

Buscador Ontológico Web basado en Agentes Inteligentes aplicado a Artículos Científicos: WSIA.

Paola Ariza Colpas

Corporación Universitaria de la Costa – CUC, Barranquilla, Colombia, pariza1@cuc.edu.co

Marlon Piñeres Melo

Universidad Autónoma del Caribe – CUC, Barranquilla, Colombia, mpineros@uac.edu.co

Wilson Nieto Bernal

Universidad del Norte – CUC, Barranquilla, Colombia, wnieto@uninorte.edu.co

RESUMEN

La Web Semántica propuesta por la W3C(Consortio Word Wide Web), tiene como finalidad realizar la automatización de la información contenida en la Web actual por medio del procesamiento semántico, basándose en ontologías que definen cuales han de ser las reglas utilizadas para realizar la representación del conocimiento.

El presente artículo resultado del proyecto de investigación: “Modelo para la representación de conocimiento basado en Ontologías Web y Agentes inteligentes de búsqueda, caso de Estudio: Artículos Científicos WSIA”, propone una arquitectura para la búsqueda de información a través de agente inteligentes y Ontologías Web de artículos científicos. Este artículo muestra la arquitectura, su implementación y la comparación de estas con aplicativos tradicionales

Palabras claves: Web Semántica, Ontologías Web, Agentes Inteligentes, WSIA.

ABSTRACT

The Semantic Web proposed by the W3C (Word Wide Web Consortium), aims to make the automation of the information contained in the current web through semantic processing based on ontologies that define what must be the rules used for the representation knowledge.

This article resulting from the research project "Model for the representation of knowledge based on Web ontologies and intelligent search agents, if required: Scientific articles WSIA" proposes an architecture for finding information through intelligent agents and ontologies Web of scientific articles. This paper shows the architecture, implementation and comparing these with traditional applications

Keywords: Semantic Web, Web Ontologies, Intelligent Agents, WSIA

1. INTRODUCCIÓN

En la actual era del conocimiento, la investigación se ha instaurado como un aspecto esencial en la búsqueda de respuestas a diferentes fenómenos de la cotidianidad, las evidencias de estas investigaciones se ven claramente visibles siendo plasmadas en artículos científicos.

La utilización de artículos científicos para evidenciar los resultados de investigaciones, ha logrado que la comunidad académica, pueda de esta forma acceder y conocer la producción existente en diferentes áreas del saber. Por lo tanto es necesario que al momento de realizar una búsqueda de información acerca de una temática en particular, el resultado de esta proporcione lo que el usuario realmente necesita.

En función de estas ideas es necesario precisar que la estructura de la web actual dificulta sobremanera este objetivo, debido a que se encuentra estructurada por medio de la utilización de hipervínculos, en el cual los usuarios navegan de una página a otra direccionados por los buscadores, pero finalmente las maquinas no entienden el significado de esta, en la figura No 1, se muestra el esquema navegacional de la Web actual.

Figura No 1 Funcionamiento Actual de la Web e.g (Cantera, Hierro and Romo, 2007)

Es por ello que el mismo autor de la web Tim Berners-Lee, propone la incursión de la Web Semántica, la cual adiciona significado a la Web, modificando además la forma de la estructura de los contenidos disponibles en la WWW. En contraste con el caos y desorden de la web actual, ésta propone un esquema de clasificación, modificación de la estructura y anotación de los recursos en forma semántica y entendible por las máquinas. En la figura No 2, nótese el esquema de la Web Semántica.

Figura No 2. Funcionamiento de la Web Semántica e.g (Cantera, Hierro and Romo, 2007)

La web semántica, tiene como fundamento las características generales en cuanto a accesibilidad que lograron el éxito de la web actual, pero le adiciona, emanado del campo de la inteligencia artificial, las ontologías, necesarias para la resolución de dificultades como las planteadas anteriormente.

Una Ontología Web representa una jerarquización de conceptos, por medio de atributos y relaciones, como finalidad la creación de las denominadas redes semánticas. Con la creación de las ontologías se estructura de clases y relaciones definidas de un dominio de conocimiento.

La finalidad de la incursión de la web semántica es la conformación de nodos que se encuentran tipificados por medio de clases y relaciones definidas en las ontologías, lo cual no solamente permite el acceso a contenidos, sino que le agrega funcionalidad y procedimientos para describir servicios web.

Este artículo de investigación se encuentra organizado de la siguiente forma, en primera instancia se definen algunos conceptos básicos necesarios para la comprensión de la propuesta de investigación. En segunda instancia, se describe la arquitectura propuesta para la solución a la problemática y en última instancia se discrimina el aplicativo Web resultante y las conclusiones de la propuesta.

2. ONTOLOGIAS APLICADAS A LA BUSQUEDA DE ARTICULOS

Existen muchos aplicativos o buscadores Web que tienen como objetivo principal buscar información sin discriminación alguna. No obstante es necesario precisar que al momento de realizar estas búsquedas en artículos de tipo científicos, los resultados que estos arrojan por lo general no satisfacen los que usuario espera.

De esta forma surge esta temática de investigación la cual tiene como contribuciones específicas:

- Presentar una arquitectura para mejorar los resultados de las búsquedas de artículos científicos en ontologías Web basándose en Agentes Inteligentes.
- Incentivar el uso de software libre para la representación de la ontología y desarrollo del aplicativo propuesto.
- Realizar una comparación al implementar la arquitectura con una solución desarrollada bajo el esquema tradicional, haciendo de bases de datos relacionales y el lenguaje de programación PHP.

3. CONCEPTOS BÁSICOS

Para ingresar a precisar aspectos muy puntuales de la arquitectura se hace meritorio dar a conocer algunos conceptos básicos y relevantes para la buena comprensión de las partes esenciales que conforman esta propuesta de investigación, los cuales son: Ontología, Razonador y Agente Inteligente

3.1. ONTOLOGÍAS WEB

Las ontologías tienen su origen general en la inteligencia Artificial. e.g (Gruber, 2003). Según Gruber: “Una ontología es una especificación formal y explícita de una conceptualización compartida, esto se entiende mejor de la siguiente manera:

- Formal: se refiere al hecho de que la ontología debe ser legible por un ordenador, excluyendo el lenguaje natural.
- Explícita: significa que los conceptos que se utilizan y sus limitaciones se definen explícitamente.
- Conceptualización: se refiere a la identificación de los conceptos más relevantes de un fenómeno del mundo.
- Compartida: quiere decir que una ontología captura un consenso de conocimiento, es decir que el conocimiento no proviene de un solo individuo sino que es aceptada por un grupo.”

Las ontologías pueden ser representadas por diferentes lenguajes a saber: RDF (Resource Description Framework), RDF Schema, XML(Extensible Markup Language), XML Schema, OWL(Ontology Web Language). De los lenguajes mencionados anteriormente OWL, es considerado el más completo debido a que tiene mayor capacidad al momento de expresar el significado y la semántica.

En la figura No 3, es mostrado un ejemplo de OWL que define la ontología Web utilizada para el desarrollo de esta propuesta, donde es especificada en la primera línea el archivo con extensión OWL “Tesis_Paola.owl”, sobre

el cual serán definidos cada uno de las etiquetas necesarias para la establecer el significado y la semántica de la Ontología en mención.

```
<rdf:RDF xmlns="http://www.semanticweb.org/ontologies/2010/8/12/Tesis_Paola.owl#"
xml:base="http://www.semanticweb.org/ontologies/2010/8/12/Tesis_Paola.owl"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl2xml="http://www.w3.org/2006/12/owl2-xml#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:autor_Costa="http://www.semanticweb.org/ontologies/2010/8/12/Tesis_Paola.o
wl#autor_Costa,"
xmlns:autor_Afonso="http://www.semanticweb.org/ontologies/2010/8/12/Tesis_Paola.
owl#autor_Afonso,"
xmlns:autor_Borges="http://www.semanticweb.org/ontologies/2010/8/12/Tesis_Paola.
owl#autor_Borges."> <owl:Ontology
rdf:about="http://www.semanticweb.org/ontologies/2010/8/12/Tesis_Paola.owl"/>
```

Figura No 3 Ontología para Buscador Ontológico

3.2. RAZONADOR

El objetivo principal de los razonadores, es realizar inferencias sobre la Ontología Web, para el caso específico de esta investigación se utilizó Jena Semantic Web Framework. Jena es un framework de Java, diseñado exclusivamente para la programación de aplicaciones que soporten la Web Semántica y es de código abierto. Es una API(Application Programming Interface), que permite trabajar con ontologías que se encuentren enmarcadas en los lenguajes: OWL y RDF Schema, adicionalmente permite el procesamiento de consultas de tipo SPARQL (Query Language for RDF), posee motores de inferencia y conectores a motores externos.

De tal manera que ayuda a establecer las diferentes consultas que se han de realizar a la Ontología Web, con el objetivo de interactuar con esta representación del conocimiento. En esta investigación fue utilizado Protegé, el cual es un editor de código abierto y a la vez es un sistema de adquisición de conocimiento y está hecho en Java usando fuertemente Swing, desarrollado por toda una comunidad de 10000 usuarios aproximadamente.

3.3 AGENTES INTELIGENTES

La incursión de los agentes Inteligentes dentro de propuesta, cobra un papel primordial debido a que es se convierte en su factor diferenciador en la Web 3.0. Puede entenderse como agente inteligente , e.g (Rodriguez, 2010) “una entidad capaz de percibir su entorno, procesar tales percepciones y responder o actuar en su entorno de manera racional, es decir, de manera correcta y tendiendo a maximizar un resultado esperado”.

En su artículo Is there and Inteligente Agent in Your Future? E.g(Hendler, 1999) Hendler da a conocer cuáles han de ser las principales características de un agente inteligente.

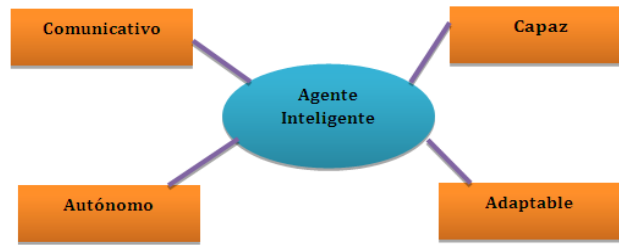


Figura No 4. Características Generales de un Agente Inteligente.

En primera instancia un agente inteligente debe ser comunicativo, puesto que debe tener la capacidad de entender cuáles son las necesidades reales del usuario, debido a que sin esta cualidad es imposible desempeñarse en forma eficiente. En segunda instancia debe ser capaz, debido a que debe desarrollar la facultad no solo de mostrar información sino también de inferir. En tercera instancia debe ser autónomo, es decir debe poseer la capacidad de tomar decisiones por si solo a través de reglas establecidas con antelación. En última instancia debe ser adaptable, acoplarse a diferentes preferencias de los usuarios en cuanto a visualización, fuentes de información, debe ser capaz de aprender del entorno.

4. ARQUITECTURA PROPUESTA PARA UN BUSCADOR SEMANTICOWEB BASADO EN AGENTES INTELIGENTES.

Resultado de este proceso investigativo se propone la siguiente arquitectura, denominada WSIA (Web Semantic Intelligent Agent), la cual se encuentra compuesta de seis capas generales a saber:



Figura No 5. Arquitectura General WSIA.

En la primera capa de la arquitectura fue definido el esquema general de la ontología Web, la cual es mostrada en la figura No 6. En esta se puede observar las clases propuestas para la solución de la problemática planteada, teniendo en cuenta las relaciones que existen entre cada una de ellas. Por ejemplo:

- Los autores son una subclase de la clase persona
- Los artículos científicos son escritos por autores y pertenecen a una línea de investigación.
- Las revistas publican artículos que son realizados por personas y estos se encuentran abscritos a una línea de investigación.

La metodología utilizada para la construcción de la ontología, fue identificar los conceptos relevantes del dominio del problema y definir las relaciones existentes entre ellos, con la finalidad de representar el área de conocimiento de los artículos científicos.

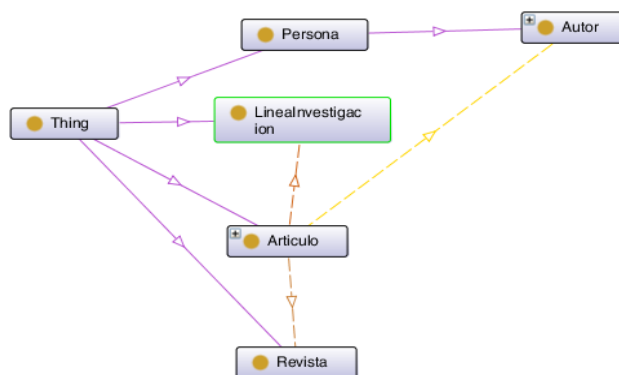


Figura No 6 Esquema General de la Ontología Web creada en Protegé 4.1

En esta investigación los artículos científicos se encuentran discriminados en tres áreas del saber: Salud, Ingeniería Básica e Ingeniería Aplicada, tomados de la revista WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY ISSUE 63 MARCH 2010. En la segunda capa de la arquitectura, fue utilizado el software Protege para poder realizar el esquema general de la ontología planteada y de esta forma poder utilizar el razonador Jena y verificar la construcción de ésta.

Fue utilizado para realizar las consultas a las ontologías SPARQL, el cual es un lenguaje de consulta RDF, factor primordial del Aplicativo Web, para mostrar el resultado de las consultas que requeridas por el usuario.

En la tercera, cuarta y quinta capa fue utilizado JADE, una API de java, cuyo principal objetivo es proveer una plataforma de agentes requerida para la creación y designación de las reglas generales a desarrollar por cada uno de los agentes inteligentes en la búsqueda de la información de los artículos científicos. En la siguiente gráfica se muestra la forma general del funcionamiento de los agentes dentro de esta arquitectura.

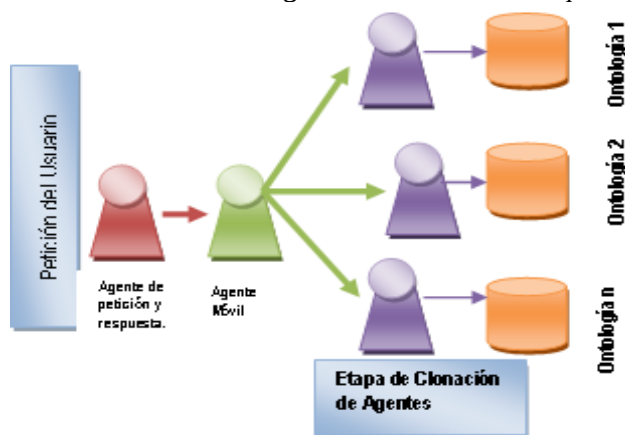


Figura No 7. Funcionamiento de los Agentes Inteligentes en WSIA.

Un agente móvil se diferencia de los agentes inteligentes comunes por su capacidad de movilizarse por distintos nodos de la red, pueden implementar perfectamente la persistencia, la comunicación y la colaboración. La característica principal de estos agentes no solo es el movilizarse en la computadora de su propietario, sino también en diferentes servidores.

La característica principal que determina la inteligencia asociada a los agentes implementados en esta arquitectura, consiste en su capacidad para comunicarse con otros y adaptarse con otros agentes en pro de encontrar los resultados coherentes con el parametro de búsqueda.

En esta investigación la función principal de los agentes inteligentes es el clonarse y reproducirse permitiéndose de esta manera la obtención de la búsqueda que realmente el usuario necesita, como es mostrado en la figura No 7.

5. CONSTRUCCION DE UN BUSCADOR SEMANTICO WEB BASADO EN AGENTES INTELIGENTES.

La arquitectura planteada anteriormente trae como consecuencia la creación de un aplicativo Web que permita la búsqueda por medio de la interacción de agentes inteligentes en diferentes ontologías.



Figura No 8. Buscador Web Semántico Basado en Agentes Inteligentes

Como es mostrado en la figura No 8, al usuario ingresar la información específica del cual desea obtener un artículo científico que concuerde con la cadena de búsqueda, internamente el aplicativo enviara esta petición a un agente de petición y respuesta que posteriormente distribuirá esta búsqueda en los agentes inteligentes que se clonarán para ingresar a la ontología propuesta para este caso de estudio investigativo. Es necesario especificar que esta arquitectura puede ser aplicada para n ontologías. El resultado es un conjunto de artículos científicos que concuerden con la búsqueda realizada por el usuario. Como puede apreciarse en la figura 9:



Figura No 9. Buscador Web Semántico Basado en Agentes Inteligentes en Ejecución.

Se desarrolló un aplicativo Web paralelo a esta solución haciendo uso de PHP y MySQL, guardando el esquema tradicional de desarrollo Web. Los resultados arrojados por este fueron contractados por los emitidos por el buscador ontológico basado en agentes inteligentes de búsqueda. Cabe anotar que ambas soluciones fueron desplegadas en ambientes locales, con una configuración de servidores con los requerimientos técnicos que son nombrados a continuación:

Tabla No 1. Configuración Básica de los servidores necesarios para realizar la prueba y requerimientos de memoria

Características necesarias para el funcionamiento adecuado de los aplicativos	Buscador Tradicional (PHP- MySql)	WSIA
Sistema Operativo	Windows	Windows
Servidor	Apache	Tomcat
Memoria	2 GB	2 GB

Resultado de estas comparaciones fueron analizadas dos variables indispensables para esta investigación, las cuales fueron: El tiempo de respuesta y la concordancia de las búsquedas realizadas. En cuanto al tiempo de respuesta la diferencia no fue muy significativa como puede apreciarse en la tabla No 2 y la Figura No 10.

Tabla No 2. Pruebas realizadas Aplicativos generales y Agentes Inteligentes.

TIEMPOS DE RESPUESTA DE APLICATIVOS CONSULTANDO	Buscador Tradicional (PHP- MySql)	WSIA
	0,0056 seg.	0,0036 seg.



Figura No 10. Rendimiento de Aplicativos Normales Versus WSIA.

No obstante, la concordancia de las búsquedas realizadas, marca un aspecto relevante en esta investigación, debido a que permite vislumbrar las diferencias reales de fondo. Por ejemplo, se ingresó la siguiente cadena de búsqueda para efectos de corroborar la eficiencia existente entre las soluciones anteriormente mencionadas: “Muéstreme los artículos escritos por Marlon Piñeres”, los resultados fueron medidos en escala de 1 a 5, teniendo en cuenta las convenciones mostradas en la tabla No 3, obteniendo los resultados mostrados en la tabla No 4.

Tabla No 3 Convenciones de concordancia de búsquedas realizadas

CONVENCION	DESCRIPCION
1	No es consistente la búsqueda
2	El resultado de la búsqueda es aceptable
3	El resultado de la búsqueda es bueno
4	El resultado de la búsqueda es muy bueno
5	El resultado de la búsqueda es excelente

Tabla No 4 Concordancia de las búsquedas realizadas

NIVEL DE COHERENCIA DE LOS RESULTADOS DE LOS APLICATIVOS.	BUSCADOR TRADICIONAL (PHP- MySQL)	WSIA
	2	4

Por lo tanto es de apreciar, que si bien es cierto que un aplicativo construido en forma tradicional, para el caso específico de este ejemplo con una base de datos MySQL y PHP puede competir en función del tiempo de respuesta a WSIA, este lo supera notablemente en cuanto a la inteligencia en los resultados mostrados en congruencia con lo que el usuario realmente requiere.

6. CONCLUSIONES

Como resultado de este proceso de investigación se puede apreciar, que haciendo uso de las Ontologías Web, Razonadores y Agentes Inteligentes de Búsqueda, pueden ser mejoradas las formas tradicionales de búsqueda en la Web. Es por ello, que se planteó una arquitectura la cual sirvió como base para la construcción de un buscador ontológico de artículos científicos WSIA (Web Semantic Intelligent Agent), haciendo uso de software libre, lo cual facilitó el desarrollo de éste.

A través de la aplicación de técnicas de pruebas de caja negra, se pudo determinar que el funcionamiento de este aplicativo en función del tiempo de respuesta y concordancia de las búsquedas realizadas, resultó bastante competidor al ser constructado con esquemas tradicionales de implementación de aplicativos Web. Sin lugar a dudas, esta afirmación se basa primordialmente en la inclusión de agentes inteligentes, que permiten incrementar en forma circunstancial la coherencia en los resultados emitidos por este.

En cuanto a los trabajos futuros, resultado de el análisis de esta propuesta de investigación, se puede apreciar la posible aplicación de la arquitectura propuesta, para lograr búsquedas eficientes en empresas del sector público tales como: alcaldías, gobernaciones, notarias públicas, entre otras.

REFERENCIAS

- Gruber, T.(2003) It Is What It Does: The Pragmatics of Ontology. Invited presentation to the meeting of the CIDOC Conceptual Reference Model committee, Smithsonian Museum, Washington, D.C., March 26, 2003. Disponible desde Internet en: http://cidoc.ics.forth.gr/docs/symposium_presentations/gruber_cidoc-ontology-2003.pdf [con acceso el 31-5-2011]
- Rodriguez, H (2010). A proposito de Convergencia hacia normas unicas: Un modelo para la aplicación de principios y reglas basadas en principios. Disponible desde Internet en: <http://www.actualicese.com/opinion/a-proposito-de-convergencia-hacia-normas-unicas-un-modelo-para-la-aplicacion-de-principios-y-reglas-basadas-en-principios-hernan-a-rodriguez-g/> [con acceso el 31-5-2011]
- Hendler, J. Is there and Intelligente Agent in Your Future? Disponible desde Internet en: <http://www.nature.com/nature/webmatters/agents/agents.html> [con acceso el 31-5-2011]
- W3C, OWL Web Ontology Language Overview. Pagina Principal. Disponible desde Internet en: <http://www.w3.org/TR/owl-features/> [con acceso el 31-5-2011].
- W3C, Resource Description Framework (RDF). Pagina Principal Disponible desde Internet en: <http://www.w3.org/RDF/> [con acceso el 31-5-2011].
- W3C, SPARQL Query Language for RDF. Página Principal Disponible desde Internet en: <http://www.w3.org/TR/rdf-sparql-query/> [con acceso el 31-5-2011].

W3C, Extensible Markup Language (XML). Página Principal Disponible desde Internet en: <http://www.w3.org/XML/> [con acceso el 31-5-2011].

JADE, Java Agent Development Framework. Pagina Principal Disponible desde Internet en: <http://jade.tilab.com/> [con acceso el 31-5-2011].

Cantera, J., Hierro, J., Romo, P., La Web Semántica, la siguiente generación de Webs. Disponible desde Internet en:

http://sociedadinformacion.fundacion.telefonica.com/DYC/SHI/seccion=1188&idioma=es_ES&id=2009100116310013&activo=4.do?elem=4299 [con acceso el 31-5-2011].

Autorización y Renuncia

Los autores autorizan a LACCEI para publicar el escrito en las memorias de la conferencia. LACCEI o los editores no son responsables ni por el contenido ni por las implicaciones de lo que esta expresado en el escrito