACIÓN Y

TELECOMUNICACIONES

COMTEL 2009

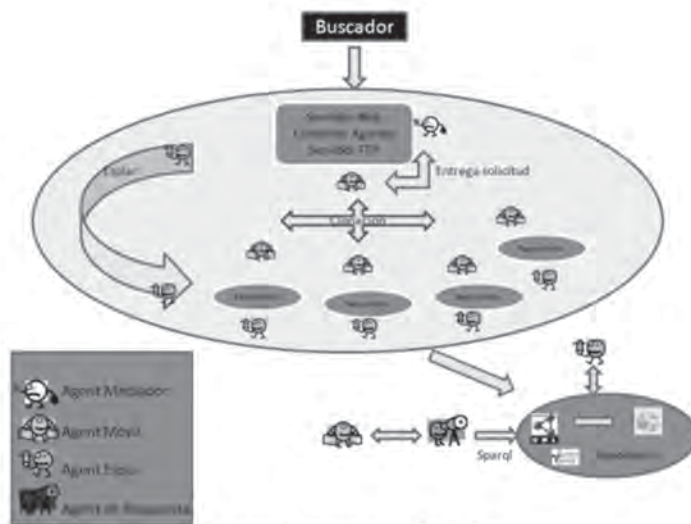


Figura 7. Descripción Prototipo

- **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.

## 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 analizó y construyó los metadatos a partir de las revistas escogidas. La segunda, a partir de estos metadatos, construyó una ontología donde creó una serie de conceptos, relaciones y reglas. En la última fase se realizó un sistema WEB de búsqueda basado en agentes sobre la ontología y utilizando consultas semánticas.

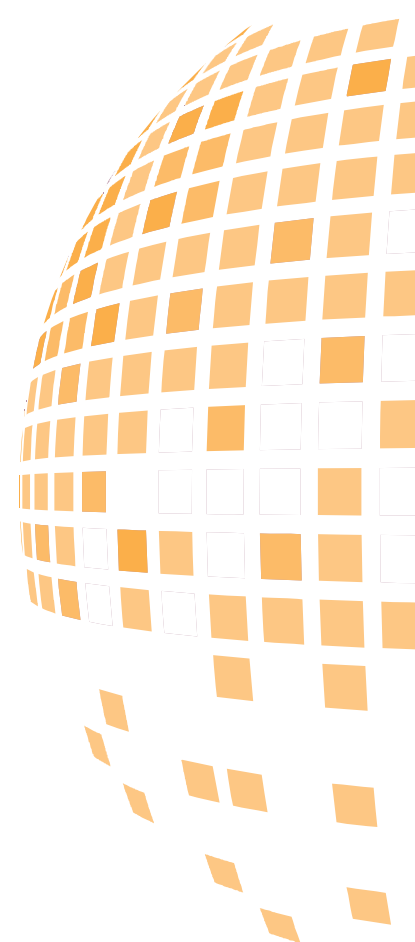
La arquitectura propuesta es ideal para la construcción de sistemas de recuperación de documentos en un entorno distribuido independiente del formato del recurso electrónico.

En esta investigación se nota cómo 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.

## 6. 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 (2008) **Modelo para la representación y visualización del conocimiento basados en ontologías web.**
- [4] Delcor J. (2006) **Descripción, Indexación, búsqueda Y Adquisición De Secuencias De Vídeo Mediante Descriptores Mpeg.** Universidad UPC.
- [5] W3C Consortium. **RDF Primer W3C Recommendation 10 February 2004.** [Internet ] <http://www.w3.org/TR/REC-rdf-syntax/>
- [6] W3C Consortium. **Recomendaciones del W3C 2004** [Internet ] <<http://www.w3.org/2007/09/OWL-Overview-es>>
- [7] Henríquez C. (2008) **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 de Maestría Universidad del Norte 2009
- [8] Castaño A y Loaiza A (2008) **Descripción de revistas y artículos científicos situados en un entorno distribuido, empleando un estándar de metadatos.** Tesis de pregrado Universidad Autónoma del caribe.
- [9] Ucross d, y Vargas I. (2008) **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 de Pregrado Universidad Autónoma del caribe
- [10]. Florez J y Messino A. (2008). **Desarrollo de un sistema basado en agentes inteligentes para la descripción, localización y acceso a artículos científicos en un sistema distribuido.** Tesis de Grado Universidad Autónoma del caribe.



"ESTRATEGIAS DE  
LAS TECNOLOGÍA DE  
LA INFORMACIÓN Y  
COMUNICACIÓN EN  
EL CONTEXTO DE LA  
CRISIS MUNDIAL"