UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA

Prototipo de un buscador web semántico para el LCE (Laboratorio de Cómputo Especializado)

Autor:
Nestor Luis Niño Salcedo
nnino@unab.edu.co

Director:
Dr. José de Jesús Pérez Alcázar
jperez@unab.edu.co

Facultad de Ingeniería de Sistemas
Línea de Tecnologías de la Información

PALABRAS CLAVES: Web Semántica, Ontología, SWQ (Smart Web Quey).

RESUMEN

La búsqueda de información en la web se ha convertido en un tema de interés general dentro del mundo de la computación, generando la introducción de nuevos paradigmas. La web semántica como nuevo paradigma de las tecnologías de la Información está apenas evolucionando y desde el finales del año 2002 se ha tomado en serio por los grupos de investigación de Computación e Informática y algunas empresas en áreas como La recuperación, extracción y mantenimiento de la información que están dentro del àr administración del conocimiento y el área de comercio electrónico. El presente artículo muestra la aplicabilidad que tiene la web semántica, el modelamiento de la web semántica, la construcción de las ontologías parte esencial de la web semántica, el método SWQ (Smart Web Query) y el desarrollo de la aplicación.

INTRODUCCIÓN

El crecido interés de las personas y el auge que ha tenido la Web en los últimos diez años ha llevado a que su utilización crezca exponencialmente.
La causa de éste crecimiento exponencial es la utilidad que presta la web a través de los sitios web que se encuentran en ella.
Los sitios guardan gran cantidad de información relacionada con la organización que la creó, con el fin de que su información sea conocida en diferentes partes del mundo.
Muchos de estos sitios son creados teniendo como prioridad su diseño, en vez del manejo de información que se tiene en él, lo que ha imposibilitado algunas veces a acceder a la información requerida por el usuario de estos sitios.

La World Wide Web (WWW) ha cambiado la posibilidad de acceder a la información. Actualmente, hay alrededor de tres billones de documentos estáticos en la WWW que son usados por más de 200 millones de usuarios internacionalmente, creciendo éste número astronómicamente. En 1990, la WWW comenzó con un pequeño número de documentos cómo una solución casera de alrededor de mil usuarios en CERN [1]. En el año 2002 , la W3C (Worl Wide Web Consortium), estimó alrededor de un billón de usuarios y el crecimiento elevado de documentos disponibles. Este suceso y el crecimiento exponencial hacen difícil, ac encontrar, presentar, y mantener la información. Actualmente las páginas son elaboradas en formato HTML y SGML que emplean protocolos que permiten a los navegadores presentar ésta información a los usuarios. Sin embargo, ésta información es presentada en lenguaje natural.

Ésta representación en lenguaje natural permite que cada vez sea más dificil encontrar cualquier tipo de información. Es por esto que se han creado lo que se denomina la Web semántica, la cuál soluciona gran parte de estos problemas.
La web semántica está en estos momentos está siendo investigada, con el fin de resolver los múltiples problemas que se presentan en àreas como la recuperación de información, comercio electrónico, negocios electrónicos y administración de la información.

Para esto, en el trabajo presentado, se intenta mostrar los avances que se han producido en el campo de la web semántica en general, además de la presntación del método SWQ (Smart Web Query) realizado por los investigadores Roger H.L. Chiang. Cecil Eng Huang Chua y Veda C Storey a través de un prototipo de búsqueda.

El trabajo se divide en cinco partes, definidas de la siguiente manera:
En la primera parte, se conceptualiza la web semántica. La segunda parte, trata de un estudio hecho al tema de ontologías. La tercera parte, explica el método SWQ con sus caracterísitcas, la cuarta muestra el desarrollo del prototipo del buscador web semántico usando el lenguaje UML (Unified Model Language) y la quinta muestra un análisis estadístico realizado con el recall and precision.

1. LA WEB SEMÁNTICA

La web semántica ha tenido un crecido interés en los procesos de recuperación, indexación y edición de la información.
La historia de la web semántica pretende comprender, estructurar y administrar los contenidos (documentos) almacenados en la web ya sea en forma de texto, imágenes o graficos a partir de una valoración semántica de esos contenidos.
Los componentes básicos necesarios para la implementación de la web semántica son (a) representación del conocimiento, expresado por lenguajes de marcado (SGML, XML, RDF); (b) ontología, una disciplina que estudia las relaciones entre conceptos por medio de reglas lógicas de raciocinio, generando lenguajes que son comprendidos por los computadores y (c) agentes, que son programas creados para recolectar contenidos de la web a partir de diversas fuentes, procesando la información y permutando los resultados con otros programas a través de lenguajes que expresan inferencias lógicas resultantes de reglas e información como aquellas que son representadas por las ontologías.

Las aplicaciones de la web semántica tienen lugar en el comercio electrónico, con la producción de catálogos electrónicos eficaces y en sitios no comerciales.
De este modo el área de aplicación de la web semántica es extenso , ya que permite consultar “inteligentemente” la información deseada. Por ejemplo, es aplicable a la indexación de material que se encuentra en las Bibliotecas, contenidos electrónicos de los sitios web especializados en investigación, y en otros sitios más.
Para lograr esto se han desarrollado metodologias y herramientas para aplicar la web semántica, las cuales tienen algo en común, el uso de ontologías.

Por otra parte, la web semántica es la realización de un aspecto de la Web que fúe parte de un sueño de 1989, pero sólo hasta finales del año 2002 se han visto algunos resultados presentados por los diferentes centros de investigación.
La web semántica es específicamente una web de máquinas que entienden el significado de la información que se encuentra en la web, siendo definida por estándares que necesitan de una infraestructura interoperable que globalice los protocolos y estándares.[2].
La capa básica de representación de datos está estandarizada por el Resource Description Framework (RDF), el cuál es una recomendación de la W3C, en 1999. RDF es un estándar basado en XML (otra recomendación de W3C), URIs (creados en el año de 1989) y Unicode.
Encima de la capa de RDF se encuentra la capa de Construcción Ontologica , diseñada en lenguaje OWL, que fue desarrollado por el Grupo de Ontologías de la W3C. Éstas capas se encuentran dentro de la arquitectura de la web semántica que será vista en otro apartado.
La World Wide Web Consortium (W3C) es el lugar que permite a algunas empresas, centros de investigación y universidades entender y consensuar en los diferentes estandares que permiten su desarrollo. El Consorcio agrupa por lo menos a 230 organizaciones.
Sí el desarrollo de softare, hardware,redes de comunicaciones, o información es vendida o usada en la web como una parte crucial de la vida de sus negocios, estas compañias o empress son conducidas por áreas emergentes tales como la publicación Web, uso de Intranet, Comercio Electrónico, y Web basada en educación y entrenamiento. Para estos campos la W3C está trabajando y guíando a sus miembors para usar sus protocolos para cualquier área. Los protocolos, en éste contexto, son las reglas que permiten a los computadores dialogar ambos sobre un tópico dado. Cuando la industia de la computación estén de acuerdo con los protocolos para un área particular de la Computación, entonces los nuevos estándares serán

usados a nivel mundial, y los nuevos programas desarrollados incorporando los protocolos pueden todos trabajar y hablar el mismo idioma. Está es la clave para el desarrollo de la web.

1.1 DEFINICIÓN DE LA WEB SEMÁNTICA

La web semántica tiene diversas definiciones, cada una dependiendo del autor o centro de investigación que haga el estudio de ella. Aunque reune dos conceptos claves para su entendimiento que son: La web y la semántica, los cuales tienen el siguiente significado: La web es el sitio que almacena y maneja la información a través de documentos y la semántica es el estudio del significado de las palabras, teniendo en cuenta lo anterior se puede definir la web semántica como una estructuración de los datos para que tengan un contenido semántico y así poder darle un significado a los documentos que se encuentran en los sitios web.
Aunque existe la definición más acertada universalmente es la de Tim Berner Lee : " *La web semántica es una extensión de la web actual que presentará una estructura que posibilitará una comprensión y/o administración de los contenidos almacenados en la web, independiente de la forma en que estos se presentan, ya sea texto, sonidos, imágenes y gráficos, a partir de una valoración semántica de sus contenidos, y a través de agentes que son programas recolectores de contenido de diversas fuentes de información, capaces de procesar la información e intercambiar resultados con otros programas".*
La extensión de la web actual conlleva a que no es el reemplazo de la web actual, sino más bien al mantenimiento que se le hace a la web actual.

1.2 ARQUITECTURA DE LA WEB SEMÁNTICA

La arquitectura de la web semántica surge por la visión de Tim Berners Lee sobre la Web, aunque en algunas compañias y centros de investigación ahn modelado su propia arquitectura, ésta es la más común y conocida a nivel mundial.
Está arquitectura esta´constituida por tres capas o niveles, cada de las cuáles puede manejar sus propias herramientas y lenguajes. De esta manera, las capas son construidas y entendidas de abajo hacia arriba, es decir se debe cumplir con los requisitos de la primera capa (que es la de abajo) para seguir con la segunda capa (la capa de más arriba).

Teniendo en cuenta lo anterior, la arquitetura de la web semántica (Figura Nº 1), se representa de esta forma:

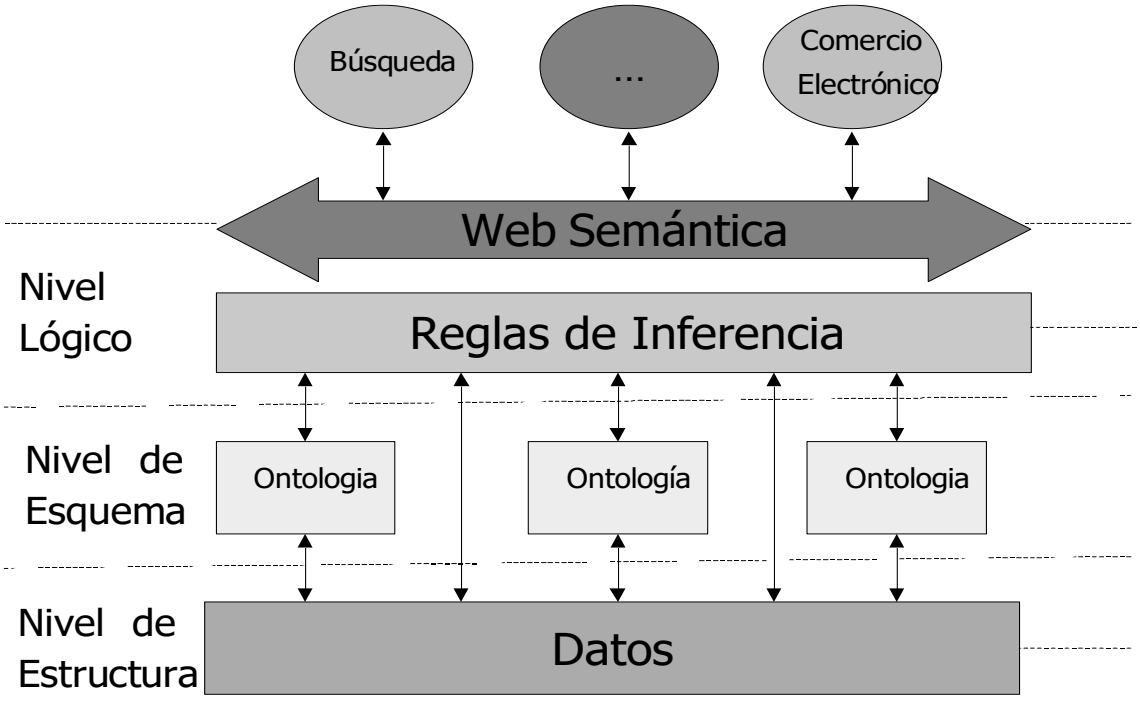


Figura Nº 1 La arquitectura de la web semántica

1.2 1. Nivel o Capa de Datos:
En está capa se estructuran los datos y se le da un significado a estos datos, teniendo en cuenta que está estructuración necesita de personas conocedoras del dominio que se requiere estructurar.
Los lenguajes conocidos para estructurar está información son XML (eXtensible Markup Language) y RDF (Resource Description Framework). Con cualquiera de los dos lenguajes se pueden estructurar los datos.
Está estructuración permite, como su nombre lo indica darle un ordenamiento a los datos para que sean fácilmente utilizados y entendidos por los mismos computadores, ya que manejan estándares diseñados para ello. Estos estándares fueron creados por la W3C (Worl Wide Web Consortium).

1.2.2 Nivel o Capa de Esquema:
Los datos anteriormente estructurados son tomados para establecerles las relaciones y las propiedades, para lograr con ello que tengan un significado.
Estas relaciones y propiedades que se le diseñan a los datos se asemeja al modelo entidad relación visto en Bases de datos y al paradigma orientado a objetos (actualmente usado para diferentes aplicaciones).
En este nivel se usan las ontologías para poder conceptualizar los datos y darle relaciones.

1.2.3 Nivel o Capa Lógica.
En esta capa se definen los mecanismos para realizar inferencia sobre los datos, lo que significa que está capa está compuesta por reglas de inferencia
Estas reglas de inferencia permiten a los agentes computacionales raciocinar sobre la estrutura de los datos.
Existen lenguajes de representación del conocimiento para lograr esto, entre los cuáles se destacan: OIL (Ontology Inference Layer) y DAMl + OIL.

1.3 APLICACIONES DE LA WEB SEMÁNTICA

La web tiene nuevas posibilidades para el acceso a la información y al comercio electrónico. Aunque el crecimiento exponencial que tiene la web ha creado problemas en el manejo de la información que se encuentra en ella.
Por esto, las áreas de aplicación d ela web semántica son similares a la de la web actual, aunque se debe cambiar algunos conceptos que se llevan en la web actual.
Entre las posibles áreas de aplicación de la web semántica se encuentra:

1.3.1 Área de administración del conocimiento:
Ésta área réune lo concerniente a la adquisición, mantenimiento y acceso al conocimiento de una organización. Ésta área apunta a aprovechar los recursos de una organización intelectual para tener una alta productividad, nuevos valores, e incrementar la competitividad.
Debido al gran número de documentos que existen disponibles en las organizaciones, han entrado al mercado sistemas administradores de estos documentos. Aunque estos sistemas tienen algunos problemas:

Buscar Información: Existe en el momento palabras claves o keywords que recuperan información irrelevante mostrando información que no pertenece al contexto requerido por el usuario.

Extracción de Información: El navegador humano es actualmente requerido para extraer información relevante de recursos de información, como agentes automáticos que pierden el sentido común del conocimiento requerido para extraer dicha información de representaciones textuales y pierden la integración de la información sobre diferentes recursos.

Mantenimiento: El mantenimiento estructurado débil de los recursos de textos es una dificultad y consumen demasiado tiempo cuando estos recursos son extensos. Guardar tal colección consistente, correcta y actualizada requiere un mecanismo de representación semántico y restringido que ayude a detectar las anomalias.

Generar automáticamente documentos: Los sitios web adaptativos permiten una reconfiguración dinámica de la información de acuerdo al perfil de usuario y otros aspectos relevantes que son útiles. La generación de información semiestructurada representada por datos semiestructurados requiere de una representación accesible por las máquinas de la semántica de estos recursos de información y tal representación actualmente no existe.

La tecnología de web semántica permiten estructurar y definir la semántica de los documentos trayendo nuevas posibilidades tales como: buscadores inteligentes en vez de escogencia de palabras claves o keywords , consultas en vez de recuperación de información.

1.3.2 Área del Comercio Electrónico.

El comercio electrónico (B2C) es un área de negocio importante y crecida. Esto se debe a dos razones: Primero, es ampliada a modelos de negocios existentes, lo cuál reduce costos y extiende la la existencia de canales de distribución, y puede introducir nuevas posibilidades de distribución. Un ejemplo de los modelos de negocios son los almacenes en línea.
Segundo, ésta área permite entrar en nuevos modelos de negocios y darle una mayor importancia. Ejemplos de nuevas áreas de modelos de negocios son: shopping agents, supermercados en línea, y compraventa de casas.
Aunque existen problemas por resolver en ésta área tales como el esfuerzo que necesita el usuario para conseguir un producto, es decir el tiempo que consume para adquirirlo. Además el mantenimiento de estos sitios requiere de un gran esfuerzo.
También se presenta que la información extraida por los shopbots es incompleta y usan tecnología limitada (básicamente en la información de los precios). Por ejemplo extraen información de los costos directos del producto más no de los indirectos.

Estos problemas son causados debido a que la mayorìa de la información de los produc
proveída en lenguaje natural, y el reconocimiento de los textos automáticamente es todavía un área con problemas significantes por resolver.
Se requiere es una máquina que procese la semántica de la información. Ësta situación cambiará cuando los formalismos de representación estándar para la estructura y la semántica de los datos estén disponibles.

2. LAS ONTOLOGÍAS

Las ontologías se han convertido en un área de investigación para diferentes entidades, que en su mayoría son universidades. Cada una de ellas ha definido una metodología para poder trabajar sobre ésta área (las ontologías).
Las universidades que realizan investigación concienzuda de los mayores problemas que aqueja la web actual están abanderando estos estudios, que han llevado a presentar productos en ésta área.
Las ontologías según aquellas investigaciones nacen del estudio hecho a lo que se denomina la web semántica, la cuál es representada por medio de las ontologías. La construcción de las ontologías dependen de la metodología que utilizan estos centros de estudios, aunque ellas poseen en gran porcentaje las mismas características, las cuáles serán descritan en este apartado.
De esta manera, las ontologías se puede decir que son la esencia de la web semántica actual, lo que indica que una web semántica bien construida debe tener ontologías.
Las ontologías en su mayoría son diseñadas por personas conocedoras del dominio del área de la cuál se va a construir la web semántica, lo que indica que cualquier persona del común no las puede construir, y que se necesita de personas expertas en estos conocimientos para su construcción.
Por otra parte las ontologías representan la solución a la cantidad de información que se encuentra en la web, (la cuál fue diseñada en su momento para la vista del usuario, y no para analizar la información que poseen ellas), a través de una estructuración significativa de esta información para poder ser analizada y valorada.

2.1 DEFINICIÓN DE ONTOLOGÍA

El témino ontología definido en su tiempo por Aristóteles en su intento de clasificar todo lo que existe en el mundo, la palabra ontología proviene de la conjunción de dos términos *Ontos* y *Logos*, que denotan existencia y mundo respectivamente, término que ha tomado la Inteligencia Artificial (AI) como base para adoptar el término ontología para describir todo lo que puede ser representado computacionalmente.
La definición más citada en el área de la Inteligencia Artificial es la de Gruber: *"Una ontologia es la especificación explícita de una conceptualización."* [3].
Para la Inteligencia Artificial lo que existe es exactamente aquello que puede ser representado computacionalmente.
Cuando el conocimiento de un dominio es representado en un formalismo declarativo, el conjunto de objetos que pueden ser representados es llamado el *universo de discurso*. Este conjunto de objetos, y las descriptibles relaciones entre ellos, es refle jado en el *vocabulario* con el cual se representa conocimiento.
En una ontología se asocian mediante definiciones los nombres de las entidades en el universo de discurso, (por ejemplo clases, relaciones, funciones o otros objetos) con cierto texto leíble. humanamente que las describe y con axiomas que restringen la interpretación y otorgan la característica de "bien formados" a dichos términos.

Teniendo en cuenta que existen diferentes centros de estudios a nivel mundial dedicados al estudio de las ontologías no existe una diferencia en el concepto mundial que se tomó para definir las ontologías.
El concepto dado por Gruber en su momento fue cambiado ligeramente por la de Borst de la siguiente manera : **"Las ontologías se definen como la especificación formal de una conceptualización compartida".**[4]
Studer, Benjamins, y Fensel [5] agregaron expresividad a las definiciones de Gruber y Borst explicitando:
Conceptualización se refiere a un modelo abstracto de algún fenómeno en el mundo proveniente de haber identificado los conceptos relevantes de dicho fenómeno.
Explícita se refiere a que el tipo de concepto usado y la restricciones para su uso son explícitamente definidas.
Formal se refiere al hecho de que la ontología debería ser leíble o interpretable por un computador.
Compartida refleja la noción de que una ontología captura conocimiento concensual, esto es, no es privado de un individuo, sino aceptado por un grupo.
De esta manera, nos quedamos que con la definición anterior, ya que es la más completa y que se asemeja más a lo que realmente significa una ontología.

2.2 COMPONENTES DE UNA ONTOLOGÍA

Los componentes de una ontología varía de acuerdo a la metodología o arquitectura definida por los investigadores para su posterior uso. Es por esto que los componentes que se describirán en este apartado son genéricos, lo que significa que son los que mayormente se utilizan para su constucción, aunque esto no implica que sí algún componente no se encuentra en la construcción de la ontología, está se encuentra mal construida porque no cumplió con alguno de ellos.
Los componentes varían de acuerdo a las diferentes investigaciones relaizadas en diferentes centros de investigación, aunque su construcción depende es del concepto definido anteriormente.
De acuerdo a lo anteriormente explicado tomamos los componentes genéricos de las ontologías y que le dan una mayor representatividad.

2.2.1 Conceptos

Un concepto puede ser cualquier cosa acerca de la cual algo se pueda aseverar, y por tanto puede ser eventualmente un objeto físico, la descripción de una tarea, función, acción, estrategia, etc.
Cada concepto tiene un término asociado como nombre y un conjunto de atributos que lo identifican, asociándose un concepto a un objeto en Lenguaje Unificado de Modelado (UML), de tal forma que las personas que han trabajado con este lenguaje se le facilitarán las cosas en

la construcción de conceptos, además de las personas que han trabajado con lenguajes orientado a objetos.

2.2.2 Relaciones

Las relaciones representan el tipo de interacción entre los conceptos de un dominio.
Existen múltiples relaciones entre los conceptos que enriquecen la construcción de las ontologías.
Como ejemplos se pueden citar las relaciones binarias subclase-de (IS-A) o es-parte-de (PART-OF), Actua_Sobre (ACTS ON)
Otros ejemplos de relaciones ontológicas son las relaciones temporal (que implica precedencia en el tiempo) y topológica (que implica conexión espacial entre objetos).
Las relaciones pueden usan los conceptos de taxonomía y mereorología.

Taxonomía

La taxonomía que clasifica los términos y los jerarquiza es utilizada en los modelos ontológicos.
En general los modelos ontológicos definen la relación taxonomía como irreflexiva, transitiva y asimétrica.
En tanto que la transitiva permite inferir en la estructura, las restantes se prestan para chequear consistencia

Mereología

Para modelar la relación de que un conjunto de conceptos son todas las partes que constituyen otro objeto (part-whole) (parte-todo), la mereología se suele definir con las propiedades irreflexiva y asimétrica, en tanto que para la transitiva hay elementos que dificultan su definición.
Cabe destacar la diferencia existente entre una relación de partonomía y una de mereología. En tanto que la relación partonomía solo define en una estructura algunos objetos que contribuyen a conformar un todo, una mereología establece todos los objetos que conforman el todo, sin alguno de los objetos que contribuyen en la formación, el todo no sería posible.

2.2.3 Axiomas

Los axiomas se usan para modelar verdades que se cumplen siempre en la realidad modelada.
Los axiomas definidos en una ontología pueden ser estructurales o no estructurales.
Un axioma estructural establece condiciones relacionadas a las jerarquías de la ontología, conceptos y atributos definidos.
Un Ejemplo de axioma estructural pueden ser “El concepto A no es una clase del Concepto A”.
Los axiomas no estructurales establecen relaciones entre atributos de un concepto, y son específicos de cada dominio.
Un ejemplo de axioma no estructural puede ser la relación **F=m*a** que debe cumplirse siempre entre los atributos **F** (fuerza), **m** (masa) y **a** (aceleración) de un determinado concepto.

2.2.4 Instancias

Se usan para representar elementos del dominio de la ontología.

2.3 CLASIFICACIÓN DE LAS ONTOLOGÍAS

Algunos autores han clasificado las ontologías existentes según diferentes criterios, a continuación se exponen algunos de ellos.

2.3.1 Clasificación por grado de axiomatización

Según Sowa [7] las ontologías se pueden clasificar en Terminológicas o Formales de acuerdo al grado de axiomatización que tenga la definición de sus categorías.

Ontologías Terminológicas

Una ontología terminológica define términos y sus relacionamientos en taxonomías que involucran relaciones de subtipo y supertipo y mereologías que relacionan partes con un todo (part-whole), pero no incluyen axiomas y definiciones expresadas en lógica o algún tipo de lenguaje interpretable por un computador que podría automáticamente ser traducido a lógica, esto hace que se disponga de menos información del universo modelado, pero permite a su vez, por la relativa simplicidad de su construcción, construir ontologías de gran tamaño.

La mayoría de los campos de la ciencia, ingeniería, negocios y jurídico, han desarrollado sistemas de terminología o nomenclatura para designar, clasificar y estandarizar sus conceptos en enormes ontologías terminológicas.
Ejemplos de estas ontologías son:
EDR. The Electronic Dictionary Research proyect, es un proyecto japonés que ha desarrollado un diccionario con mas de 400.000 conceptos, con su mapeamiento a palabras en inglés y japonés con muy poco detalle acerca de cada uno de sus términos.
WordNet, es una jerarquía de 166.000 palabras, que de hecho es la ontología mas usada para procesamiento de lenguage natural

Ontologías Formales
Una ontología formal tiene sus categorías restringidas por axiomas y definiciones expresadas en lógica o en algún tipo de lenguaje interpretable por computador. Las ontologías formales tienden a tener menos cantidad de conceptos que las terminológicas, pero sus axiomas y definiciones pueden soportar computaciones e inferencias mas complejas.
La diferencia entre una ontología terminológica y una formal, según Sowa, es de grado, teóricamente en tanto se adicionen axiomas a una ontología terminológica, eventualmente podrá evolucionar a una formal, pero la definición de axiomas no es una tarea trivial, y en por eso en la práctica es difícil tal evolución.

2.3.2 Clasificación por dependencia del contexto
Según Mizoguchi, Vanwelkenhuysen e Ikeda [8] las ontologías pueden clasificarse según el grado de dependencia del contexto que presenten, en el sentido de que aquellas menos dependientes del contexto serán las candidatas a ser reusadas. La clasificación que surge es:

Ontologías de dominio
Expresan conceptualizaciones que son específicas a un dominio particular, colocando restricciones en la estructura y contenido de un dominio de conocimiento mediante axiomas que se cumplen siempre entre los elementos de dicho domínio. Su principal cometido es permitir el reuso de la ontología para diferentes aplicaciones que involucren al mismo domínio.
Ejemplos de estas ontologías son:
The EngMath ontology, disponible en la biblioteca de ontologías de Ontolingua.
The Enterprise Ontology que se ubica en el dominio del modelado de procesos empresariales, es por tanto una colección de términos y definiciones relevantes al entorno empresarial.
Las Ontologías TOVE (Toronto Virtual Enterprise) fueron desarrolladas en la universidad de Toronto y también están destinadas al modelado de procesos empresariales, a saber: Enterprise Design Ontology, Project Ontology, Scheduling Ontology y Service Ontology

Ontologías generales o de sentido común
Incluyen vocabulario relacionado a cosas, eventos, tiempo, espacio, causalidad, comportamiento, función, etc. La ontología CYC disponible en http://www.Cyc.com es una ontología de este tipo, que provee una gran cantidad de conocimiento humano fundamental. Está además dividida en micro
teorías que permiten su mejor administración.

Meta Ontologías, genéricas o Core Ontologías
Son similares a las de dominio, pero los conceptos que definen se consideran genéricos a través de diferentes áreas de conocimiento y por ello reusables en diferentes domínios.
Típicamente, las ontologías genéricas definen conceptos como estado, evento, proceso, acción, etc.
Los conceptos de las ontologías de dominio son a menudo definidos como especializaciones de conceptos existentes en ontologías genéricas. Cabe destacar que el límite entre que considerar una ontología genérica es vago, pero la distinción es intuitivamente significativa y útil cuando es necesario organizar ontologías en bibliotecas.
Un ejemplo de este tipo de ontologías es The Mereology Ontology propuesta por W. N. Borst. [6] y que define la relación part-of y sus propiedades.

Ontologías de tareas
Proveen un sistemático vocabulario de los términos usados para resolver problemas asociados con tareas particulares, ya sean dependientes o no de dominio. Por ejemplo, al respecto de la

tarea diagnóstico, tendremos términos que involucren a los conceptos de “observación” “hipótesis” y “objetivo”, así como el término que refiera a la acción “generación de hipótesis”.

2.3.3 Clasificación por el sujeto de conceptualización

Según Van Heijst, Schreiber y Wielinga [HSW97] las ontologías pueden clasificarse de acuerdo al sujeto de su conceptualización en:

Ontologías de aplicación

Contienen todas las definiciones que son necesarias para modelar el conocimiento requerido para una aplicación particular en un dominio dado. Típicamente son una mezcla de conceptos provenientes de ontologías de dominio y de ontologías genéricas. Las ontologías de aplicación no se construyen con el propósito de lograr usabilidad en diferentes dominios.

Ontologías de dominio

Coincide con la clasificación propuesta por Mizoguchi, Vanwelkenhuysen e Ikeda referenciada antes en la clasificación por dependencia del contexto.

Ontologías genéricas

Coincide con la clasificación propuesta por Mizoguchi, Vanwelkenhuysen e Ikeda [4] referenciada antes en la clasificación por dependencia del contexto.

Ontologías de representación de conocimiento

Describen las primitivas de representación usadas para formalizar conocimiento en paradigmas de representación de conocimiento, esto es que explican la conceptualización que subyace a un formalismo de representación de conocimiento. Se pretende que sean neutrales con respecto a las entidades de mundo, esto es, proveen un marco represetacional sin hacer aseveraciones acerca del mundo.

Las ontologías genéricas y de dominio son descriptas usando ontologías de representación.

El ejemplo mas representativo de esta clase de ontologías es la Frame-Ontology disponible en el Ontolingua Server: http://www-ksl-svc.stanford.edu.

Ontologías de Dominio-Tarea

Según Benjamins y Gómez Pérez [7] se pueden distingir en este tipo aquellas ontologías ya clasificadas como ontologías de tareas por [4] pero que se distinguen por ser reusables en un determinado dominio, pero no a través de otros dominios.

2.3.4 Ontologías Top-Level

Pretenden establecer una estructura básica, bajo la cual todos los términos en cualquier ontología existente deberían poder relacionarse. Hasta ahora el principal problema es que no existe una ontología única de este tipo

2.4 AREAS DE APLICACIÓN

A continuación se citan algunas de las áreas en que actualmente se usan ontologías.

Ingeniería del Conocimiento

En los 1990, la ingeniería del conocimiento ha emergido como un campo maduro, diferente, pero estrechamente relacionado al de la ingeniería de software. Entre sus aspectos diferentes hay un conjunto de técnicas de elicitation (extracción) y modelado del conocimiento, un conjunto de formalismos para representar conocimiento, y un conjunto de mecanismos para automatizar razonamiento.

Según Preece y colegas [9] el aporte de las ontologías en el proceso de ingeniería del conocimiento está en las siguientes etapas:

modelado conceptual, donde se crea un glosario de la terminología del dominio de la aplicación (los conceptos), se definen relaciones entre dichos términos y restricciones en su uso. Este modelo conceptual explícito es la *ontología.*

*Construcción de la base de conocimiento.*Usando la ontología definida en la etapa anterior como un conjunto de esquemas o contenedores de conocimiento, se puebla la base de conocimiento con instancias del dominio bajo la forma de reglas, hechos, eventos y restricciones.

Las ontologías, conjuntamente con los métodos de solución de problemas (PSMs) son prometedores candidatos para posibilitar el reuso de componentes en Ingeniería del Conocimiento. En tanto que las ontologías definen el conocimiento declarativo del dominio a un nivel genérico, los PSMs especifican
conocimiento de razonamiento sobre dicho dominio.

Procesamiento de lenguaje natural
En procesamiento de lenguaje natural, una ontología puede mantener la definición de elementos gramaticales del lenguaje y sus relacionamiento, permitiendo por ejemplo el análisis sintáctico de un texto.

Interoperabilidad entre sistemas heterogéneos
En la interoperabilidad de sistemas, las ontologías se presentan como una importante solución para lograr una integración inteligente, en particular en el área de bases de datos, una ontología puede ser un elemento clave asociado a un mediador que integra datos provenientes de wrappers que recogen información de fuentes heterogéneas. Eventualmente el mediador podrá estar destinado a realizar operaciones de integración destinadas a la carga o proceso de refresh de un datawarehouse.
Con una ontología terminológica, se pueden organizar los términos que son usados en interacciones entre sistemas heterogéneos, de manera de reconocer cuando una aplicación está usando un término que es mas general o mas específico que otro que está en uso por otra aplicación.
Si la ontología es formal, se puede contar con una definición mas completa de cómo se relaciona un término de un origen con el de otro, y eventualmente usar axiomas definidos que los vinculen por igualdad o que expresen un término exactamente en función del otro, lo que permitiría establecer correspondencias seguras y automáticas entre ellos.

Indexación de sitios Web
Desmontils y Jacquin en [10] presentan un procedimiento para indexar un sitio Web con apoyo de una ontología terminológica (figura 1). En una primera etapa se extraen los términos mas relevantes de cada página con ayuda de las marcas de HTML , luego se asocian a dichos términos, conceptos candidatos a través del uso de un Thesaurus. A continuación para cada concepto candidato, su capacidad para representar a la página es evaluada determinando su nivel de representatividad dentro de ésta, y finalmente el índice es construído. A cada concepto de la ontología son asociadas las páginas del sitio web donde está contenido.

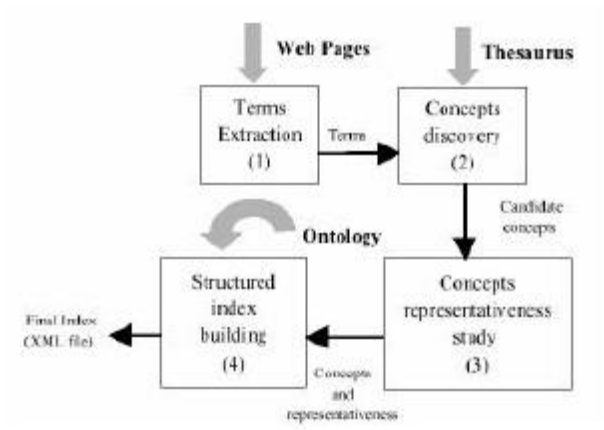


Figura 2. Procedimiento de indexación de un sitio web

De esta forma las consultas no serán procesadas a un nivel terminológico, sino conceptual, con los beneficios que eso supone en el grado de acierto de las respuestas retornadas.

Modelado de empresas
En el área de modelado de empresas, las ontologías desempeñan entre otros, el rol de mantener una memoria organizacional colectiva que permita a los distintos niveles de la empresa interoperar en un lenguaje común y con reglas únicas.

Las Ontologías TOVE (Toronto Virtual Enterprise) desarrolladas en la universidad de Toronto y la Enterprise Ontology desarrollada en la Universidad de Edimburgo, son ejemplos de este tipo de ontologías.

2.5 METODOLOGÍAS DE DISEÑO

Teniendo presente que no existe un consenso de una metodología que reúna todas las condiciones que debe tener una ontología, se han desarrollado varías metodologías. Estas metodologías dependen de los grupos de investigación involucrados en la construcción de la web semántica a través de las ontologías, es decir tantos grupos haya investigando el área en cuestión (web semántica), así será el número de metodologías.
Las metodologías más importantes vistas hasta el momento son las siguientes:

2.5.1 Metodología de Gruninger y Fox

Existen unos pasos [11] para ésta metodología, ellos son:

1 Definición de los escenarios motivadores.
El desarrollo de ontologías es motivado por escenarios, que surgen en las aplicaciones. En particular, tales escenarios pueden ser presentados como la descripción de problemas que no han podido ser resueltos hasta el momento.Un escenario motivador suele proveer un conjunto de intuitivamente
posibles soluciones a los problemas enumerados. Esas soluciones dan una semántica informal para los objetos y relaciones que serán luego incluidos en la ontología.
Cualquier propuesta de creación de una ontología o extensión de una existente, debería describir uno o mas escenarios motivadores, y el conjunto de soluciones proyectadas a los problemas de dichos escenarios.

2 Especificación informal de las 'competency questions'.
Dado el escenario motivador, un conjunto de preguntas surgirán como demandas que la ontología a construir debería resolver. Se pueden considerar esas preguntas como requerimientos en forma de interrogantes (competency questions). Una ontología debe ser capaz de representar esas interrogantes usando su terminología, y ser capaz de caracterizar las respuestas a ellas usando axiomas y definiciones. Se consideran informales, pues no están aún expresadas en el lenguaje formal de la ontología.

Las competency questions o interrogantes, deberían ser definidas de una manera estratificada, de forma que la respuesta a una interrogante puede ser usada para responder interrogantes mas generales de la misma u otra ontología por medio de operaciones de composición y descomposición, esta estratificación, en realidad es una forma de identificar conocimiento para reuso.
Las interrogantes no generan compromisos ontológicos de diseño, sino que son usadas luego, para evaluar los compromisos ontológicos que se han hecho en cuanto a la capacidad de la ontología para satisfacer los requerimientos que plantean.

3 Especificación de la terminología.
A partir de las interrogantes formuladas en el paso anterior, se puede extraer el conjunto de términos usados para expresarlas y considerarlo como base para la terminología a usar en un lenguaje formal.
Si se está diseñando una nueva ontología, para cada interrogante informal planteada, debe haber objetos, atributos, y relaciones en la ontología en construcción que son intuitivamente requeridos para resolver la interrogante.
El primer paso en la especificación de la terminología de la ontología es identificar los objetos en el dominio de discurso, serán representados por constantes y variables del lenguaje.
Atributos de objetos y relaciones serán definidos luego, usando predicados en lógica de primer orden, o en un formalismo equivalente como podría ser KIF.

5 Especificación de axiomas formales.
Los axiomas en la ontología especifican las definiciones de términos y restricciones en su interpretación, son definidos como sentencias en lógica de primer orden. Para los autores [12]

un conjunto simple de objetos, o un conjunto de cláusulas básicas en lógica de primer orden, no constituye una ontología.
Deben proveerse axiomas para definir la semántica del conjunto. Si el conjunto de axiomas es insuficiente para representar las competency questions y caracterizar sus soluciones, es un indicativo de que nuevos objetos o axiomas deben ser adicionados a la ontología hasta lograrlo. Se puede ver entonces que la construcción de axiomas y la evaluación de las competency
questions es un proceso iterativo.

6 Verificar la completitud de la ontología
Una vez que las competency questions han sido formalmente enunciadas se deben definir las condiciones bajo las cuales las soluciones son consideradas completas.
La figura 3 ilustra las etapas de la metodología de Gruninger y Fox.

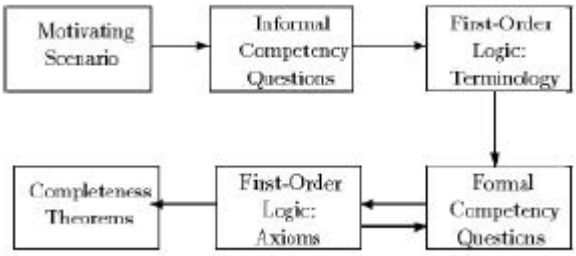


Figura 3. Etapas en la metodología de Gruninger y Fox

Ontologías desarrolladas usando la metodología
Esta metodología fue usada para construir las ontologías del proyecto TOVE (Toronto Virtual Enterprise) en el Enterprise Integration Laboratoy de la Universidad de Toronto. Dichas ontologías constituyen un modelo integrado y formalizado usando lógica de primer orden que incluye Enterprise Design Ontology, Project Ontology, Scheduling Ontology y Service Ontology.

2.5.2 Metodología de Unschold y King
Los pasos propuestos por esta metodología son [11]:

1 Identificar el propósito
Esto es, clarificar porque la ontología está siendo construída y que uso se pretende darle.

2 Construir la ontología
Donde se distinguen tres pasos:

2.1 Capturar la ontología

- Identificando los conceptos claves y relaciones en el dominio de interés.
- Produciendo definiciones textuales precisas y carentes de ambigüedad que describan tales conceptos y relaciones.
- Identificando lo términos para referirse a tales conceptos y relaciones

Los autores usan un enfoque middle -out para ejecutar esta etapa, y recomiendan que antes de buscar los conceptos mas generales o particulares como conceptos clave, se identifiquen los conceptos mas importantes, que luego serán usados para obtener el resto en la jerarquía por medio de generalización y especialización.

2.2 Codificación
Involucra explícitamente representar el conocimiento adquirido en el paso anterior, en un lenguaje formal.

2.3 Integrar ontologías existentes.
Durante los procesos de captura y codificación plantearse la pregunta de si corresponde y cómo, usar ontologías que ya existen.

3 Evaluación
Los autores adoptan la definición de Asunción Gómez Pérez, N. Juristo y J.Pazos “hacer una evaluación de las ontologías, sus ambientes de software asociados, y documentación con

respecto al marco de referencia, donde el marco de referencia puede ser especificaciones de requerimientos, competency questions, y /o el mundo real"

4 Documentación
Se recomienda el establecimiento de pautas para evaluar la ontología construída, posiblemente diferentes de acuerdo al tipo y propósito de la ontología.

Ontologías desarrolladas usando la metodolo gía
El proyecto mas importante que se desarrolló usando esta metodología es The Enterprise Ontology, que es una colección de términos y definiciones relevantes a empresas de negocios. La ontología fue desarrollada bajo el Enterprise Proyect del Artificial Intelligence Applications Institute de la Universidad de Edinburgo, con la colaboración de IBM, Lloyd's Register, Logica UK Limited, y Unilever.

2.5.3 METHONTOLOGY

Esta metodología fue desarrollada dentro del laboratorio de Inteligencia Artificial de la Universidad Politécnica de Madrid. El framework METHONTOLOGY permite la construcción de ontologías en el nivel de conocimiento de Newel [14]. Se identifican las siguientes etapas:

1 Establecimiento del proceso de desarrollo de la ontología
En esta etapa se identifican que actividades serán realizadas al construir la ontología.

> **Actividades de gerenciamiento del proyecto**
> Incluyen planificación, control y aseguramiento de la calidad.
> ***Planificación***, identifica cuales tareas tienen que ser ejecutadas, como serán organizadas , cuanto tiempo y que recursos se necesitan para completarlas. Esta actividad es escencial para ontologías que necesitan usar otras ontologías que ya han sido construídas o ontologías que requieren niveles de abstracción y generalidad.
> ***Control*** , para garantizar que las tareas planificadas son completadas en la manera en que fueron planificadas.
> ***Aseguramiento de la calidad***, para asegurar que la calidad de todos y cada uno de los productos liberados, ya sea ontología, software o documentación, es satisfactoria. [13] describe como esas actividades se pueden ejecutar.
> **Actividades orientadas al de sarrollo del proyecto**
> Incluyen especificación, conceptualización, formalización, implementación y mantenimiento.
> ***Especificación de requerimientos*** dice porqué la ontología está siendo construída y cuales son los usos que se pretende darle, asi como también quienes son sus usuarios finales.
> ***Conceptualización*** estructura el conocimiento de dominio como un modelo significativo al nivel de conocimiento [13].
> ***Formalización*** transforma el modelo conceptual en un modelo formal o semi-computable.
> ***Implementación*** construye modelos computables en un lenguaje computacional.
> ***Mantenimiento*** actualiza y corrige la ontología. [14]

Actividades de apoyo
Incluye una serie de actividades ejecutadas al mismo tiempo que las actividades orientadas al desarrollo, sin las cuales la ontología podría no llegar a ser construída. Ellas incluyen adquisición de conocimiento, evaluación, integración , documentación y administración de versiones.

> ***Adquisición de conocimiento*** En esta etapa se adquiere el conocimiento del dominio.
> ***Evaluación*** hace un juzgamiento técnico de la ontología, su ambiente de software y documentación con respecto al marco de referencia durante cada fase y entre fases de su ciclo de vida.
> ***Integración de Ontologías*** es requerida cuando se construye una nueva ontología reusando otras ya disponibles.
> ***Documentación*** detallada, clara y exhaustiva es necesaria para cada una de las fases completadas y productos generados.

Administración de versiones (Configuration Management) Un registro de todas las versiones de la documentación, código de software y código de la ontología, para control de cambios, debe ser mantenida.

Cabe agregar que hasta el presente ninguna de las herramientas disponibles para construcción de ontologías resuelven satisfactoriamente el tema del versionamiento.

2 Ciclo de vida de la ontología

El ciclo de vida propuesto para METHONTOLOGY incluye las etapas descriptas como actividades en la figura Nº 4. como actividades orientadas al desarrollo, estas son: especificación de requerimientos,conceptualización, formalización, implementación y mantenimiento

En la figura 3 se muestra la representación gráfica que realizan sobre tal ciclo.

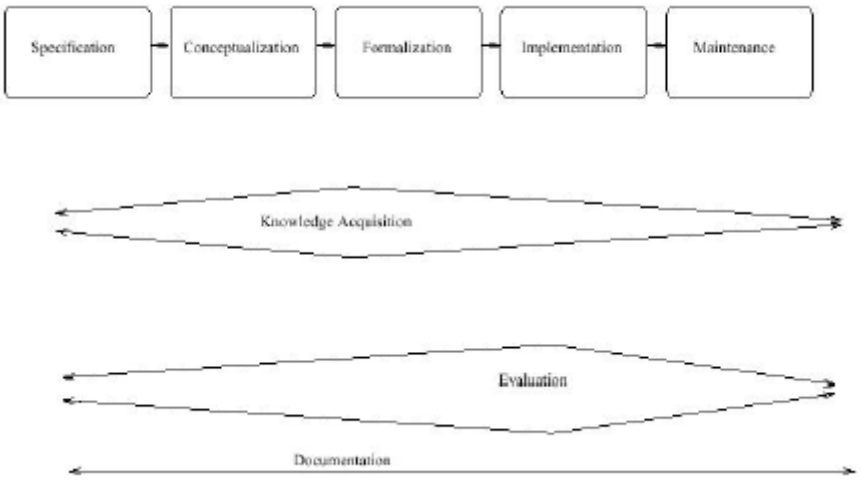


Figura 4 Ciclo de vida en METHONTOLOGY

Ontologías desarrolladas usando la metodología

CHEMICALS que contiene conocimiento sobre el dominio de elementos químicos y estructuras cristalinas.

Environmental pollutants ontologies que representa los métodos para detectar los diferentes componentes contaminantes de diversos medios como agua, aire, suelo, etc. y las máximas concentraciones permitidas para esos componentes, tomando en cuenta toda la legislación vigente (Unión Europea, España, Alemania, reglamentaciónes de Estados Unidos, etc.)

The Reference-Ontology, una ontología en el dominio de las ontologías (una meta-ontología) que juega el rol de una especie de páginas amarillas de ontologías. Reúne, describe y relaciona ontologías existentes, usando una organización lógica común.

La versión estructurada de la ontología (KA)2 que contiene conocimiento acerca de la comunidad científica en el área de adquisición de conocimiento, particularmente: científicos, tópicos de investigación, proyectos, universidades, etc.

Aplicaciones que usan ontologías desarrolladas con esta metodología

(Onto)Agent es un www broker basado en ontologías, que usa The Reference-Ontology como origen de su conocimiento y devuelve descripciones de ontologías que satisfacen un determinado conjunto de restricciones propuestas.

Chemical OntoAgent, es un broker para la enseñanza de química a través de internet que permite a los estudiantes aprender química y testear su aprendizaje en este dominio. Usa CHEMICALS como fuente de conocimiento.

Ontogeneration, es un sistema que usa la ontología de dominio CHEMICALS y la ontología linguística GUM para generar descripciones textuales en español en respuesta a consultas de estudiantes en el dominio de la química.

3. A SMART WEB QUERY METHOD FOR SEMANTIC RETRIEVAL OF WEB DATA (SWQ).[1]

El método SWQ es el resultado de la investigación desarrollada por los investigadores Roger H.L. Chiang. Cecil Eng Huang Chua y Veda C Storey.
Este método nace gracias a la dificultad que existe en crear buscadores web semánticos. En este método se usan dominios semánticos representados en un contexto ontólogico para especificar y formular apropiadamente la consultas web. El método SWQ usa filtros de de búsqueda semántica para identificar y rankear las páginas web semi-automáticamente. Además, permite estructurar las ontologias en una estructura flexible representada en el contexto ontológico.

3.1. RECUPERACIÓN DE INFORMACIÓN SEMÁNTICA

Probablemente el recurso más fácilmente accesible es la World Wide Web (WWW).
Los buscadores web, sin embargo, son algunas veces difíciles de crear por diversas razones. Primero, las máquinas de búsqueda (buscadores web) son todavía muy primitivas. La búsqueda requerida debe ser especificada como una colección de palabras claves (keywords), separados a su vez por operadores booleanos. Razonablemente los niveles de precisiòn y recall pude requerir de poner muchas palabras claves.
Segundo, la búsqueda con contenido sensitivo no es alcanzable en el momento. Los buscadores web pueden solamente recuperar datos sobre una base puramente sintántica. No es poseible obtener un dominio específico de un dominio en la búsqueda web.
Tercero, los datos web aparecen en una extensa variedad de formatos heterogéneos tales como: docuemnto PDF´s , imágenes y sonido.
Finalmente, no es factible establecer un riguroso estandar para el modelamiento de datos web y su desarrollo (tal como el modelo de datos relacional y XML), ya que las páginas web son desarrolladas llevando un proceso no regular. Por esto es imposible desarrollar un lenguaje formal de búsqueda web para la WWW.

3.2 MARCO TEÓRICO

El SWQ (Smart Web Query) es una investigación desarrollada por Roger H. L. Chaing, Cecil Eng Huang Chua y Veda C. Storey. Es un método diseñado para realizar recuperación de información semántica, basada en una máquina SWQ.
La máquina SWQ facilita la formulación de consultas web, capturando no solamente la búsqueda del usuario, sino el dominio semántico relacionado a la búsqueda del usuario.
Especificamente, el SWQ (a) desarrolla un método que refina las consultas especificadas del usuario (Keywords), explorando los dominios semánticos.(es decir, conocimiento e información), (b) presenta una estrutura flexible para organizar dominios semánticos a través de un contexto ontológico. (c) Propone filtros de búsqueda semántica para identificar y rankear las páginas relevantes, empleando dominios semánticos organizados en el contexto ontológico. y (d) desarrolla una máquina SWQ para probar y evaluar el propósito de éste método.

3.3 APLICACIONES DEL MÉTODO SWQ

El desarrollo y la aplicación del método SWQ está relacionado en la investigación de dos (2) áreas:
a Tecnología de búsqueda web y
b Desarrollo y uso de ontologías en buscadores web.

a Tecnología de búsqueda web
Existen cuatro formas principales para buscar datos web no estructurados. (1) Búsqueda sintáctica, (2) Búsqueda de metadatos, (3) QBE, (4) Búsqueda autónoma navigacional.
(1) La búsqueda sintáctica es la búsqueda más común sobre datos semiestructurados o no estructurados y es usada por los buscadores más comunes. Los documentos son encontrados, clasificados y referenciados en una base de datos por una persona experta o por agentes/spiders. La instancia de la base de datos es accedida a través de consultas

[1] H. L. Chaing, ROGER. Data & Knowledge Engineering.38 (2001). P. 63-84

booleanas (palabras claves (keywords) separadas por un oprador booleano como OR,AND), y por búsquedas jerárquica.

(2) La búsqueda de Metadatos.. La búsqueda de metadatos es una mejora de la búsqueda sintáctica , ya que usa aspectos relevantes de un documento esquema.
Existen dos (2) tipos de búsqueda de metadatos - , sintáctica y semántica. En la búsqueda de metadatos sintáctica, los elementos sintácticos de un documento como el titulo, el encabezado y los links son considerados.
En la búsqueda de metadatos semánticos es usualmente desarrolloda a través de dominios especializados. El conocimiento especifico del dominio es usado para incrementar el recall y la precisión de la consulta.

(3) Query by Example (QBE). En QBE, la máquina de búsqueda presenta documentos muestra (ejemplo). La máquina de búsqueda, entonces recupera el documento relevante. El usuario puede expandir el documento y refinar la búsqueda por medio de una selección apropiada de los documentos relevantes retornados por la máquina de búsqueda. QBE es usado para búsquedas nop textuales, tal como algunas imágenes web.

(4) Búsqueda Navigacional. La búsqueda navigacional simula el proceso de búsqueda en el browser. Un robot o agente muestra una página ejemplo que contiene links. El robot evalua cada página encontrando la búsqueda más relevante.
La búsqueda de metadatos semánticas es las más acertada para la recuperación de documentos web. La búsqueda sintáctica es limitada, porque el contexto de la búsqueda no es considerada.
Sin embargo, QBE es útil en dominios especializados tales cómo consultas de imágenes pero es díficil desarrollar para una búsqueda genérica.

b Desarrollo y uso de ontologías en buscadores web

Las ontologías se han empleado en varias áreas, por ejemplo en Química, Derecho , Directorios telefónicos y catálogos de productos que son una importante prueba para capturar dominios semánticos.
Cualquier sistema basado en conocimiento es creado sobre cosas abstractas, vista simplificada del mundo que es llamada conceptualización. Una ontología es una especificación explícita de una conceptualización, que puede ser usada para soportar comunicación en diferentes partes con diferente conocimiento interno de esquema de organización. La mayoría de ontologías están constituidas en términos, sus definiciones y axiomas relacionandolos, con términos usualmente organizados en una taxonomía.
La investigación de las ontologías se han unido para estruturar las ontologías de forma taxonomica/jerárquica o por redes de términos. El más famoso de estos es WordNet, él organiza términos usando un modelo lexicográfico de la mente humana. Sin embargo WorNet es una ontología genérica de palabras.
Una pequeña investigación ha creado una flexible y adaptable estructura para organizar contenidos de ontologías.
Existen problemas conocidos de las ontologías asociadas a la estructura jerarquica o a las redes de términos que evitan su uso en los buscadores web.
Estos problemas son los siguientes:
1.Dificultad de obtener dominio de humanos expertos. Tradicionalmente las ontologías son desarrolladas normalmente por humanos expertos. Los usuarios de internet algunas veces consultan la web por un rango grande de tópicos.Es difícil y costoso obtener un dominio experto para la abundancia de tópicos que los usuarios buscan. Sin embargo, muchos aspectos de las ontologías requieren solo una especialización de un dominio limitado.
2.- Extensibilidad. Las ontologías basadas en taxonomías son dificiles de extender. Frecuentemente,la adición de nuevos términos necesita la reestructuración de la jerarquía. Para la búsqueda web las ontologías deben ser extendibles.
3. Dificultad de validación. La validación de las ontologías es usualmente realizada por expertos de ese dominio, lo que conlleva a que se presenten dificultades en sus apreciaciones.

3.4 LA ARQUITECTURA DEL MÉTODO SWQ

La arquitectura del método SWQ está basada en:
1. un motor SWQ y sus componentes como un parser para la consulta y un motor que determina el contexto ontológico;
2. el contexto ontológico con el dominio de la aplicación; y
3. filtros de búsqueda semánticos para mejorar la precisión.

Esta arquitectura es representada de la siguiente forma (fig. Nº 5):

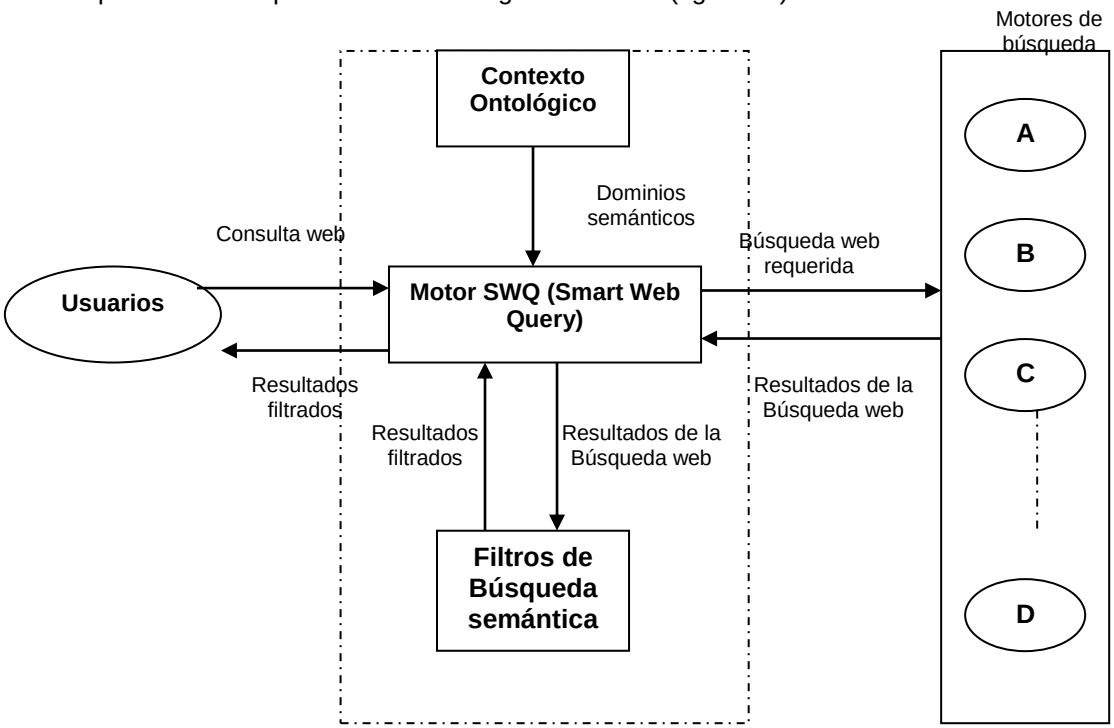


Figura Nº 5 La arquitectura SWQ (Smart Web Query)

Está arquitectura es descrita de la siguiente forma:

3.4.1 El motor SWQ (Smart Web Query)

El motor SWQ está compuesto por una serie de procesos, que permiten entender cual(es) son sus funciones:

3.4.1.1 Analizar la búsqueda web

El usuario digita una palabras claves con una serie de operadores booleanos (por ejmplo y, o) y algunos parámetros de búsqueda. El motor SWQ, entonces construye un árbol analítico de las búsquedas del usuario.

3.4.1.2 Determinar la ontología

El motor SWQ soporta múltiples contextos ontológicos, el usuario debe seleccionar la ontología más relevante para guíar la búsqueda.

Para facilitar la determinación de ontologías, el motor escoge las palabras claves del usuario en las ontologías y presenta las ontologías apropiadas rankeadas desde la ontología que posee la mayor parte de los términos escogidos hasta la ontología que tiene menores términos escogidos.

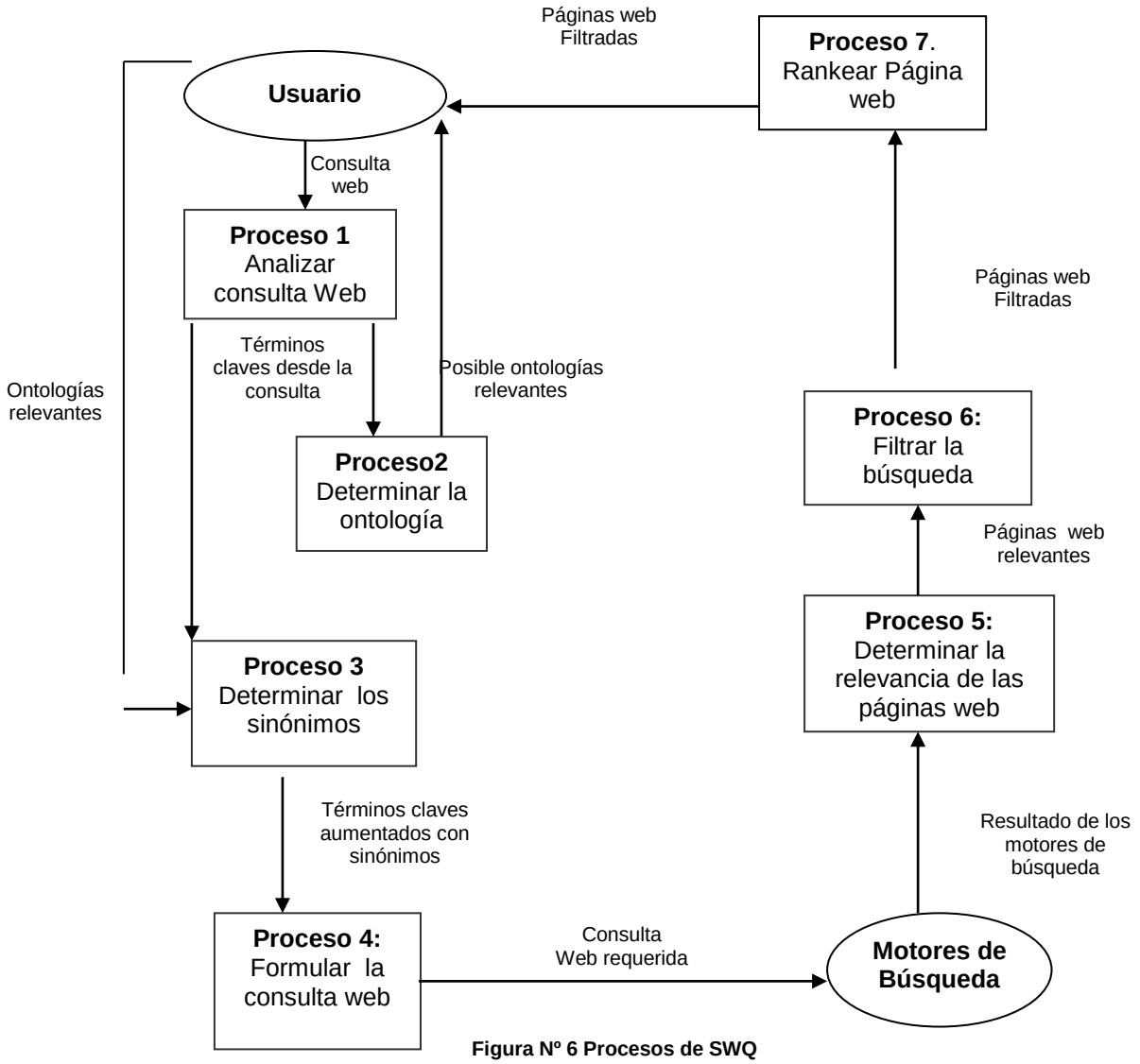


Figura Nº 6 Procesos de SWQ

3.4.1.3 Determinar los sinónimos

Se encuentran un conjunto de términos sinónimos a los términos de la ontología relevante seleccionada por el usuario. La finalidad de encontrar estos términos sinónimos es ampliar el rango de búsqueda.

3.4.1.4 Formular la consulta web

El motor SWQ primero selecciona el contexto ontológico para refinarlo y extender los términos claves identificados por el usuario. Con el conjunto de términos encontrado se hace una consulta Web en buscadores tradicionales pasándolos como argumentos y recuperando la lista de respuesta que ellos arrojan.

3.4.1.5 Determinar la relevancia de las páginas web

Muchos motores de búsqueda retornan el URL y un snippet (pequeño trozo de texto que pertenece a la página encontrada, en el cual se encuentra el conjunto de términos claves introducidos al principio por el usuario). Estos textos, snippets, son parseados por un buscador de los términos encontrados en las ontologías. El porcentaje de términos encontrados se

compara con en numero total de palabras en el texto y es usado para ordenar las páginas web por relevancia.

3.4.1.6 Filtrar la búsqueda

Para reducir la lista de páginas y poder devolver la más relevantes, se puede usar uno o varios filtros, entre ellos: el de legibilidad, estructura del documentos y sentido de las palabras (word sense).

3.4.1.7 Rankear páginas web

Después de pasar el filtro estás páginas web son rankeadas.

3.4.2 El Contexto Ontológico

El contexto ontológico define los términos básicos y sus relaciones. Esto comprende el vocabulario y la semántica de un dominio. El contexto ontológico apunta a capturar y organizar dominios semánticos en una estructura flexible. Estos pueden ser reusados y distribuidos por los motores de búsqueda (buscadores web).

De esta manera un contexto ontológico contiene tres (3) componentes:

1. Términos: Una serie de términos básicos que comprenden el vocabulario de un domini.
2. Relaciones: Una serie de relaciones entre estos términos.
3. Propiedades: Una serie de propiedades de la ontología (domino), términos y sus relaciones.

3.4.2.1 Estructura flexible del contexto ontológico.

El desarrollo ontológico está representado por las ontologías como una jerarquía de términos (por ejemplo una taxonomía). Con una estructura taxonómica, los términos son organizados primariamente con un oreden jerárquico.

Por ejemplo cuando se desea incluir nuevos términos en una ontología con una estructura jerárquica es necesario establecer el orden jerárquico.

En esta investigación no se adopta uan estructura taxonómica en términos organizados pertenecientes a un domino particular. Más bien, los términos pueden ser oraganizados en el contexto ontológico sin ningún orden jerárquico. Además, las realciones entre términos pueden ser adicionadas después en cualquier secuencia. Es aceptable en SWQ para los términos no tener relación con otros términos.

Esta estructura en términos organizados hace del contexto ontológico más adaptable y extensivo con un despliegue rápido para la máquina SWQ.

3.4.2.2 Organización del Contexto ontológico

Las ontologías están realcionadas por el número de términos que ellas tengan en común.

Las ontologías pueden ser sub-ontologías de otras ontologías (super-ontologías). Una ontología es una sub-ontología sí y solamente sí todos sus términos están contenidos en otra ontología.

Los dominios de las propiedades se la sub-ontología son siempre restricciones de los dominios de sus propiedades super-ontológicas.

3.4.2.3 Ontológico y propiedades de términos

Las ontologías y los términos pueden tener propiedades correspondientes llamadas ontológico y propiedades de términos respectivamente. Cada ontología tiene propiedades idénticas, aunque con diferentes valores. Por ejemplo, cada ontología tiene una propiedad de legibilidad. La legibilidad (readability) cuantifica la educación formal que un lector requiere para comprender un documento.

Similarmente, cada término ontológico tiene uan definición y un sentido de la palabra (word sense), que significa que el término puede ser un verbo, nombre , adjetivo,etc.

3.4.2.4 Relación entre término y sus propiedades

Los tèrminos pueden interactuar con otros términos para formar realciones entre términos. Hay dos tipos de relaciones entre los términos y sus propiedades que son: genéricos y específico.La realción del término genérico se encuentra en todos los términos relacionados. Por ejemplo algunos términos relacionados son simétricos y otros son asimétricos.

Está característica de la relación de términos y sus propiedades es llamada la propiedad Dirección.

La relación dirigida permite a la ontología SWQ establecer un orden parcial entre términos, teniendo en cuenta que el orden de los términos depende de la relación de términos y no de una jerarquía.

La relación de términos en la ontología SWQ puede ser transitiva, la cuál tiende a ser un estado explícito.
Una propiedad solamente puede ser relevante a únicamnete a algún término.
La relación – término *Sinónimo* tiene la propiedad de *Distancia Semántica* la cuál no es hallada en otra relación – término.
La Distancia Semántica es el grado de sinonímia de dos (2) términos, con valores entre cero (0) y siete (7). Un puntaje de siete (7) significa que existe una sinonímia muy fuerte, cuatro (4) indica sinonímia media y uno (1) en sinonímia muy débil. No es necesario especificar las propiedades de las relaciones entre términos, dejar la propiedad de la relación Distancai Semántica no especificada no impide la funcinalidad del motor SWQ. Además, las ontologías SWQ pueden ser ampliadas, incrementadas y refinadas incrementalmente con términos adicionales, relaciones y propiedades.
El estándar lexicográfica de las relaciones (sinónimos,antónimos,homónimos) , y otras relaciones pueden ser modeladas de esta manera.
La habilidad para modelar todas las relaciones y los términos idénticamente
hacen del SWQ muy escalable. La figura Nº muestra el modelo ontológico SWQ.

3.4.2.5 Esquema relacional y ontologías SWQ

Las ontologías SWQ y el esquema de base de datos relacional comparte características comúnes. Por ejemplo, ambas formas de representación no orden adatos explícitamente. Así, una administración de Base de datos relacional (DBMS) fue escogida para organizar y almacenar las ontologías SWQ.
La figura Nº 7 muestra el esquema relacional para organizar ontologías SWQ.

Ontología:	*Ontología* (Nombre Ontología)
	Propiedades Ontología (Nombre Ontología, Nombre Propiedad; Valor)
	Relación Ontología (Nombre Ontología,Nombre Super Ontología)
Término:	*Términos* (Nombre Ontología, Término, Definición).
Relaciones Término:	*Relaciones* (Nombre Ontología, Código relación, Tipo Relación).
	Propiedades Relaciones (Código relación, Nombre Propiedad Relación, Valor Propiedad relación).
	Relación Términos (Código relación, Término).
	Relación Origen (Código relación, Término).

Figura. Nº 7 El esquema relacional del motor SWQ.

La relación *Ontología* captura el nombre de la ontología. Cada una de las propiedades de las Ontologías (por ejemplo Puntaje de legibilidad) y sus valores son almacenados en una tupla individual, en la relación *Propiedades Ontología.* Las ontologías son tratadas como subclases o superclases, sí son presentadas como la relación *Relación Ontología.* Los términos (por ejemplo Ciencia, Computación) y las propiedades de los términos tal como el Sentido de la Palabra (word sense), (es decir, si el término es un nombre o un verbo) usado en la ontología son almacenados en la relación *Términos.* Las relaciones entre términos son almacenadas en la relación *Relaciones.* La relación *Relación Términos* almacena la intersección de datos de estas dos relaciones. Para direccionar las relaciones la *Relación Origen* identifica los términos que direccionan otros términos. Finalmente, la relación *Propiedades Relaciones* almacena las propiedades de las relaciones entre términos.

3.4.2.6 Beneficios y características de una estructura flexible

El principal beneficio de una estrutura flexible es que es fácil desplegar una ontología sobre un sujeto de dominio. Como mínimo una ontología puede ser empleada introduciendo una serie de términos ontológicos sin ninguna relación entre ellos.
Para muchos dominios, los términos ontologicos pueden ser descubiertos, identificando un diccionario de dominio. Humanos expertos en dominios son solamente requeridos para definir las propiedades y las relaciones, las cuales incrementan pero no son requeridas por un buscador SWQ.
En muchas ocasiones, la ontología puede estar definida semi-automáticamente.

Por ejemplo, un desarrollo ontológico puede identificar una colección electrónica de palabras (diccionario en línea) que encapsula una ontología.
Según la figura Nº , los rectángulos representan términos ontológicos. Los ovalos representan la relación entre términos. Las palabras subrayadas representan propiedades (para las relaciones y las relaciones entre términos).
Segmentos diferentes de la ontología pueden ser representados por personas diferentes, porque los términos están organizados usando una estructura flexible, es decir las relaciones definidas por un usuario no comprometen la relación definida por otro usuario. El borrado o modificación de una u otra relación no afecta la otra,
Aunque las ontologías SWQ pueden ser extendidas por múltiples individuos, algunos conflictos pueden ocurrir. En muchas ontologías ocurren estos conflictos cuando son guiadas o inferidas a un dominio que no es. Sin embargo las ontologías SWQ son usadas para identificar páginas web que pertenecen a un dominio en particular.

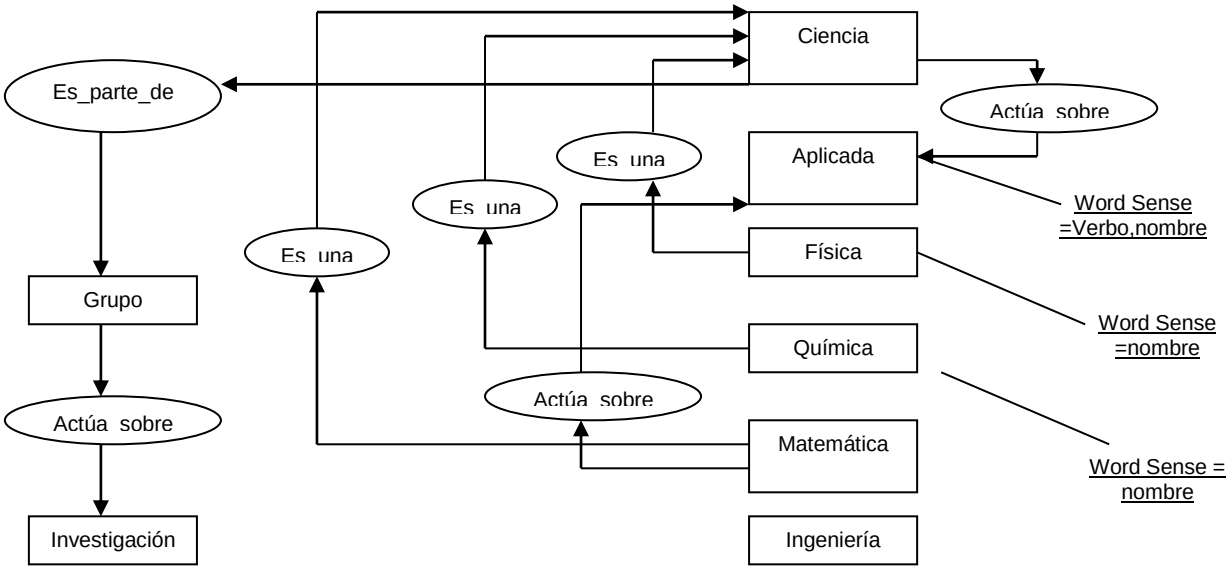


Figura Nº 8 Ontología Ciencias Aplicadas en SWQ

Este es un ejemplo de ontología, tal y cómo se representa en el método SWQ, y la cúal hace parte del Laboratorio de Cómputo Especializado. En ella se puede observar los términos más relevantes (keywords) que la conforman.

Éstos términos están con sus propiedades, y sus relaciones, las cuáles poseen también propiedades.

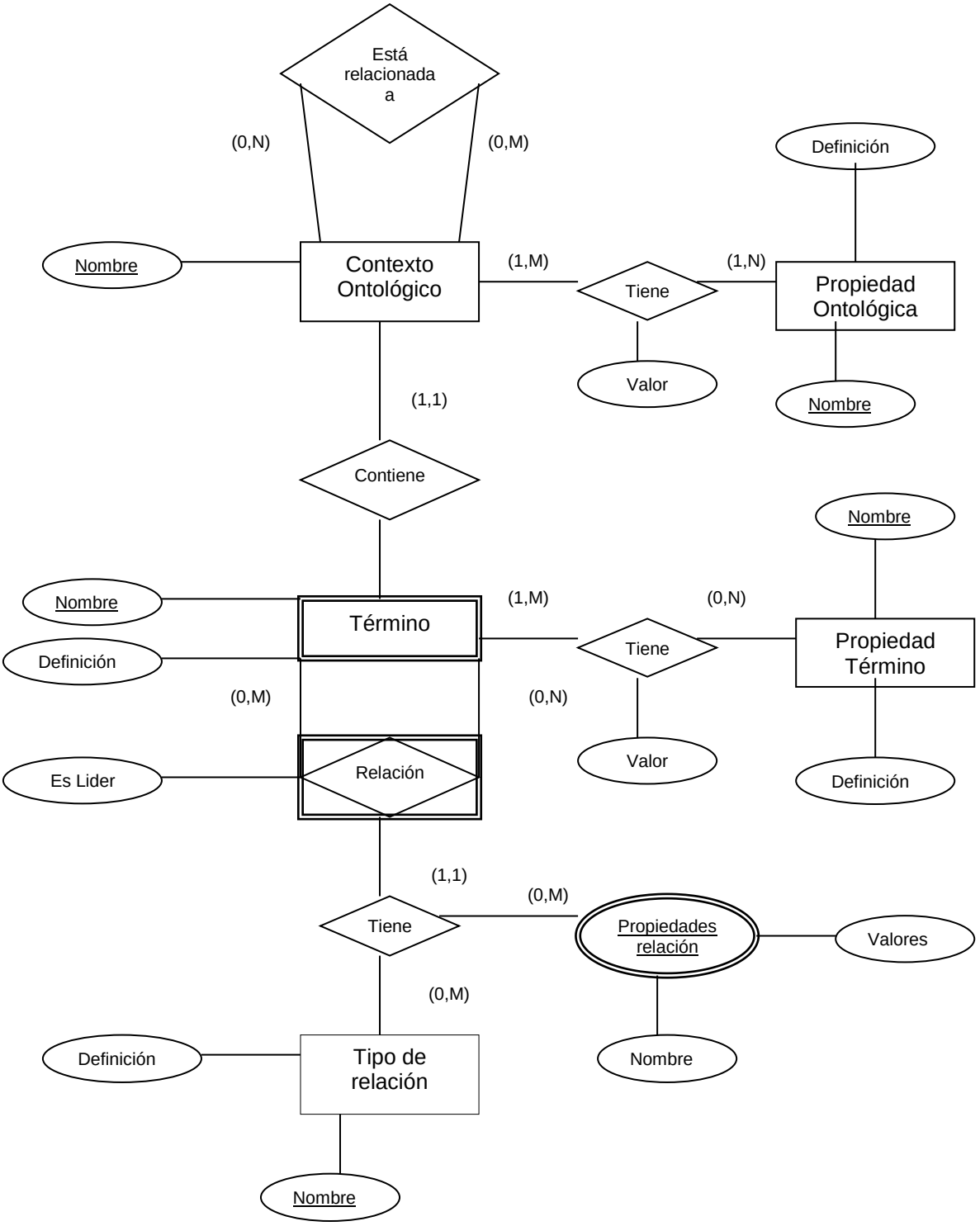


Figura Nº 9 El ESQUEMA RELACIONAL SWQ

3.4.2.7 Comparación entre el contexto ontológico y las ontologías genéricas para la búsqueda web

Aunque se han hechos algunos intentos para desarrollar ontologías universales, el contexto ontológico es más efectivo para la recuperación semántica de los datos web.

El contexto ontológico puede incrementar búsquedas web en direcciones, mientras las ontologías genéricas no lo pueden hacer.

Primero, es más fácil identificar frases y oraciones, por ejemplo, información lexicográfica que son correctas para el dominio.

Hay que tener en cuenta que existen términos que pueden ser adjetivos, nombres o verbos en un dominio específico, mientras en otros no.
Segundo, esto permite la automatización de ciertos procesos en la búsqueda web. Por ejmplo, los sinónimos pueden ser automáticamente determinados sí el contexto ontologico es conocido.
Finalmente, en una ontología genérica, el conocimiento del dominio y la información lexicográfica puede no ser simultáneamente representada y modelada, mientras que en el contexto ontológico puede ser como una parte de dominios semánticos (ejemplo: propiedades de términos.
Esto hade de el contexto ontológico una herramienta poderosa para buscar dominios especializados.

3.4.3 Filtros de búsqueda semántica

Para el modelo SWQ una serie de filtros semánticos son integrados para incrementar la precisión de la búsqueda web. Los filtros de búsqueda semántica son módulos de búsqueda especializada que emplea una o unas pocas relaciones o propiedades en la ontología para evaluar la relevancia de la página web. Actualmente, se han integrado tres (3) filtros de búsqueda: (1) el filtro de legibilidad, (2) el filtro documento estructurado y (3) el filtro sentido de las palabras (word sense).

3.4.3.1 El filtro de Legibilidad

El filtro de legibilidad identifica la relevancia de las páginas web usando el puntaje de Flesch – Kinkaid. Este puntaje es hallado (ecuación Nº 1) , en dónde *W* es el número de sílabas por cada cien (100) palabras , y *S* es el número de palabras en una oración. El filtro de legibilidad de Flesch – Kinkaid es uno de los más populares y es incorporado en dominios tal como el derecho y procesador de palabras.

$$R = 206.835 – 0.846W – 1.015S$$

Las páginas web encontradas por las motores de búsqueda son evaluadas por su nivel de calidad.Este nivel de calidad es comparado con la tabla de niveles de calidad estandarizada, las cuál contiene los respectivos puntajes.
Cuando las páginas web no caen dentro de algún nivel de calidad es rechazada. Ver Tabla N º 1 Nivel de calidad Flesch Kinkaid

Puntaje mínimo	Puntaje máximo	Nivel de calida Flesch
Menor de 30		Post- Grado
30	49	Universidad
50	59	Media Vocacional
60	69	Grado Octavo y Noveno
70	79	Grado Séptimo
80	89	Grado Sexto
Por encima de 90		Por debajo Quinto Grado

Tabla N º 1 Nivel de calidad Flesch Kinkaid

4. DESARROLLO DEL PROTOTIPO

La metodología usada para desarrolar el prototipo de buscador del Laboratorio de Cómputo Especializado está basada en el Proceso Unificado de Desarrollo de Software (RUP). Este RUP como su nombre lo dice es una metodología a seguir en el Proceso de Desarrollo de Software y al mismo tiempo ayuda a obtener diseños sólidos usando UML, simbología la mayoría desarrolladores y clientes aceptan como estándar.

El diseño del prototipo de basa en la investigación realizada por Roger Chiang, Cecil Eng Huang Chua y Veda Storey. Ellos encontraron una forma de implementar un búscador web semántico de una forma práctica, usando como base un motor SWQ (Smart Web Query) o Motor de búsqueda inteligente.
Éste motor posee una serie de procesos que están basados en el modelo de bases de datos relacional, junto con una serie de algoritmos o filtros que refinen la búsqueda.

De esta manera, el desarrollo de este prototipo se adecuó al lenguaje de modelado unificado (UML), para que sea entendido por los desarrolladores de Sotware.

4.1 ETAPA DE ANÁLISIS DE REQUERIMIENTOS

A continuación se dispone a recolectar información del LCE.
El Laboratorio de Cómputo Especializado (LCE) de la Universidad Autónoma de Bucaramanga, UNAB, nació en 1996 para responder a las necesidades surgidas del convenio de cooperación académica firmado con el ITESM (Instituto Tecnológico de Estudios Superiores de Monterrey), de Méjico. El acuerdo con esa institución educativa permitió a nuestra universidad, entre otras cosas, crear las maestrías en Administración y en Ciencias de la Computación y, tras ellas, generar condiciones para que los estudiantes de dichas maestría realizaran sus trabajos de tesis.

Hoy, el LCE es una dependencia de la UNAB con autonomía administrativa y presupuestal. Los objetivos son:

- Promover el desarrollo tecnológico de la Universidad y de su entorno.
- Promover y realizar investigación científica y tecnológica en ciencias básicas y en ingeniería.
- Brindar soporte técnico y académico a la maestría en Ciencias Computacionales y a otras maestrías afines a la ingeniería que se lleguen a crear en la UNAB.

La misión del LCE es: Realizar proyectos de investigación y desarrollo tecnológico en las áreas de ingeniería y afines de la UNAB para así fortalecer su infraestructura académica, científica y tecnológica, además de contribuir a la solución de los problemas del entorno. En el desarrollo de ésta labor se valoriza la innovación, el trabajo participativo y la ética.

El LCE cuenta con un director y con un recurso humano calificado para su correcto desarrollo, cada uno de ellos pertenece a algún grupo de investigación

El LCE cuenta con una infraestructura tecnológica para sus labores de I+D, con una red local constituida por 20 computadores personales multimedia y 1 servidor. Adicionalmente se cuenta con una buena cantidad de periféricos y software especializado.

Este Centro está conformado por los grupos de investigación:

- Pensamiento Sistémico
- Tecnologías de Información
- Ciencias Aplicadas

Los grupos de investigación cuentan con un director e investigadores asociados al grupo. El grupo de investigación cuenta con proyectos en desarrollo y ya proyectos desarrollados o

concluidos; estos proyectos de investigación pertenecen a alguna de las líneas de investigación que conforman el grupo de investigación.

Cada proyecto, tanto en desarrollo como ya desarrollado, cuenta con una ficha técnica en donde se resaltan los siguientes puntos:

- Nombre
- Descripción
- Fecha inicio
- Fecha finalización
- Tipo de Investigación
- Investigadores participantes
- Cooperación a través de proyectos
- Línea de investigación principal del proyecto

A continuación se detallan los grupos de investigación:

- **Ciencias Aplicadas**: El grupo de Ciencias Aplicadas surge de la fusión del grupo de Matemática Aplicada, con el objetivo de apoyar también los proyectos e investigaciones en Física y Química, áreas en las cuales la división de ciencias naturales e ingeniería posee un recurso humano valioso.

- **Pensamiento Sistémico**: El grupo de investigación en Pensamiento Sistémico y Educación nació de la confluencia de dos líneas de interés que se gestaron en la Facultad de Ingeniería de Sistemas de la UNAB a mediados de la década del 1990, la línea de informática Educativa y la línea de Pensamiento Sistémico. Es así como a partir del año 1996 profesores de ambas líneas de investigación, iniciamos dirigiendo algunas tesis de pregrado y esto nos permitió encontrarnos con temas de interés común a ambas líneas
- **Tecnologías de Información:** El grupo de Tecnologías de Información surge con la creación del Laboratorio de Cómputo Especializado (LCE). En el año de 1996, la Universidad Autónoma de Bucaramanga (UNAB) firma un convenio de cooperación académica con el Instituto Tecnológico de Estudios Superiores de Monterrey de México (ITESM), convenio que trajo entre otras consecuencias la creación de la Maestría en Ciencias Computacionales. El grupo surgió para promover y realizar investigación científica en el área de las tecnologías de la Información y brindar soporte técnico y académico a la maestría en Ciencias Computacionales.

El LCE cuenta con un sitio web: http://serverlab.unab.edu.co donde se encuentra la información correspondiente a sus actividades.

4.2 ETAPA DE DISEÑO

Está información es analizada y presentada en un diagrama de casos de uso y el diagrama de secuencia. Para esto se revisará algunos conceptos:

4.2.1 Diagrama de Caso de Uso: Un diagrama de Casos de Uso muestra las distintas operaciones que se esperan de una aplicación o sistema y cómo se relaciona con su entorno (usuarios u otras aplicaciones).

4.2.2.1 Caso de uso

Se representa en el diagrama por una elipse, denota un requerimiento solucionado por el sistema. Cada caso de uso es una operación completa desarrollada por los actores y por el sistema en un diálogo. El conjunto de casos de uso representa la totalidad de operaciones desarrolladas por el sistema.

4.2.2.2 Actor

Es un usuario del sistema, que necesita o usa algunos de los casos de uso. Sé representa mediante una figura, acompañado de un nombre significativo, si es necesario.

4.2.2.3 Relaciones en un diagrama de casos de uso

Entre los elementos de un diagrama de Casos de uso se pueden presentar tres tipos de relaciones, representadas por líneas dirigidas entre ellos (del elemento dependiente al independiente).

La figura Nº muestra el diagrama de casos de uso del Prototipo de buscador web semántico usando el método SWQ.

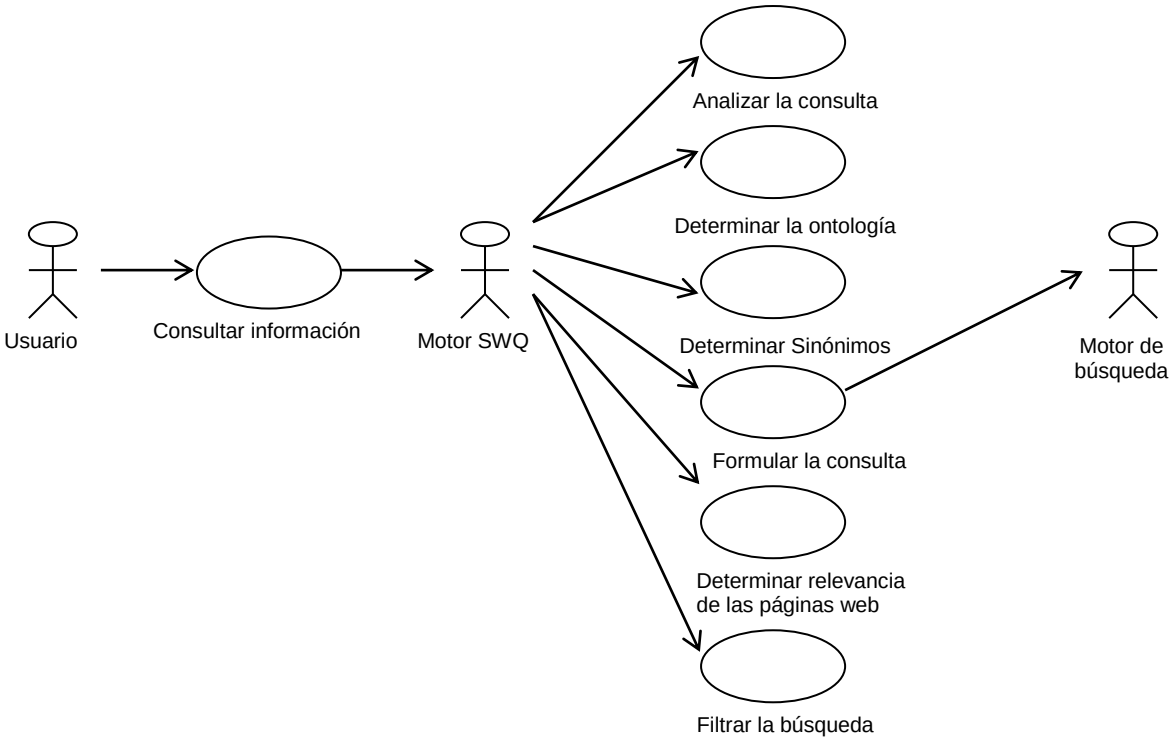


Figura Nº 10 Diagrama de casos de uso

4.2.2.4 Plantilla de casos de uso

En esta plantilla se detallan las actividades que conllevan los casos de uso. Se muestra la plantilla del caso Consultar Información.

Caso de Uso Consultar Información

1. Consultar Información

1.1 Breve Descripción

Este Caso de Uso describe como el usuario para consultar información, necesita digitar la información que necesita por medio de palabras claves (es decir, palbras que representen para el usuario relevancia a la hora de buscar ésta información), las cuáles quedan en motor SWQ, para después ser analizadas por él.

2. Flujos

El Caso de Uso Validar usuario inicia cuando el usuario ingresa las palabras claves (keywords).

2.1 Flujo Básico

1. El Sistema analiza éstas palabras claves en el contexto ontológico.

2.2 Flujo Alternativo

2.2.1 No se encuentra la palabra clave en el contexto ontológico.

Si el usuario escribe una palabra clave que no éste relacionada con el dominio , entonces el sistema le regresará un mensaje....ésta palabra no se encuentra en el contexto ontológico.

3. Requisitos Especiales

Los requisitos para este caso de uso es conocer el dominio de la aplicación.

4. Condiciones Previa

Las condiciones para este caso de uso, es estar relacionado con el Laboratorio de Cómputo Especializado, para entender que dominios se encuentran allí.

5. Condiciones Posteriores

Al finalizar el caso de uso Consultar Información ,el sistema muestra las ontologías relacionadas con la palabra(s) clave(s).

Cuadro Nº 2 Caso de uso "Consultar Información"

4.2.2 Diagrama de Secuencia

Muestran las interacciones entre un conjunto de objetos, ordenadas según el tiempo en que tienen lugar. En los diagramas de este tipo intervienen objetos, que tienen un significado parecido al de los objetos representados en los diagramas de colaboración, es decir son instancias concretas de una clase que participa en la interacción. El objeto puede existir sólo durante la ejecución de la interacción, se puede crear o puede ser destruido durante la ejecución de la interacción. Un diagrama de secuencia representa una forma de indicar el período durante el que un objeto está desarrollando una acción directamente o a través de un procedimiento.

Los diagramas de secuencia permiten indicar cuál es el momento en el que se envía o se completa un mensaje mediante el *tiempo de transición*, que se especifica en el diagrama.

Los diagramas de secuencia se utilizan para describir la vista dinámica del sistema.

Para el método SWQ, se muestra el diagrama de secuencia (Figura Nº 11) para el caso de uso *Consultar Información*. Los demás diagramas se encuentran en el Anexo B.

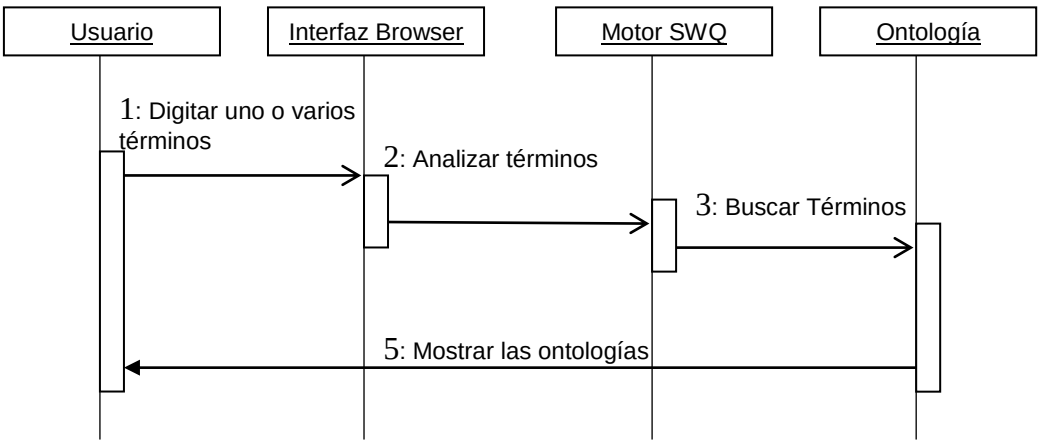
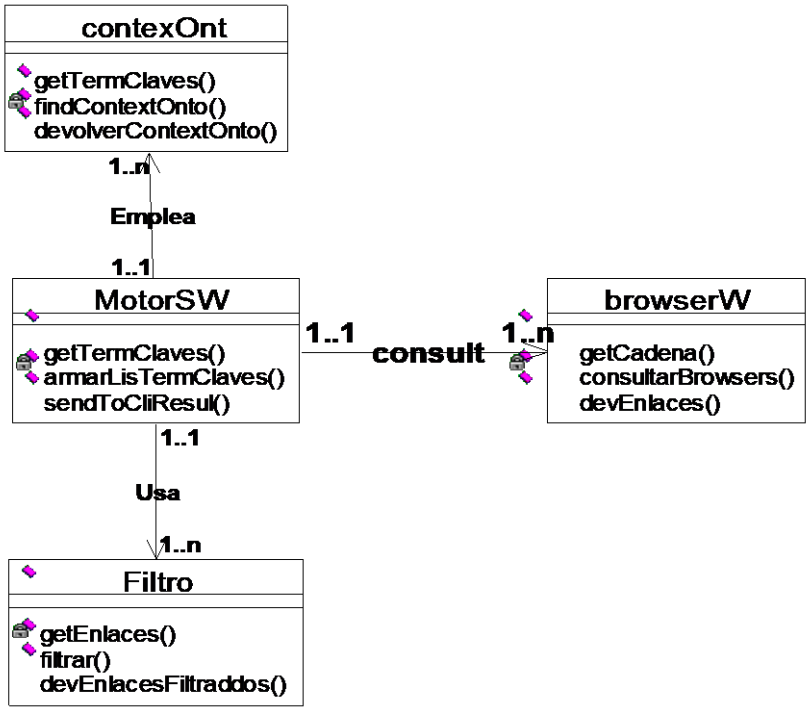


Figura Nº 11 Diagrama de Secuencia Consultar Información

4.2.3 Diagrama de clases

Muestra el conjunto de clases y objetos importantes que hacen parte de un sistema, junto con las relaciones existentes entre estas clases y objetos. Muestra de una manera estática la estructura de información del sistema y la visibilidad que tiene cada una de las clases, dada por sus relaciones con las demás en el modelo.
En la figura Nº 12 se muestra el diagrama de clases del sistema.

Figura Nº 12 Diagrama de clases



4.3 ETAPA DE IMPLEMENTACIÓN

En esta etapa se explica se explica brevemente en que ambiente, herramientas y lenguajes se utilizarón para la implementación del prototipo.

4.3.1. Ambiente de desarrollo.

Para desarrollar el prototipo se tenían dos ambientes principales de desarrollo. El primero era el ambiente cliente que consistía en desarrollar el prototipo de un buscador web, que usara web semántica con la arquitectura SWQ, en un computador cliente y desde allí el usuario podría hacer sus consultas Web. La gran desventaja que ofrece este ambiente es que es monousuario, se debe instalar en cada máquina que se quiera usar y no tiene mucha semejanza a un buscador tradicional.

El segundo ambiente con que se contaba era con un ambiente orientado a la web. Esto es un programa que corriera en una maquina que funcionara como servidor web y que los interesados la consultaran desde su computador cliente a través de un navegador de internet que soportara graficos (como Internet Explorer, Netscape, Mozilla, etc.). Como desventaja de este tipo de ambientes, es la dificultad que trae inherente el desarrollo de una aplicación en que corra en un servidor web.

Por motivo de que la aplicación que se pretende desarrollar está orientada a la web y se desea que se parezca (en su interfaz) a los buscadores web tradicionales se eligió el segundo ambiente de desarrollo.

4.3.2 Lenguaje de desarrollo.

En la actualidad existe una gran variedad de lenguajes y de opciones a la hora de hacer una aplicación web que corra en un servidor web. Entre ellas se tienen los CGI's, ISAPI's, Lenguajes que corren en el servidor como ASP o ASP.net, JSP, PHP. Cada uno de estos lenguajes y tecnologías tiene ventajas y desventajas.
La gran desventaja de los CGI's es que cada vez que un usuario solicita un servicio se monta el CGI en la memoria, consumiendo recursos, es decir, si 100 usuarios activan el CGI, habra 100 CGI ejecutando en la memoria. Esta desventaja es superada por los ISAPI, ya que básicamente son como librerías dinámicas en memoria, se cargan una vez y son usados por muchos; la desventaja es que no todos los ISAPI's corren en diferentes browser.
Nos quedan las tecnologías de servlet. Las dos más grandes tecnologías de servlet son ASP y JSP, ambas son excelentes. ASP es una excelente tecnología y la gran diferencia con JSP es el precio. Además JSP es basado en Java, el cual es un lenguaje de gran aceptación mundial, que corren en todas las plataformas (teóricamente por lo menos) el mismo algoritmo, y que no tiene precio de uso.

A continuación explicaremos muy brevemente Java y JSP.

4.3.2.1. Java. Java es un lenguaje moderno, publicado por Sun Microsystems en la ultima mitad de 1995, pero a pesar de ser un lenguaje muy joven, ha tenido gran aceptación por la comunidad de desarrolladores debido a diversas razones:

- La sencillez y elegancia de cómo se escriben los programas en Java. A ello se une que es un lenguaje orientado a objetos que evita muchas preocupaciones a los programadores. En el proceso de compilación se realizan multitud de comprobaciones que permiten eliminar muchos errores posteriores.
- Java tiene una gran biblioteca de librerías ya definidas que proporciona una gran ayuda al programador ya que no tiene que volverlas a hacer. Y si por alguna razón el programador no encuentra una biblioteca que se ajuste a lo que está haciendo, muy posiblemente la encontrará en la comunidad ya que frecuentemente se están escribiendo multitud de líneas de código Java para hacer una multitud de cosas y lo más maravillosos es que la gran mayoría de esas bibliotecas son gratis.
- Java es multiplataforma, esto es, el mismo programa Java puede correr en multitud de plataformas y sistemas operativos sin hacerle ningún cambio. Esta cualidad es muy conocida y se cumple en todas las plataformas que se conocen (Windows, Linux, Apple).

- Java proporciona mecanismos para programar en internet. Con Java se puede hacer desde una simple calculadora que sume y reste y que corra en un navegador a internet, hasta simuladores, editores de texto, buscadores web, etc.

La evolución de Java ha sido muy rápida. Desde que se hizo público el lenjuage y un primer entorno de desarrollo, el JDK (Java Development Kit), hasta el momento actual, la plataforma Java ha crecido constantemente y a un ritmo cada vez mayor.

Java no es solo un lenguaje. Es una plataforma de desarrollo de programas que consta de:

- Un lenguaje de programación: el lenguaje Java, del mismo nombre que la plataforma.
- Un conjunto de bibliotecas estándar que se incluyen con la plataforma y que deben existir en cualquier entorno con Java. También se denomina Java Core. Estas bibliotecas comprenden: strings, procesos, entrada y salida, propiedades del sistema. Fecha y hora, etc. Applets, API de red, Internacionalización, Seguridad, Componentes, Serialización, JDBC, etc.
- Un conjunto de herramientas para el desarrollo de programas. Entre ellas cabe citar el compilador de Java a código bytes, el generador de documentación, el depurador de programas en Java, etc.
- Un entorno de ejecución cuyo principal componente es una máquina virtual para poder ejecutar los programas en código de bytes.

4.3.2.2. JSP. Primero que todo JSP significa Java Server Page (Servidor de Páginas Java). JSP en una solución al problema de las páginas dinámicas. El primer intento de generar contenido dinámico fue la interfaz de puerta de enlace común (CGI, Common Gateway Interface), que dependia de scripts de Perl o programas de C que se ejecutaban en el servidor. Esta solución encontró problemas como el uso de memoria, la dificultad para el diseñador web que no era programador, el manejo de sesiones y que la lógica de trabajo se encontraba en el mismo código que era responsable de la capa de presentación.

La solución a los CGI´s son los servidores de páginas web. El más conocido junto con JSP es ASP (Active Server Pages) de Microsoft. Una de las diferencias fundamentales entre JSP y ASP es el precio ya que la plataforma para correr ASP tiene precio y JSP no lo tiene. Otra diferencia fundamental es que JSP por ser Java puede correr en cualquier plataforma (Solaris, Linux, Windows, Mac, AIX, HP-UX y otras variantes UNIX) en cambio las ASP solo corren en plataformas de Microsoft (Windows); además JSP puede correr sobre cualquier Servidor Web (los más populares son Apache, Netscape, IIS, Tomcat) y las ASP solo corren en Microsoft IIS o Personal Web Server.

Entre los beneficios que ofrece usar JSP están: la capacidad de separar la lógica del programa y su aspecto (beneficio más importante); otro beneficio muy importante esta de que se escribe una vez y se ejecuta en cualquier lugar, esto debido a que cuando se compila lo que se produce es un código de bytes Java que son corridos por una plataforma de ejecución Java y existe una plataforma de ejecución Java en todos los sistemas operativos; JSP es independiente de la plataforma. Otro beneficio es que JSP facilita la inserción de componentes reutilizables. Componentes JavaBeans y Enterprise JavaBeans, que realizan tareas especializadas, pueden desarrollarse una vez e incorporarse en cualquier página JSP.

4.3.3. Arquitectura del prototipo.

La Web Semántica aún no está lo suficientemente extendida, sin embargo, se han propuesto varias arquitecturas para facilitar la recuperación de información en la Web Semántica**.**
En el desarrollo del prototipo de un buscador web semántico se escogió la arquitectura SWQ (A Smart Web Query method for semantic retrieval of web data). Esta arquitectura nos permite construir un buscador web usando ontologías, términos de un dominios y relaciones entre estos términos y ontologías.
Uno de los componentes principales del método SWQ es la consulta a buscadores web tradicionales, así que se procedió a hacer varias pruebas en buscadores de extenso uso como Google, Altavista y Yahoo, pero se encontró que ninguno tenia la página del Laboratorio de Computo Especializado (LCE) indexada. Se procedió a consultar como se podía hacer para

que estos buscadores web indexaran la página del LCE y luego de leer las instrucciones, se anexo la ruta web del sitio web del LCE. Después de un largo tiempo prudencial (donde los robots de los buscadores web pudieran indexar el sitio del LCE) se volvió a consultar, de forma tradicional (esto es ir al sitio del buscador web e introducir términos claves indicando que buscara en el sitio del LCE), el sitio web del LCE y se encontró que no lo habían indexado.

Se repitió el procedimiento de agregar el sitio del LCE a los buscadores web y se espero seis (6) meses para que estos buscadores indexaran. Para sorpresa se encontró que no lo hacían. Dado las circunstancias se procedió a construir un índice del sitio web del LCE para que simule el proceso de consultar los buscadores web tradicionales.
Para la construcción del índice del sitio web del LCE se tuvieron que desarrollar varios programas adicionales para que extrayeran la información; otro programa para que analizara y depurara la información extraída; y otro programa para que indexara la información extraída y depurada.

4.3.4. Herramientas utilizadas.
Las herramientas que se usaron para la implementación fueron:

- Tomcat 5 como servidor web de páginas JSP.
- Java de la Sun versión del JDK 1.4.2. Con esta versión de la plataforma Java se compilo los diferentes componentes del prototipo
- MySQL como motor SQL de bases de datos para que sirviera como contenec prototipo.
- Otras herramientas que se usaron fueron Dreamweaver como editor de páginas JSP, Jcreator Lite versión 3.1 como entorno de edición y complicación de Java y MySQL Front como herramienta para crear bases de datos, tablas y relación en MySQL.

RECALL AND PRECISION.

Después de terminar una búsqueda, el usuario se le viene a la mente la siguiente pregunta: ¿Será que estoy encontrando el material pertinente o se están pasando por alto el buscador, material que es importante para mí?
Además el usuario espera no encontrar mucha “basura” (información no relevante para el usuario) cuando realiza una búsqueda, cosa que pasa en la realidad, ya que al realizar una búsqueda siempre se encuentra basura.
Teniendo en cuenta estos dos parametros se puede verificar que tan efectiva fue una búsqueda.
El Recall and precision son las dos medidas básicas para evaluar las estrategias de búsqueda.

$$recall = |R^*| / |R|$$

$$precision = |R^*| / |A|$$

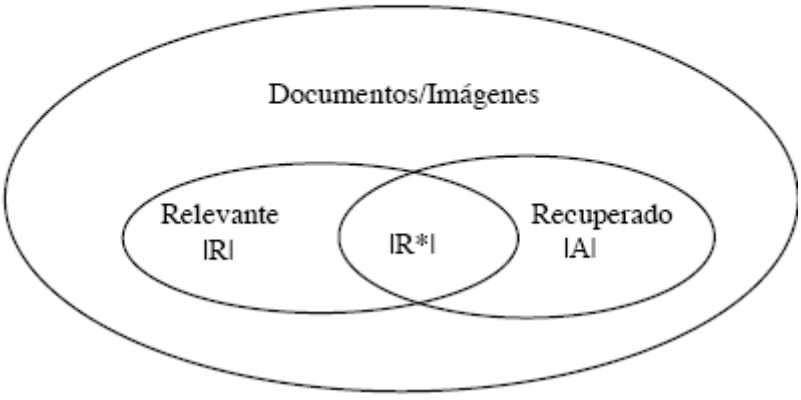


Figura 13. Recall and Precision

5.1 RECALL

Es la proporción del número de documentos relevantes recuperados, sobre el total de documentos relevantes de la base de datos, se expresa normalmente como un porcentaje.[15]

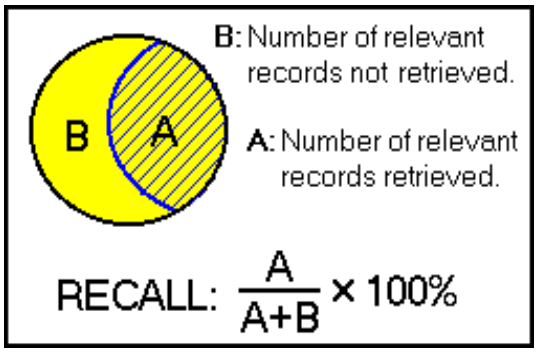


Figura 14. Recall

5.2 PRECISION

Es la proporción del número de documentos relevantes recuperados sobre el número total de documentos relevantes y no relevantes recuperados. Es usualmente representado en porcentaje.

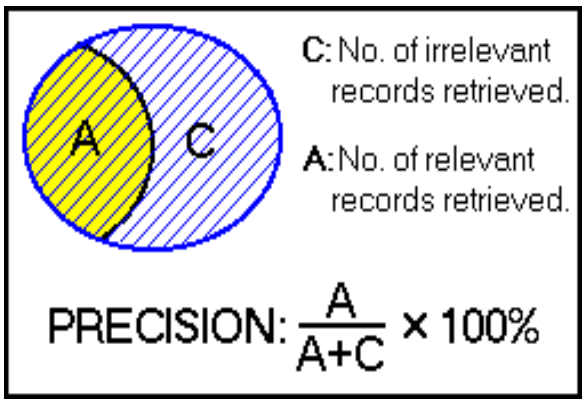


Figura 15. Precision.

CONCLUSIONES

La web fue diseñada como espacio de información, con el objetivo de que debería ser útil no solamente para la comunicación entre humanos, sino también para que las máquinas pudieran participar y ayudar.

La Web Semántica es una visión: la idea del contar en la web con información definida y ligada de tal manera que pueda ser utilizada por las máquinas - no sólo para finalidades de exposición, sino para utilizarla en diversos usos. La Web solamente puede alcanzar su potencial máximo si se convierte en un lugar en donde los datos se puedan compartir y procesar por las herramientas automatizadas tanto como por las personas.

La web semántica es un espacio conceptual de información en el cual los recursos identificados por URIs (Identificador Universal de Recursos) se pueden procesar por máquinas. Funciona mediante los principios de 'comprensión parcial' y de 'inferencia' (pudiendo deducir nuevos conocimientos a partir de datos ya entendidos), y por lo tanto, de la evolución y la transformación.

Teniendo como base el concepto descrito por Gruber sobre ontología que es “una especificación explícita de una conceptualización”. Esta definición menciona que una ontología esta especificando una forma de entender las cosas, está proporcionando una forma de ver el conocimiento, de representarlo, de especificarlo, en definitiva, una forma de entendimiento.
Este concepto conlleva a pensar que las ontologías son la esencia de la web semántica.

Las ontologías definen conceptos y relaciones de forma compartida, esta conceptualización debe ser legible, formal y utilizable por los computadores.

El método de Recall and Precision mide que tan relevante es la información recuperada, y por tal motivo verificar las ventajas que existe entre el buscador web convencional y el buscador web semántico.

El buscador web convencional se basa en la sintaxis de las palabras, mientras que el buscador web semantico en la semántica de ellas.

REFERENCIAS BIBLIOGRÁFICAS

[1] Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolgang Wahlster. Spinning the Semantic Web. Introduction. Page 1

[2] Tim Berners Lee . Spinning the Semantic Web Foreword.

[3] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220, 1993.

[4] V. Richard Benjamins and Asuncion Gomez Perez: Knowledge-System Technology: Ontologies and Problem-Solving Methods

[5] R. Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering, principles and methods. *Data and Knowledge Engineering*, 25(1-2):161–197, 1998.

[6] W. N. Borst. *Construction of Engineering Ontologies*. PhD thesis, University of Twente, Enschede, 1997.

[7] John F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, CA 2000. Disponible en http://www.jfsowa.com/ontology/index.htm

[8] R. Mizoguchi, J. Vanwelkenhuysen, and M. Ikeda. Task ontology for reuse of problem solving knowledge. In N. J. I. Mars, editor, Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing., pages 46–57. IOS Press, Amsterdam, NL, 1995.

[9] A. Preece, A.Flett, D Sleeman, D Curry, N Meany and P Perry. Better Knowledge Management through Knowledge Engineering. In *IEEE Intelligent Systems* January/February 2001, pp 36-42.

[10] E. Desmontils, C. Jacquin. Indexing a Web Site with a Terminology Oriented Ontology. *International Semantic Web Working Symposium (SWWS'2001), Stanford, CA, USA, july 30 - august 1, 2001, pp. 549-565.* [DSWKB00] Duineveld, A., Stoter, R., Weiden, M., Kenepa, B., and Benjamins, V. R., *Wondertools? A comparative study of ontological engineering tools*. International Journal of Human Computer Studies, 52(5), pp. 1111-1133, 2000

[11] Michael Gruninger, Mark S. Fox, Methodology for the Design and Evaluation of Ontologies, 1995

[12] M. Uschold. M. King. Towards a Methodology for Building Ontologies. *Workshop on Basic Ontological Issues in Knowledge Sharing.* 1995

[13] Rojas, M.D. Ontologías de iones monoatómicos en variables físicos del medio ambiente. *Final-Year Proyect. Facultad de Informática de la Universidad Politécnica de Madrid, 199º*

[14] A. Newell. The knowledge level *Artificial Intelligence*, 18(1), 87-127

[15] Creighton University Health Sciences Library and Learning Resources Center. Author: Richard Jizba, © 2000 Creighton University