

USO DE ONTOLOGÍAS Y WEB SEMÁNTICA PARA APOYAR LA GESTIÓN DEL CONOCIMIENTO

USE OF ONTOLOGÍES AND SEMANTIC WEB TO SUPPORT THE KNOWLEDGE MANAGEMENT

Ana Efigenia, Sandoval Cantor

Ingeniera de Sistemas., Especialista en Telecomunicaciones Móviles., Investigador Grupo de Investigación en Gestión Empresarial e Innovación Tecnológica GEIT, Universidad Distrital Francisco José de Caldas. Profesional Especializado en la División de Informática., Universidad Militar Nueva Granada, Bogotá – Colombia
anasandovalc@gmail.com

Fecha de recepción: Marzo 7 de 2007

Fecha de aprobación: Agosto 30 de 2007

RESUMEN

Los servicios web semánticos son aplicaciones componentes de la Web 2.0, se consolidan como una evolución de los servicios web tradicionales, al estar enriquecidos con descriptores que permiten especializar las búsquedas de información, según el significado y enfocadas en las necesidades del usuario.

Las ontologías son fundamentos semánticos que tienen algunas aplicaciones prácticas desarrolladas en las ciencias de la salud y la filosofía, especialmente.

Los Servicios Web se han consolidado como una tecnología esencial para la cooperación en Internet, pero requieren mecanismos para su integración, estableciéndose como herramienta tecnológica que contribuya a la globalización y gestión del conocimiento en las actividades organizacionales o en el campo de la investigación, con servicios que mejoren los tiempos de respuesta a los usuarios, en términos de búsquedas eficientes y rápidas.

Esta Web extendida se apoya en lenguajes universales como la lógica descriptiva, los agentes inteligentes y las ontologías, resolviendo las carencias semánticas que hoy hacen difícil y dispendioso el acceso a la información en Internet.

Mientras los estándares relacionados con la composición o colaboración de Servicios Web dan un primer acercamiento, su convergencia con las tecnologías actuales, ofrecen un gran panorama hacia la obtención de un entorno en el cual la búsqueda, efectividad y ejecución de servicios sea completamente automatizada.

De acuerdo a lo anterior, la Web Semántica es una Web perfeccionada, dotada de mayor significado con el cual, cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla, gracias a una mejor definición de la información.

PALABRAS CLAVE: Web Semántica, Servicios Web, Ontología, Ontology Web Language (OWL), Resource Description Framework (RDF), Dominio, Really Simple Syndication (RSS), Metadatos.

ABSTRACT

The semantic Web Services make part of the Web 2.0 applications consolidating as a traditional web services evolution, when being enriched with describers which allow to specialize the researches according to the meaning and focused in the user's needs.

The ontologies belong to the Semantic Web Services basics and have some practical applications that had been developed in the health science and philosophy fields.

The web services have been consolidated as an essential technology to the cooperation through internet, but they require mechanisms for their own integration, establishing as a technological tool which contributes to the globalization and the knowledge management in the organizational activities or in the researches field with services that improve the user's answer time in terms of efficient and fast searches.

This extended web relies on universal languages like descriptive logic, intelligent agents and ontologies solving the semantic lack which make the internet information access difficult.

While the standards related to the web services composition or collaboration give a first time approach, their convergence to the nowadays technologies offer a huge view heading out to an automatic search, effective and execution environment.

According to this ideas the Semantic Web has been assumed as an improved Web with a higher meaning in order to any internet user could find answers to their questions in a faster and easy way, thanks to a better information definition.

KEY WORDS: Semantic Web, Web services, Ontology, Ontology Web Language (OWL), Resource Description Framework (RDF), Domine, Really Simple Syndication (RSS), Metadata.

INTRODUCCIÓN

Los factores que han generado el éxito de Internet, también han originado sus principales problemas: sobrecarga de información, heterogeneidad de fuentes y problemas consiguientes de interoperabilidad. La Web Semántica ayuda a resolver estos problemas, al permitir a los usuarios delegar tareas en herramientas de software.

Incorporando semántica en la Web, el software es capaz de procesar contenidos, razonar con ellos, combinarlos y realizar deducciones lógicas para resolver problemas automáticamente. La habilidad automática es el resultado de la aplicación de técnicas de inteligencia

artificial, las cuales exigen la participación de agentes inteligentes que perfeccionen las búsquedas, agregando valores de razonamiento y toma de decisión a los servicios Web que almacenan altos contenidos.

Con el fin de dar un carácter especializado a los contenidos temáticos, categorizando y catalogando la información a través de la generalización de áreas globales y especializadas, es indispensable hacer uso de las ontologías. El concepto de ontología está basado de tiempo atrás, en la filosofía y recientemente se utiliza en Informática para definir vocabularios que las máquinas puedan entender y que sean especificados con la suficiente precisión como para permitir diferenciar términos y referenciarlos de mejor manera.

Las ontologías incluyen definiciones de conceptos básicos relacionados con un dominio, así como las relaciones entre ellos, de tal forma que los computadores pueden codificar el conocimiento y también el conocimiento extendido, haciendo reutilizable el conocimiento.

Integrando todos los aspectos anteriores, las organizaciones y entidades, cuentan con herramientas tecnológicas eficientes de comunicación y organización de información, de tal forma que el conocimiento organizacional podrá ser gestionado, categorizado, retroalimentado y ampliamente divulgado.

1. DESARROLLO DE HERRAMIENTAS Y APLICACIONES CON ONTOLOGÍAS

Para realizar la implementación de ontologías en los servicios web, se requiere hacer uso del lenguaje OWL (Ontology Web Language), el cual proporciona terminologías interpretables por la Web para la creación de estructuras ontológicas, brindando integración e interoperabilidad de datos descriptivos para el trabajo entre diversas comunidades [9].

OWL proporciona las siguientes capacidades a las ontologías:

- Capacidad de ser distribuidas a través de varios sistemas.
- Es escalable a las necesidades de la Web.
- Es compatible con los estándares Web de accesibilidad e internacionalización.
- Es abierto y extensible.
- Da utilidad de las Ontologías para la Web.

El Grupo de Trabajo de Ontologías para la Web del Profesor Nicola Guarino [8], identificó los principales casos de uso de ontologías en la Web y los describió en el documento de "Casos de Uso y Requisitos"; se realizó un estudio sobre veinticinco servicios implementados con lenguajes de ontologías para Web poco avanzados, de donde se obtuvo la siguiente clasificación:

- Portales Web.
- Reglas de categorización utilizadas para mejorar la búsqueda.

- Colecciones multimedia.
- Búsquedas basadas en contenido para medios no textuales.
- Administración de Sitios Web Corporativos.
- Organización taxonómica automatizada de datos y documentos.
- Asignación entre Sectores Corporativos.
- Documentación de Diseño.
- Explicación de partes “derivadas” (Ej.: el tornillo de una pieza mecánica).
- Administración explícita de restricciones.
- Agentes Inteligentes.
- Expresión de las preferencias y/o intereses de los usuarios.
- Mapeo de contenidos entre sitios Web.
- Composición y descubrimiento de Servicios Web.
- Administración de derechos y control de acceso.

2. FUNCIONAMIENTO DE LA WEB SEMÁNTICA

Se parte del supuesto que esta nueva Web tiene la capacidad de construir una base de conocimiento sobre las preferencias de los usuarios y que, a través de una combinación entre su capacidad de comprensión y la información disponible en Internet, sea capaz de atender de forma exacta las demandas de información por parte de los usuarios en relación, por ejemplo, a reserva de restaurante, programación de vuelos, consultas médicas, compra de libros, etc. Si esto ocurriese de tal forma en la vida real, el usuario, obtendría resultados exactos sobre su búsqueda, sin mayores complicaciones, pero aunque la realidad es otra, la Web Semántica aporta un camino para razonar en la Web al ser una infraestructura basada en metadatos (datos altamente estructurados que describen información), extendiendo así sus capacidades.

No se trata de una inteligencia artificial mágica que permita a los servidores Web entender las palabras de los usuarios, es sólo la construcción de una habilidad dispuesta en una máquina, con el fin de resolver problemas bien definidos, a través de operaciones igualmente bien definidas que se llevarán a cabo sobre datos existentes.

Para esta labor, la Web Semántica utiliza Resource Description Framework (RDF) y Ontology Web Language (OWL), los cuales son dos estándares que ayudan a convertir la Web en una infraestructura global en la que es posible compartir, y reutilizar datos y documentos entre diferentes tipos de usuarios [2].

RDF proporciona información descriptiva simple sobre los recursos que se encuentran en la Web y que se utiliza, por ejemplo, en catálogos de libros, directorios, colecciones personales de música, fotos, eventos, etc.

OWL es un mecanismo para desarrollar temas o vocabularios específicos que se encuentran asociados a diversos recursos; lo que hace OWL es proporcionar un lenguaje para definir ontologías estructuradas que pueden ser utilizadas a través de diferentes sistemas.

Las ontologías, que se encargan de definir los términos utilizados para describir y representar un área de conocimiento, son aplicadas por los usuarios, las bases de datos y las herramientas que necesitan compartir información específica, es decir, en un campo determinado como puede ser el de las finanzas, medicina, deporte, etc. [3]

Las ontologías incluyen definiciones de conceptos básicos en un campo determinado y establece la relación entre ellos.

Dos de los ejemplos más conocidos de aplicación de Web Semántica es el servicio Really Simple Syndication (RSS), el cual es un vocabulario RDF (Resource Description Framework) basado en XML (eXtensible Markup Language) que permite realizar una catalogación de información, noticias, datos, eventos, etc., de tal manera que sea posible encontrar información precisa adaptada a las preferencias de los usuarios [4].

Los archivos RSS contienen metadatos sobre fuentes de información que han sido especificadas como de interés por parte de los usuarios, su función principal es notificar cambios acerca de los recursos seleccionados, sin necesidad de comprobar directamente ingresando a la página Web, es decir, los RSS notifican de forma automática cualquier cambio que se realice en esos recursos de interés.

FOAF (FRIEND OF A FRIEND) es un proyecto de Web Semántica, que permite crear páginas Web para describir personas, vínculos entre ellos, y cosas que hacen y crean. Se trata de un vocabulario RDF, que permite tener disponible información personal de forma sencilla y simplificada para que pueda ser procesada, compartida y reutilizada.

Dentro de FOAF se puede destacar FOAF-a-Matic, que se trata de una aplicación Javascript que permite crear una descripción FOAF de uno mismo; con esta descripción, los datos personales serán compartidos en la Web pasando a formar parte de un motor de búsqueda donde será posible descubrir información a cerca de una persona en concreto y de las comunidades de las que es miembro de forma sencilla y rápida.

Los buscadores semánticos son un ejemplo más de aplicaciones específicas, cuyo objetivo es satisfacer las expectativas de búsqueda de usuarios que requieren respuestas precisas.

3. ESTADO DEL ARTE

Tal como lo menciona Aldo Gangemi [5]: “la Web semántica cuenta con una poderosa visión que traerá cambios significativos a la forma en la que se desarrollan los servicios web actuales”, ya que con el valor agregado de la semántica y el aporte de las ontologías, estos servicios podrán ser orientados a campos específicos como la genética, la biología, la medicina, entre otras áreas de investigación, ya que contribuyen a la construcción de fuentes de información y documentación relevante, organizada, clasificada y actualizada para las comunidades virtuales de trabajo e investigación.

En términos prácticos, el significado de web semántica, está explicado claramente por tres investigadores (entre ellos Tim Berners-Lee [1] considerado inventor y protector de la web), quienes hacen referencia a un Internet novedoso en el que los computadores muestran la información en páginas web con la capacidad de poder “entender” dicha información. Esto significa que los equipos de cómputo, ya sean personales o cualquier otro dispositivo conectado a Internet, podrán realizar múltiples tareas sin necesidad de intervención humana, minimizando esfuerzos al usuario en términos de tiempo y haciendo posible que las herramientas tecnológicas y las personas trabajen en cooperación.

El consorcio W3C (World Wide Web Consortium) es dirigido por Tim Berners-Lee, en donde su grupo de trabajo afirma que: “la web semántica es una visión: la idea de tener datos en la web definidos y enlazados de forma que puedan ser utilizados por máquinas no sólo con el fin de presentarlos sino también para la automatización, integración y reutilización de la información a través de varias aplicaciones”.

En concordancia con lo anterior, se documentan casos aplicados de web semántica, algunos relacionados con aspectos del área médica (terminología, diagnóstico, pronóstico, sistemas terapéuticos) y uno sobre biodiversidad.

3.1 ANÁLISIS ONTOLÓGICO PARA LA UNIFICACIÓN DE LA BIOLOGÍA

En el ámbito de la biología y en particular, en cuanto al estudio de la biodiversidad, la taxonomía y la sistemática son el fundamento de la clasificación de la vida en la tierra. El proceso de coleccionar, catalogar y describir las especies se ha llevado a cabo durante varias décadas, y han resultado en colecciones biológicas de gran valor. Sin embargo, la capacidad para llevar a cabo este inventario de la biodiversidad, así como la conservación de las colecciones biológicas y el acceso a esta información es limitada [7, 8, 9].

Este impedimento taxonómico ha limitado el diseño de programas, planes y políticas orientados al conocimiento, conservación y uso sostenible de los recursos biológicos.

La Iniciativa Mundial en Taxonomía (IMT) [13] adoptada en Mayo de 2000, fue desarrollada como una estrategia para superar el impedimento taxonómico, y apoyar de tal manera, la implementación del Convenio sobre Diversidad Biológica de las Naciones Unidas.

El plan estratégico de la Iniciativa Mundial en Taxonomía (IMT) incluye los siguientes elementos:

- a) Evaluar las necesidades en taxonomía.
- b) Conservar las colecciones y ejemplares biológicos.
- c) Facilitar el acceso a la información de colecciones.
- d) Incluir elementos de taxonomía en los programas de trabajo del convenio sobre biodiversidad.

La Web Semántica tiene como objetivo ser una base de conocimiento, al estar dotada de mayor significado da la posibilidad a cualquier usuario en Internet de encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una mejor definición de la información [19]. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web actualmente carente de semántica y en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante [3, 15].

De acuerdo a estudios de expertos y tal como menciona Pisanelli [14], se da testimonio de la urgencia de unificar información en campos del conocimiento como la Biología, puesto que expertos como bioquímicos y genetistas ahora reconocen un solo universo de conocimiento, en su caso particular y en lo relacionado con el estudio de genes, proteínas y manejo de secuencias de genomas enteros, se requiere contar con mecanismos automatizados que hagan eficiente el manejo de la información, esto ha de ser viable con las amplias posibilidades de clasificación, descripción, organización y funcionalidad que ofrecen las ontologías, los lenguajes descriptivos y las reglas semánticas en las que se basan las ontologías y la Web Semántica, sin dejar de lado el valioso aporte que brindan los agentes inteligentes en los procesos de búsqueda especializados y la toma de decisiones.

Aunque una unificación no es tarea fácil, aun cuando se habla entre expertos y dadas las ambigüedades o dudas que puedan darse de un idioma a otro, surge como gran alternativa, la adopción y ejecución de las aproximaciones ontológicas para sistemas de representación de terminologías.

Las ontologías no solo redefinen el conocimiento de manera fácil, sino que son también el resultado de esfuerzos conjuntos para alcanzar la estandarización, debido a que se hacen conceptualizaciones explícitas detrás de terminologías y modelos.

3.2 LAS ONTOLOGÍAS COMO MODELO DE CONOCIMIENTO

La ontología trata de describir o proponer las categorías y relaciones básicas del ser o la existencia para definir las entidades y de qué tipo son [15]. Las entidades comprenden los objetos, las personas, los conceptos, las ideas, las cosas, y todo aquello de lo que se puede cuestionar su existencia, en cierto modo reflexiona sobre las concepciones de la realidad y sobre cómo son definidas las entidades de la realidad.

El concepto de ontología se ha basado desde hace mucho tiempo en la filosofía y recientemente se utiliza en informática para definir vocabularios que las máquinas puedan entender y que sean especificados con la suficiente precisión como para permitir diferenciar términos y referenciarlos de mejor manera.

Una definición más precisa, extraída del documento de análisis ontológicos [16] expresa: "Una ontología define los términos a utilizar para describir y representar un área de conocimiento". Las ontologías son utilizadas por las personas, las bases de datos, y las aplica-

ciones que necesitan compartir un dominio de información (un dominio es simplemente un área de temática específica o un área de conocimiento, tales como las ciencias básicas, la fabricación de herramientas, la comercialización de bienes inmuebles, el campo automotriz, la gestión financiera, etc.).

Las ontologías útiles para los computadores son aquellas que incluyen definiciones de conceptos básicos del dominio y sus relaciones, codificando el conocimiento de dicho dominio y su relación extendida con otros dominios, de tal manera que pueda hacerse reutilizable el conocimiento.

3.3 ONTOLOGÍAS Y TERMINOLOGÍAS MÉDICAS

El proyecto GALEN General Architecture for Languages Encyclopedias and Nomenclatures in Medicine Proyecto financiado por la Unión Europea con el objetivo de desarrollar herramientas y métodos para la elaboración y mantenimiento de clasificaciones de procedimientos quirúrgicos que los distintos grupos de trabajo se encargan de representar conceptualmente mediante un lenguaje específico denominado GRAIL (Galen Representation and Integration Language).

Actualmente el proyecto GALEN desarrolla un modelo semántico para la gestión de terminología clínica denominado CORE (Coding Reference) el cual cuenta con tres módulos en donde se incluyen relaciones que reflejan posibles combinaciones de términos y conceptos complejos como por ejemplo: 'fracture of the left humerus'. La manipulación de estos conceptos y relaciones se hace a través del lenguaje GRAIL y el CM (Concept Model), herramienta conceptual que permite generar modelos con conceptos y relaciones. Esta capacidad relacional implica que no es necesario enumerar de forma explícita todos los términos posibles, ya que, como en el caso del ejemplo, el sistema puede generar y clasificar aquellos conceptos que representen fracturas óseas. GALEN también tiene en cuenta los sistemas de clasificación vigentes en ciencias de la salud, relacionando los conceptos de estos sistemas con los conceptos estructurados en el modelo CORE que actúan a modo de traductor y que da lugar al tercer elemento: el Módulo de Conversión (Code Conversion Module).

La propuesta de GALEN supuso un avance al basarse en sistemas funcionales y dinámicos con relación a otras propuestas que contaban únicamente con almacenamiento estático de datos. Sin embargo, en la actualidad, esta herramienta se convierte progresivamente en un recurso difícil de controlar, ya que debido a su alto potencial de combinación se puede generar redundancia si se expresa un mismo concepto con más de una combinación posible, adicionalmente, existen dificultades de procesamiento gracias al gran volumen de datos y el número de algoritmos que se requieren para llevar a cabo las distintas operaciones [17].

3.4 SISTEMA DE LENGUAJE MÉDICO UNIFICADO

El lenguaje médico unificado es un sistema que permite contar con referencias cruzadas entre diversos vocabularios y clasificaciones correspondientes a estándares de terminologías médicas, que se obtienen gracias al análisis léxico de dichos términos. El UMLS (Unified Medical Language System) es un proyecto de la National Library of Medicine del National Institutes of Health de Estados Unidos y emplea técnicas de procesamiento basadas en unidades léxicas, en donde las frases se comparan en función de lo que parecen significar, en el orden que aparecen las palabras, en el tipo de palabras y en constructores gramaticales utilizados, aún teniendo en cuenta la función de lo que podrían o no podrían significar.

Este sistema no ha sido diseñado para consulta humana, los usuarios finales son expertos en programación que al trabajar con expertos en el dominio médico podrán definir consultas y relaciones basadas en los elementos del UMLS. El lexicón (diccionario) especializado del UMLS es un léxico en lengua inglesa que incluye una gran cantidad de términos del dominio biomédico, la versión actual incluye más de cien mil informes léxicos, con más de ciento ochenta mil cadenas de términos [18].

En el sistema, las entradas incluyen información sintáctica, morfológica y ortográfica, en donde se puede contar con un término formado por un sólo elemento léxico o por varios. La forma base no tiene flexión, singular en el caso de los sustantivos, infinitivo en el caso de los verbos y en forma positiva en el caso de los adjetivos y adverbios. La información léxica, incluye la categoría sintáctica, la variación de la inflexión (singular o plural para los sustantivos, conjugación de los verbos, el comparativo, superlativo para los adjetivos y adverbios), y posibles patrones de complementación (objetos y otros argumentos que los verbos nombres y adjetivos pueden regir). El lexicón reconoce once categorías sintácticas o partes del discurso: verbos, nombres, adjetivos, adverbios, auxiliares, modales, pronombres, preposiciones, conjunciones y determinantes.

Los patrones básicos de la oración se determinan por el número y la naturaleza de los complementos que rigen los verbos, donde las entradas verbales contemplan las formas verbales (regulares o irregulares). En lo relacionado con los sustantivos, se tienen en cuenta los patrones de pluralización y de nominalización.

En términos generales, se puede observar que el sistema implanta una gramática electrónica incluida en la información de entrada, aunque algunos autores [19] defienden la necesidad de incluir toda la información gramatical en un módulo aparte; esta es una gramática electrónica rigurosamente dependiente de la lengua, la cual es aplicada en un módulo que incluye información estrictamente conceptual y denotativa.

Otro de los elementos relevantes en UMLS es el Metathesaurus, el cual se define como una base de datos que contiene información sobre conceptos que aparecen en uno o más conjuntos de sesenta vocabularios controlados y clasificaciones usados en el campo de la biomedicina. Los conceptos se encuentran organizados y poseen un identificador único,

asociado a distintos términos y variantes existentes para referirse al mismo, adicionalmente, ofrece relaciones entre los conceptos de los vocabularios de origen y nuevos vocabularios.

La Red Semántica de UMLS es uno de los tres proyectos sobre fuentes de conocimiento que actualmente desarrolla la National Library of Medicine, usa ciento treinta tipos semánticos y garantiza una categorización consistente de todos los conceptos representados en el metatesauro. Los enlaces entre los tipos semánticos establecen la estructura de la Red y representan las relaciones más importantes en el dominio biomédico. Toda la información sobre los conceptos específicos se encuentra en el Metatesauro; la Red facilita la información sobre los tipos semánticos básicos que se han asignado a estos conceptos y define las relaciones que se pueden establecer entre los tipos semánticos.

En cuanto a la estructura y contenido, existen nodos (tipos semánticos) y enlaces (relaciones entre ellos), existen agrupaciones básicas de los tipos semánticos sobre organismos, estructuras anatómicas, funciones biológicas, productos químicos, eventos, objetos físicos y conceptos o ideas. El actual campo de aplicación de los tipos semánticos es bastante extenso, ya que permite la categorización semántica de un amplio grupo de terminologías en múltiples dominios de especialidad.

3.5 NOMENCLATURA SISTEMATIZADA DE MEDICINA

La saturación de la web actual ha derivado problemas de comprensión en distintos campos del conocimiento, uno de los casos recientes, ha sido la falta de consenso general sobre la nomenclatura y clasificación de las neuronas corticales, con el fin de estudiar enfermedades como la esquizofrenia, la depresión, el trastorno bipolar y el Alzheimer, pese a que existen innumerables investigaciones relacionadas con el cerebro, no se ha dado unanimidad a la hora de clasificar tales neuronas [20].

En este sentido, el Colegio Americano de Patologías ha desarrollado una ontología médica denominada SNOMED (Systematized Nomenclature of Medicine), la cual agrupa una serie de nomenclaturas médicas.

El sistema integrado de terminologías clínicas médicas ha sido validado por el Programa Nacional de Salud del Reino Unido y la Biblioteca Nacional de Medicina de Estados Unidos. La terminología integra un lenguaje común que ofrece un método consistente para capturar, compartir, agregar y actualizar información sobre especialidades y cuidados médicos. El sistema ofrece un registro médico electrónico que permite manejar información sobre monitoreo en unidades de cuidados intensivos, toma de decisiones clínicas, estudios de investigaciones, ensayos clínicos, ordenes de entrada, cuidado permanente de enfermedades e indexación de imágenes.

4. PROTOTIPO DE WEB SEMÁNTICA PARA LA GESTIÓN DEL CONOCIMIENTO EN EL SECTOR PRODUCTIVO Y DE INVESTIGACIÓN CIENTÍFICA DE LA MALACOFAUNA TERRESTRE COLOMBIANA

El prototipo de web semántica que se describe, fue desarrollado con el fin de aplicar gestión del conocimiento en las actividades productivas y de investigación científica sobre cinco variedades de caracol declaradas aptas para la producción y comercialización en Colombia. El proyecto contó con la participación de la Bióloga Clara Inés Medina Bermúdez, líder del Grupo de Investigación en Malacofauna Terrestre de Colombia.

4.1 OBJETIVO

- Establecer las deficiencias de los sistemas computacionales utilizados por profesionales vinculados a procesos del sector productivo y de investigación científica en biodiversidad.
- Aprovechar los esquemas de búsqueda de los sistemas Web Semánticos basados en ontologías, con el fin de promover la creación de comunidades de conocimiento, mediante la especificación y consolidación de información especializada, vigente y eficaz.

4.2 CONFIGURACIÓN DEL SISTEMA

- Implementación del servicio web, generación de secciones, esquemas de acceso y navegación, configuración de contenidos.
- Creación de los esquemas RSS y etiquetas para facilitar el intercambio de información sobre los siguientes datos taxonómicos relevantes de las especies: familia, especie, origen, localización en Colombia, altitud, referencias para identificación y acción en el medio.
- Sindicación de los contenidos mediante el uso de tags propietarios, que permiten la identificación de los descriptores que componen la base de conocimiento en que se apoya la lógica descriptiva y el razonador del sistema web semántico.
- Generación de modelos para la creación de agentes inteligentes, funcionales en los esquemas de búsquedas avanzadas, basadas en ontologías.

Para crear el modelo ontológico en la investigación, se usó Protégé; herramienta de software que permite desarrollar ontologías y sistemas basados en el conocimiento, con el fin de resolver problemas y tomar decisiones en dominios particulares. Este programa fue desarrollado en JAVA por la Universidad de Stanford y emplea una interfaz de usuario que facilita la creación de una estructura de marcos con clases, propiedades, reglas e instancias de manera integrada.

La figura 1 muestra el esquema de ontologías diseñado para el prototipo, en el cual se categoriza la información, generando clases y niveles de conocimiento relacionado con el tema objeto de estudio.

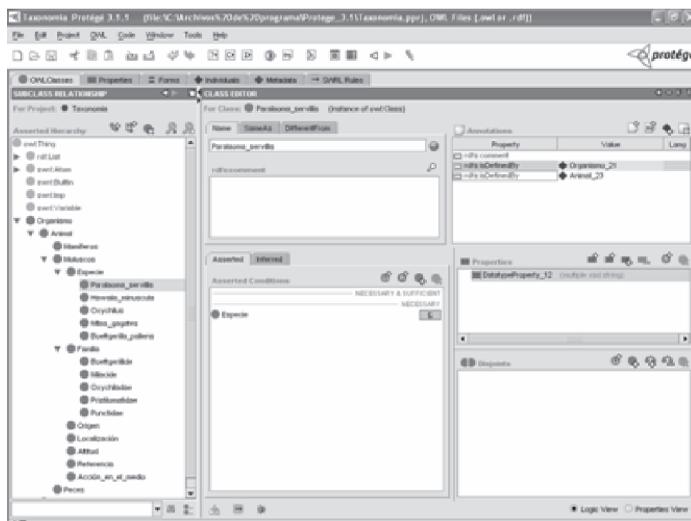


Figura 1. Definición ontológica diseñada en Protégé

La identidad ontológica elaborada en Protégé, fue generada a partir de la clasificación convencional de las especies y seres vivientes, es un esquema de correspondencias lógicas que permite identificar cinco tipos de especies de la malacofauna terrestre colombiana. El modelo describe los siguientes conceptos:

Organismo

Plantas

Animales

Mamíferos

Peces

Moluscos

Especie

Paralaoma servilis

Hawaiiia minuscula

Vitrea

Oxychilus alliarius

Milas gagates

Familia

Punctidae

Pristilomatidae

Oxychilidae

Milacidae

Boettgerillidae

Origen

Localización

Altitud

Referencias para identificación

Acción en el medio.

4.3 LÓGICA DESCRIPTIVA

La lógica descriptiva se basa en la representación del conocimiento, en donde el dominio se define como un área temática específica o un área de conocimiento. También hace uso de una terminología o vocabulario del dominio (TBOX) y un conjunto de afirmaciones (ABOX).

El vocabulario se compone de conceptos y roles. Los conceptos corresponden a conjuntos de elementos y los roles a relaciones binarias entre elementos, en donde se pueden definir dos tipos de conceptos y roles: atómicos y complejos.

Para el caso particular se toma las siguientes definiciones:

- Conceptos atómicos:
 - Animal
 - Molusco
 - Babosa
 - Caracol
 - Concha
 - Hembra
 - Macho
 - Gasterópodo
 - NombreComún
 - NombreCientífico
- Roles atómicos:
 - tieneConcha
 - segregaBaba
 - esHermafrodita
- Relaciones:
 - Babosa \equiv Molusco \cap \neg Caracol
 - Molusco \equiv Animal \cap Concha
 - Gasterópodo \equiv Molusco \cap \neg Concha
 - Macho \equiv esHermafrodita.Babosa
 - Caracol \equiv \neg Babosa

La información recogida en la ABOX es:

- tieneConcha(Caracol)
- esHermafrodita(Babosa)
- esMolusco(Babosa,Caracol)
- esGasterópodo(Babosa)
- tieneNombreComun(Babosa,Caracol)
- tieneNombreCientífico(Babosa,Caracol)

Existen razonadores que permiten trabajar con los TBOX y ABOX para verificar si el contenido de la TBOX es probable, o para contrastar las relaciones que están incluidas en otras. En el prototipo, se fijaron reglas fijas de inferencia para ser aplicados por el agente inteligente de búsqueda.

4.4 INTERFAZ DE USUARIO

La figura 2 muestra la página inicial del prototipo (portal web especializado en información y catalogación de especies de la malacofauna terrestre colombiana). La interfaz de usuario es una página web que puede ser consultada a través de cualquier navegador de Internet.

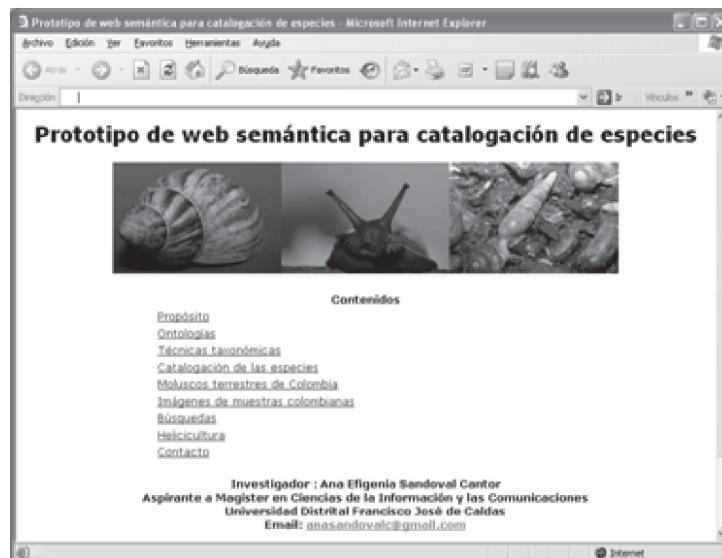


Figura 2. Interfaz de usuario, página principal del servicio web semántico

Cada una de las secciones cuenta con sindicación RSS (figura 3). El prototipo está dividido en varias secciones internas que corresponden al modelo ontológico, mediante el cual se brinda acceso a información clasificada, catalogada y especializada.

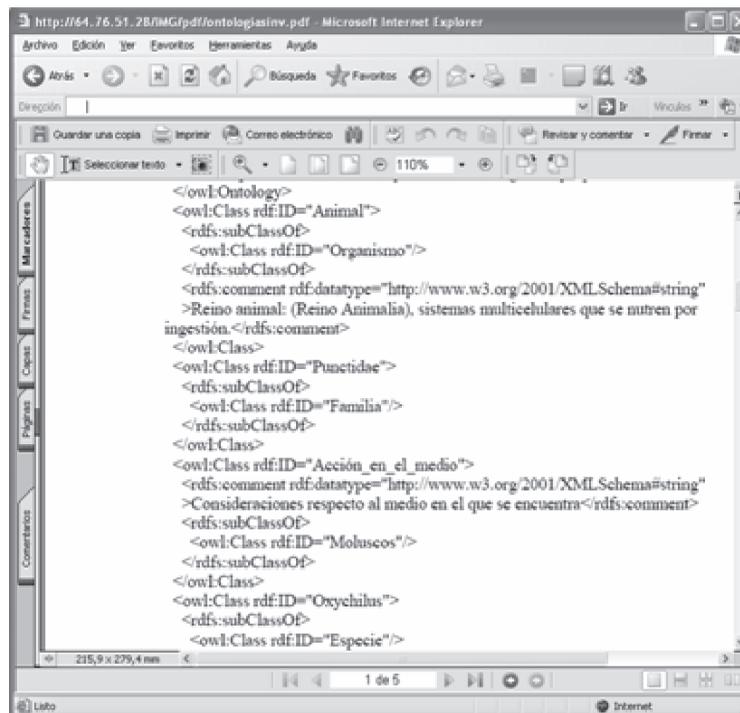


Figura 3. Definición del modelo ontológico diseñado en Protégé

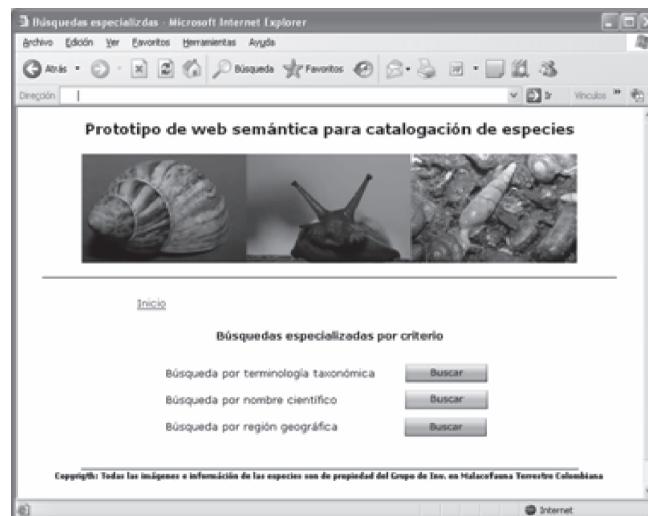


Figura 4. Interfaz de usuario, página para búsquedas especializadas

El prototipo fue evaluado de forma individual por diez personas de diferentes niveles educativos y formación profesional, todos ellos vinculados con el área de la helicultura. El cuadro de evaluación aplicado se muestra en la tabla 1, donde se piden valoraciones cuantitativas y/o cualitativas.

Tabla 1. Cuadro de evaluación funcional.

Prototipo de Web Semántica, como apoyo a la investigación científica y al sector productivo empresarial en Malacofauna Terrestre de Colombia	
Evaluador:	
Por favor evalúe de 1 a 5 cada aspecto, siendo 1 la calificación menos favorable y 5 la más favorable. Si lo prefiere, evalúe únicamente el aspecto general que aparece con el ítem "Observaciones"	
Ítem	Valor
Usabilidad	
Comprensibilidad Global del Sitio	
Mecanismos de ayuda y Retroalimentación en línea	
Aspectos de Interfaces y Estéticos	
Usabilidad de los Textos	
Elementos predictivos de navegación	
Flexibilidad	
Consistencia	
Funcionalidad	
Mecanismo de Búsqueda en el Sitio	
Nivel de Interconexión	
Indicador del Camino	
Etiqueta de la Posición Actual	
Confiabilidad	
Enlaces Rotos	
Enlaces Inválidos	
Enlaces externos a instituciones prestigiosas	
Utilización de estándares del W3C (HTML y/o RSS)	
Protege contra copia o modificación los contenidos susceptibles de la institución	
Resguarda la propiedad de autores foráneos con citas documentadas	
Documenta la fuente de las imágenes y documentos descargables de otras instituciones	
Eficiencia	
Accesibilidad de Información	
Soporte a Versión sólo Texto	
Legibilidad al desactivar la Propiedad Imagen del Browser	
Rendimiento	
Proporciona información pertinente	
Organiza información específica	
Suministra información debidamente clasificada	
Tiempo de descarga	
OBSERVACIONES	

Fuente: el autor.

5. CONCLUSIONES

En términos de búsqueda, la web semántica no genera un impacto visual que lo diferencie de otras técnicas preexistentes, como son los buscadores convencionales. Sin embargo, su real potencial se encuentra en la dinámica de su funcionamiento, su concepto, sus fundamentos, convirtiéndose en un mecanismo eficiente de clasificación, organización y dinamizador de la información.

Existe la necesidad de dar a conocer la potencialidad de la web semántica entre expertos de diversos campos del saber, ya que su funcionamiento se basa en la automatización y acceso a colecciones estructuradas de información, razonamientos y procesos cognitivos.

Esta web dotada de significado, al apoyarse en los sistemas agentes y multiagentes de la Inteligencia Artificial, requiere la creación de modelos de construcción de conocimiento, donde se definan conceptos, se asocien significados, se construyan valores de verdad, reglas y lógicas de inferencia, que den poder a la toma de decisión sin intervención humana.

El uso de las nuevas tecnologías ha de apoyar los procesos de representación del conocimiento, aunque hoy en día se cuenta con innumerables repositorios y fuentes de información, éstos en su mayoría carecen de orden, validez, homogeneidad y actualización permanente por parte de quienes lideran el dominio del conocimiento. Las ontologías son la herramienta con la que se podrán modelar y organizar los conceptos, significados y sus asociaciones.

Por su parte, los agentes inteligentes permitirán automatizar los procesos de enriquecimiento del conocimiento, resolviendo búsquedas especializadas, generando asociaciones, nuevas construcciones de conceptos complejos y toma de decisiones.

Las instituciones educativas y gubernamentales han de agremiarse con el fin de promover el desarrollo de productos relacionados con las políticas de Tecnología e Innovación del estado Colombiano, considerando planes de gestión en grupos multidisciplinarios que hagan uso de las tecnologías del conocimiento y las telecomunicaciones, construyendo servicios web eficientes que brinden fuentes y medios de información útiles, válidos, autogestionables y dinámicos.

El prototipo de web semántica fue evaluado funcionalmente por diversos tipos de usuarios, mediante un sistema de valoración basado en cinco criterios: usabilidad, accesibilidad, funcionalidad, confiabilidad y eficiencia. En el ejercicio de valoración, se tomó como altamente favorable, la confiabilidad y accesibilidad de información específica a un área temática. Como deficiencia se estimó la carencia de iniciativas tecnológicas promovidas por entidades gubernamentales que ayuden a minimizar la actual brecha digital en las pequeñas industrias de helicicultura.

REFERENCIAS BIBLIOGRÁFICAS

- [1] BERNERS Tim L. (2001) The Semantic Web. - Publications Scientific American.
- [2] CABRAL Liliana, DOMINGUE John, MOTTA Enrico, (2004). Approaches to Semantic Web Services: An Overview and Comparisons. IAM, University of Southampton, Southampton, UK.
- [3] COLOMB Robert (2002). The physical being of institutional facts - National Research Council - Institute of Biomedical Engineering - ISIB-CNR - Corso Stati Uniti, 35127 Padova.
- [4] COLOMB Robert (2002). Use of Upper Ontologies for Interoperation of Information Systems - National Research Council - Institute of Biomedical Engineering - ISIB-CNR - Corso Stati Uniti, 35127 Padova.
- [5] GANGEMI Aldo (1998). Ontology Integration: Experiences with Medical Terminologies - National Research Council, Institute for Cognitive Science and Technologies (ISTC-CNR).
- [6] GANGEMI Aldo (1999) Applying Ontologies to the Integration of Medical Terminologies - National Research Council, Institute for Cognitive Science and Technologies (ISTC-CNR).
- [7] GANGEMI Aldo (2000) Understanding Systematic Conceptual Structures in Polysemous Medical Terms - National Research Council, Institute for Cognitive Science and Technologies (ISTC-CNR).
- [8] GUARINO Nicola, (1994) Formalizing Ontological Commitments - LADSEB-CNR, National Research Council, Padova, Italy.
- [9] GUARINO Nicola, (1995). Formal Ontology, Conceptual Analysis and Knowledge Representation - LADSEB-CNR, National Research Council, Padova, Italy.
- [10] GUARINO Nicola, (1998). Formal Ontology and Information Systems - LADSEB-CNR, National Research Council, Padova, Italy.
- [11] GUARINO Nicola, (2000) A Formal Ontology of Properties - LADSEB-CNR, National Research Council, Padova, Italy.
- [12] GUARINO Nicola (2002) Evaluating ontological decisions with Ontoclean - LADSEB-CNR, Padova, Italy and Program of Computer science at Vassar College in Poughkeepsie, NY.
- [13] Meeting: Global Taxonomy Initiative Coordination Mechanism - 23 November 2000 Montreal, Canada.

-
- [14] PISANELLI Domenico M (1998). An Ontological Analysis of the UMLS Metathesaurus - INMM National Research Council, Roma, Italy.
- [15] PISANELLI Domenico M (2000) - The Role of Ontologies for an Effective and Unambiguous Dissemination of Clinical Guidelines - INMM National Research Council, Roma, Italy.
- [16] PISANELLI Domenico M., GANGEMI Aldo, BATTAGLIA Massimo (2004) - Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies - ISTC INMM National Research Council, Roma, Italy pisanelli@acm.org.
- [17] PISANELLI Domenico M. (2002) Ontological Analysis for the Unification of Biology - INMM National Research Council, Roma, Italy.
- [18] PULLAN Martin R. (2000). The Prometheus Taxonomic Model: a practical approach to representing multiple classifications - Royal Botanic Garden Edinburgh EH3 5LR, U.K.
- [19] REINOUT Van R. (2003) Clarity in the usage of the terms Ontology, Taxonomy and Classification. - Civil engineering informatics, Delft University of Technology. Cambridge: Addison-Wesley.
- [20] WELTY Christopher, (2001) Supporting ontological analysis of a taxonomic relationships - Computer Science Department, Vassar College USA.

