ANTEPROYECTO DE GRADO

DESARROLLO DE UN MÓDULO ONTOLÓGICO PARA LAS CONSULTAS AVANZADAS DE INFORMACIÓN DE LA BIBLIOTECA DIGITAL DE LA ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN DE LA UNIVERSIDAD DEL VALLE

Presentado Por:

DIANA PAOLA HURTADO BUSTAMANTE
Código 9709557

Directora:

MARTHA MILLÁN Ph.D

UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
2006

TABLA DE CONTENIDO

INFORMACIÓN GENERAL 3

RESUMEN 4

INTRODUCCIÓN 5

OBJETIVOS 6
Objetivo general 6
Objetivos específicos 6

FORMULACIÓN DEL PROBLEMA 7
Definición del problema 7
Justificación 8
Estado del arte 9

METODOLOGÍA 10

MARCO TEÓRICO 12
La Web Semántica 12
XML 13
RDF 14
Ontologías 17
Lenguajes ontológicos 19
Herramientas para ontologías 20
Metodologías para la construcción de ontologías 23
Agente de software inteligente 24
Proceso de desarrollo de software 25

ASPECTOS ADMINISTRATIVOS 27
Plan de actividades y contenido 27
Cronograma de actividades 29
Parte 1 29
Parte 2 30

PRESUPUESTO 31

REFERENCIAS BIBLIOGRÁFICAS 34

INFORMACIÓN GENERAL

Titulo:

Desarrollo de un módulo ontológico para las consultas avanzadas de información de la Biblioteca Digital de la Escuela de Ingeniería de Sistemas y Computación de la Universidad del Valle

Proyecto de grado que se desarrollará para aplicar al titulo de Ingeniera de Sistemas de la Universidad del Valle.

Responsable del desarrollo del proyecto:

Diana Paola Hurtado Bustamante
Código 9709557
Estudiante de Ingeniería de Sistemas
Universidad del Valle
Teléfono (s): +57 (2) 337 49 05 – 311 308 1829
Dirección: Calle 14 # 46ª - 90 Torre 9D – 102
E-mail: dipahur@univalle.edu.co

Directora del Proyecto:

Martha Elena del Socorro Millán González
Ph.D en Informática, Universidad Politécnica de Madrid
Directora de la Escuela de Ingeniería de Sistemas y Computación
Profesora Titular de la Escuela de Ingeniería de Sistemas y Computación
Universidad del Valle
Teléfono (s): +57 (2) 3392355 Ext. 108
E-mail: millan@eisc.univalle.edu.co

RESUMEN

El desarrollo e implementación de Bibliotecas Digitales se ha convertido casi en una necesidad hoy en día, debido al creciente volumen de información con la que cuenta la mayor Biblioteca Digital existente: la WWW. Grandes cantidades de información son publicadas día a día por desarrolladores en el mundo y su impacto ha sido tal que es difícil obtener buenos resultados a la hora de hacer una consulta en ella.

La idea de una Biblioteca Digital en un dominio específico, es parte de una solución que pretende mejorar, entre otros, la recuperación de información válida para el usuario y que tratar de eliminar los factores de riesgo comunes, cuando, a diario, los usuarios se enfrentan a las búsquedas en Internet. Muchas veces, la información consultada por el usuario versus la información que se recupera, mediante algún tipo de buscador como por ejemplo Google [1], no es información satisfactoria y entonces se requiere que el usuario deba invertir una buena cantidad de tiempo en la selección de la información de su interés frente a otra gran cantidad de información que es completamente inútil.

A pesar de ello, mediante el desarrollo y uso de una ontología es posible mejorar los resultados de una consulta y ayudar al usuario para que el proceso de búsqueda de información sea un poco más fácil y automático. Todo ello depende de que la información se describa mediante un mecanismo de representación de conocimiento donde los datos pasen a ser metadatos de información con anotaciones en los sitios Web. Así, se hace más fácil identificar la información correspondiente a conceptos claros del dominio que se ha especificado y en el caso particular de este proyecto, para el dominio de las Ciencias de la Computación.

El propósito del presente trabajo es desarrollar una aplicación basada en ontologías que sea una interfaz gráfica Web que sirva de apoyo al módulo de consultas, permitiendo que la búsqueda de información, mejore sustancialmente el proceso de recuperación de documentos para el usuario para que de esta manera la obtención de la información sea más rápida y más efectiva.

En particular, en este proyecto la información se recupera de la Biblioteca Digital de la EISC. En esta Biblioteca, se pretende recopilar información de documentos, tesis y publicaciones relacionados con el área de Ciencias de la Computación. Tales documentos, en su mayoría, se han producido por parte de los estudiantes y personal docente de la Escuela de Ingeniería de Sistemas y Computación de la Universidad del Valle. La documentación asociada comprende, entre otras, las áreas de inteligencia artificial, bases de datos, lenguajes de programación y desarrollo de software.

INTRODUCCIÓN

El propósito de este proyecto es mostrar cómo por medio de las tecnologías asociadas a la Web Semántica, es posible contribuir al mejoramiento en la recuperación de la información en aplicaciones Web para un dominio específico.

Para todos los usuarios de la WWW es conocido que el contenido de información y otros recursos existentes es cada vez más gigantesco. Aunque Internet es llamada la gran Biblioteca mundial de la era Digital, esa comparación no está muy ajustada a la realidad puesto que tanta cantidad de información no es siempre accesible al usuario.

De este modo, la WWW ha evolucionado desde su aparición de manera que se puede considerar un gran repositorio caótico de todo tipo de información en forma Digital: textos cortos, libros, datos en bruto, propaganda comercial, video, audio, documentos efímeros, documentos duraderos, noticias, información gratuita, información vendida, información trivial, información valiosa y útil, etc. Por lo que más que una Biblioteca, es una gran plaza de información y comunicación.

Para dar sentido a la información que contiene la Web es necesario basarse en la capacidad de asociar semánticamente datos independientes entre sí, pero relacionados, en un determinado contexto. Sin embargo, las máquinas aún no son completamente capaces de procesar la información de ese modo, debido a la inexistencia de un mecanismo que permita el procesamiento semántico.

En los últimos años, se ha comenzado a experimentar con el uso de nuevas tecnologías y entre ellas, de mecanismos de representación del conocimiento para extender la Web actual con metadatos y anotaciones. De esta manera, se dota una nueva Web de más significado, donde las máquinas no sólo son capaces de presentar toda la información contenida en ella, es decir, la sintaxis, sino que además, puedan entenderla y gestionarla de forma "inteligente", es decir, la semántica. Esto es lo que actualmente se conoce como la Web Semántica.

Uno de los objetivos que se procuran con la Web Semántica consiste en definir y describir los datos contenidos en la Web, de tal forma que se puedan utilizar y comprender por las máquinas sin necesidad de intervención humana. De este modo, la Web se puede convertir en un espacio navegable y comprensible, donde es posible la relación entre términos independientes de una misma búsqueda, es decir, términos que se generen en una sola búsqueda y que tengan el mismo significado o uno equivalente, ofreciendo como resultado sólo aquellas páginas Web que realmente se adaptan a los requisitos especificados durante la misma.

En la Escuela de Ingeniería de Sistemas y Computación se pretende implementar un módulo de la Biblioteca Digital Web que sirva como soporte y apoyo a las búsquedas y a la recuperación de la información basado en ontologías, de modo ésta se despliegue al usuario como una interfaz gráfica Web, que cuenta con las características mencionadas anteriormente sobre la representación de la información y la accesibilidad de ésta.

La idea es que los usuarios al hacer uso de ella, se encuentren con una aplicación Web amigable y eficaz que arroje los resultados esperados y ahorre tiempo por medio de los mecanismos semiautomáticos implementados que procesan las consultas.

OBJETIVOS

OBJETIVO GENERAL

Desarrollar un módulo ontológico integrado a la Biblioteca Digital de la EISC que sirva de soporte a las consultas avanzadas y a la recuperación de información de documentos, en el dominio de Ciencias de la Computación. La idea principal es que por medio del uso de las tecnologías de la Web Semántica, se puedan obtener los resultados específicos a las consultas realizadas por parte de los usuarios.

OBJETIVOS ESPECÍFICOS

1. Explorar la posible existencia de una ontología en el dominio de Ciencias de la Computación y en caso tal adaptarla a los requerimientos de la Biblioteca Digital de la EISC. En caso contrario, proceder al desarrollo de la ontología de acuerdo a los parámetros requeridos para clasificar, describir y estructurar la información en este mecanismo de representación de conocimiento.

2. Proveer una interfaz gráfica que permita navegar a través de la ontología.

3. Enlazar los conceptos de la ontología con sus respectivas instancias en la base de datos de manera que se garantice el acceso por parte de algunas de las tecnologías asociadas a la Web Semántica y por consiguiente garantizar que los resultados esperados a las consultas son los satisfactorios para el usuario.

4. Integrar el módulo ontológico la Biblioteca Digital de la EISC.

5. Analizar, diseñar y ejecutar un plan de pruebas que compare las búsquedas en el módulo ontológico con el módulo de consultas basado en descriptores. De igual modo, realizar los ajustes necesarios al módulo ontológico que se tengan que llevar a cabo de acuerdo a los resultados encontrados.

FORMULACIÓN DEL PROBLEMA

DEFINICIÓN DEL PROBLEMA

En la actualidad la World Wide Web ha evolucionado de tal manera que ha generado un gran impacto en los usuarios. La mayoría de éstos han sido testigos de cómo las tecnologías han ido cambiando considerablemente en el transcurso del tiempo y de cómo los avances realizados para interactuar con esta gran Biblioteca Digital, han estado orientados a facilitar las labores de búsquedas en la Web, más no a mejorar la calidad significativa de los resultados.

Una de las más importantes aplicaciones y desarrollos producidos en esta nueva generación de la Web tiene relación con la estructuración de la información. Lenguajes como XML, XMLS, RDF, RDFS, XLS, las ontologías, los lenguajes ontológicos, los agentes inteligentes, lenguajes más formales como FLogic, Description Logic, entre otros, han sido propuestos por el consorcio W3C [2], y son comúnmente usados para representar y manejar la información adecuadamente y con mayor valor significativo. El W3C es un consorcio internacional encargado de proponer y desarrollar estándares Web, que está actualmente bajo la dirección de Tim Berners-Lee [3], creador de la WWW.

Con relación a las ontologías y a los lenguajes ontológicos, éstos permiten representar la información de manera consistente y válida, a partir de la cual, es posible aplicar procesos de inferencia sobre el conocimiento representado. De esta manera, se contribuye con las nuevas tendencias en la estructuración y manipulación de la información, de modo que se genere la semántica que se desea implementar en lo que se denomina la Web semántica.

La idea principal de la Web semántica es que los datos y la información, pasen a ser metadatos contenidos en las páginas y sitios Web. De esta manera, se puede usar y derivar la información, no sólo sintácticamente como se hace actualmente, sino con un alto nivel de semántica por medio de las anotaciones. Esto permitirá, por ejemplo, que los agentes inteligentes interactúen con tales metadatos con base en el significado que se ofrece en un dominio especificado. Así, procesos tales como la búsqueda y recuperación de recursos Web se pueden hacer de forma casi semiautomática.

El problema objeto de estudio que aquí se propone, tiene que ver con la problemática actual que se presenta en la Escuela de Ingeniería de Sistemas y Computación (EISC) de la Universidad del Valle relacionada con la manipulación de la información producida y consultada por docentes, investigadores y estudiantes de pregrado y postgrado. Esta información, se necesita tanto para ser representada y almacenada adecuadamente, como para ser consultada con un alto grado de rapidez y precisión.

Se propone, por lo tanto, desarrollar un módulo ontológico que pueda manipular la información de una manera semántica, para que mejore los resultados esperados de una consulta, lo que implica ofrecer e implementar opciones de búsqueda inteligente de forma que para el usuario sea amigable y fácil. Se propone que en tal módulo, las búsquedas se puedan desplegar gráficamente en el dominio en cuestión.

JUSTIFICACIÓN

La aplicación basada en ontologías que se pretende implementar para la Biblioteca Digital de la EISC, implica tener en cuenta importantes aspectos de desarrollo para facilitar la búsqueda de información significativa de forma sencilla. Esto ofrece, por un lado, la facilidad para analizar y hacer búsquedas apoyadas en una ontología en el área de las Ciencias de la Computación, y por otro, incrementa la calidad y confiabilidad de la información que se almacena.

Por otro lado, la recuperación de la información debe ser un proceso más amigable al usuario, con resultados más acertados y acorde con la consulta realizada, más aún porque es una consulta generada a partir de una interacción gráfica dirigida, permitiéndole al usuario mejorar su tiempo de búsqueda y recopilación de información útil y relevante.

El aspecto fundamental de la Web Semántica es la conversión de la estructura actual de la Web como almacén de datos, en una estructura de almacenamiento de información como conocimiento. En el primer caso sólo los humanos son los capaces de poner los datos en contexto, mientras que en el segundo caso el conocimiento se puede usar para hacer más eficaz el tratamiento de la información de una forma semiautomática gracias a que la información es también interpretable por las maquinas.

En el trabajo que aquí se presenta, se hará un análisis de la representación de la información en el dominio de las Ciencias de la Computación, por medio del desarrollo y/o adaptación de una ontología que cumpla con los requerimientos particulares de la Biblioteca Digital de la EISC. Así mismo, será necesario establecer los mecanismos gráficos necesarios para permitir la navegación en la ontología para obtener la información precisa requerida por el usuario y verificar las mejorías en la obtención y manipulación de la información.

ESTADO DEL ARTE

Algunos ejemplos de Bibliotecas Digitales en las cuales se utilizan mediadores para resolver los problemas causados por la heterogeneidad en la Biblioteca son: la Biblioteca Digital de *ERCIM* [4], *NDLTD* [5], el proyecto de Bibliotecas Digitales de Stanford [6], *Aleph* [7] y *AQUARELLE* [8].

Stanford propone la utilización de mediadores (*Library Service Proxies*) que se ocupan de las labores de traducción necesarias para integrar servidores implementados inicialmente sobre protocolos diferentes.

En *NDLTD –Networked Digital Library of Theses and Dissertations-* un especial interés se hace en los metadatos que describen los servidores; sin embargo, uno de sus fines primordiales es permitir consultas multilingües [33].

El proyecto *AQUARELLE* propone una solución para Bibliotecas distribuidas, utilizando el protocolo Z39.50. Cada servidor de datos que se incorpora al sistema debe implementar un servidor Z39.50.

Por último, el proyecto *CICYT* "*Biblioteca Virtual de Textos Españoles de Siglo de Oro accesibles vía Internet*" [9] consiste en la implementación de una Biblioteca Digital distribuida federada, uno de cuyos objetivos es conseguir la fácil integración de servidores heterogéneos, como son bases de datos relacionales y servidores de documentos XML.

En el área de la implementación de agentes mediante ontologías para la representación de la información se encuentra como tema relacionado [11], una tesis que describe la arquitectura de un sistema basado en agentes para la recuperación de recursos de información en un entorno de educación. En esta, se construyó una ontología en la cual se representan los documentos utilizados por estudiantes y profesores. Se utilizó el lenguaje *Resource Description Framework Schema* (RDF-S) para representar la ontología de documentos.

En la arquitectura del sistema se desarrollaron tres tipos de agentes (*UserAgent, SearchAgent* y *OntologyAgent*) que se encargan de la interacción con el usuario, la recuperación de la ontología y la búsqueda de metadatos. Con base de esta arquitectura se desarrolló un prototipo llamado *oSEMA* [10] para recuperar información en un entorno distribuido. Se encontró que, una organización descentralizada de agentes con sus propios repositorios RDF presenta un mejor tiempo de respuesta. Además, resultó ventajoso usar esta ontología para describir y compartir los documentos.

Adicionalmente existe una herramienta que se llama *JEROMEDL* [34] la cual es un sistema muy completo de bibliotecas digitales de código abierto. En JEROMEDL se ejemplifica como la administración del contenido de las bibliotecas digitales pueden beneficiarse de las tecnologías de la Web Semántica, en lo que tiene que ver con los perfiles de usuario, con la descripción semántica de recursos, con las características de consultas y desplegado de la información, y con respecto a la interconexión con otras bases de datos heterogéneas.

METODOLOGÍA

La metodología que se pretende llevar a cabo para contribuir al entendimiento y cumplimiento de cada uno de los objetivos es:

•Objetivo 1:

- Buscar qué tipo de ontologías existen en el área de Ciencias de la Computación y cuál es la más adecuada.
- Analizar cómo sería el proceso de adaptación de ésta de acuerdo a los requisitos actuales exigidos por la Biblioteca Digital de la EISC y qué procedimiento se requiere para que se lleve a cabo.
- Utilizar la herramienta *protégé* para la manipulación de la ontología.

•Objetivo 2:

- Construir la taxonomía con base en la jerarquía de conceptos dada por la ontología.
- Utilizar la herramienta adecuada para desplegar la taxonomía gráficamente.
- Analizar cómo es el tratamiento de una ontología en un lenguaje de desarrollo Web como php o java y realizar la respectiva implementación.

•Objetivo 3:

- Estudiar el uso de la herramienta R2O "Relacional To Ontology", la cual es la herramienta que propone el enlace entre los datos almacenados en la base de datos y las instancias correspondientes a la ontología.
- Implementar y probar la herramienta enlazando los nodos claves de la ontología a la información relevante almacenada en la base de datos.

•Objetivo 4:

- Revisar la documentación asociada a los módulos que componen la Biblioteca Digital de la EISC, además del módulo de consultas y verificar cuáles afectan el presente módulo.
- Estudiar todos los escenarios de funcionamiento de la Biblioteca Digital de la EISC, para analizar qué ajustes deben realizarse en la ejecución del módulo ontológico.
- Elaborar el escenario perteneciente al módulo ontológico.
- Elaborar las interfaces propias del módulo ontológico que estén enmarcadas dentro de la arquitectura de la Biblioteca Digital de la EISC.

•Objetivo 5:

- Establecer las medidas de comparación que permitan hacer las pruebas de las consultas por el método tradicional basado en descriptores y por el método que en el presente trabajo se propone.
- Construir una serie de consultas estándares las cuales serán la base para establecer tales medidas de comparación.
- Realizar pruebas experimentales que certifiquen la mejoría de las búsquedas a los usuarios, así como también la amigabilidad de la interfaz en la búsqueda dirigida. Tales pruebas experimentales se llevaran a cabo de manera dinámica por medio de unas encuestas que se realicen al usuario.
- Realizar los ajustes de implementación posibles necesarios de acuerdo a los resultados encontrados en las pruebas.

MARCO TEÓRICO

Este proyecto tiene como propósito desarrollar un módulo ontológico para las búsquedas y recuperación de información, para la Biblioteca Digital para la Escuela de Ingeniería de Sistemas y Computación de la Universidad del Valle. Con este propósito es necesario construir un marco teórico básico que incluye las tecnologías asociadas a la Web Semántica y el Proceso de Desarrollo de Software.

Aquí se presentarán los conceptos básicos asociados a cada una de éstas áreas.

LA WEB SEMÁNTICA

La idea de la Web semántica surgió como respuesta al caos de información que se encuentra actualmente en la Web, debido a que la representación actual de la información no se adecua al procesamiento de manera que permita obtener el mayor rendimiento a los usuarios.

La Web Semántica es una extensión de la actual Web la cual debe permitir encontrar, compartir y combinar la información eficazmente. Es un área que propone nuevas técnicas y paradigmas para la representación del conocimiento que faciliten la localización, compartición e integración de recursos a través de la Web. Estas nuevas técnicas se basan en la introducción de conocimiento semántico explícito que describa y/o estructure la información y servicios disponibles, de forma legible para ser procesada automáticamente por un programa. Por lo tanto, de acuerdo a la idea de Berners-Lee *"La información debe ser reunida de forma que un buscador pueda 'comprender', en lugar de ponerla simplemente en una lista".* [11]

Elementos Básicos de la Web Semántica

Para que la Web Semántica se desarrolle es necesario establecer ciertos estándares. En primer lugar, vale la pena mencionar la importancia de las URI (Unique Resource Identifier) que permiten tanto la identificación única de un recurso, como el establecimiento de las relaciones existentes entre los recursos.

Dado que la base de la Web Semántica es desarrollar la capacidad de establecer tipos y relaciones, se requiere de lenguajes para llevar a cabo tal fin y además vocabularios y reglas básicas que permitan definir las relaciones entre unos significados y otros.

La siguiente figura [12] presentada por Tim Berners-Lee en la *XML Conference de 2000*, puede servir como aproximación visual al conjunto de tecnologías importantes para el desarrollo de la Web Semántica.

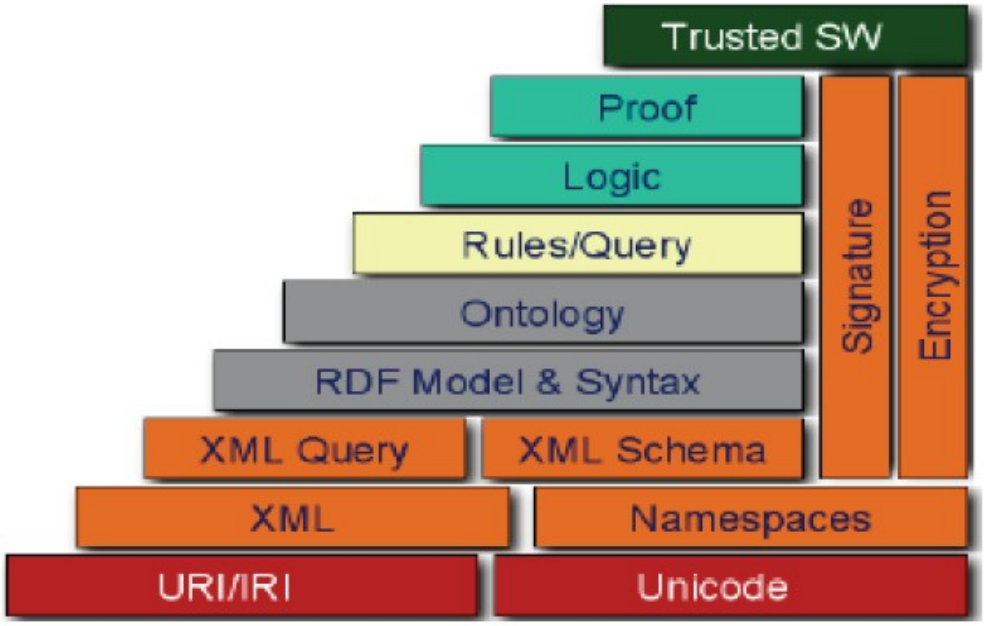

Figura 1. Capas de la Web Semántica

Los URI proporcionan el mecanismo para identificar de forma inequívoca cualquier recurso en la Web. En la Web Semántica, cumplen con el papel de identificadores de objetos del mundo real. Un subconjunto muy conocido de éstos, son los URL (Uniform Resource Locators) los cuales se usan a diario en la distinción de direcciones electrónicas en la Web.

Unicode, es simplemente una codificación del texto que permite utilizar los símbolos de diferentes idiomas sin observar caracteres extraños. Esto permite expresar información en la Web Semántica en cualquier idioma.

Otras tecnologías asociadas de mayor relevancia son:

- XML
- RDF
- ONTOLOGÍAS
- AGENTES.

XML

El XML [13] (eXtensible Markup Language), o Lenguaje Extensible de Marcas, es un lenguaje de marcas como HTML, que se enfoca más en la descripción y almacenamiento de los datos que a la forma cómo éstos son mostrados. Una de las ventajas de XML es que no hay etiquetas pre-definidas como en HTML. De este modo, se ofrece completa libertad al usuario para que cree sus propias etiquetas y estructuras según su conveniencia. Es importante destacar que XML fue diseñado para describir datos y HTML para presentar los datos y esto debe quedar muy claro a la hora de interactuar con ellos.

Una de las funcionalidades de XML, como se ha mencionado anteriormente, es que sirve para describir los datos, la relación existente entre ellos y la manera como éstos

son almacenados de manera estructurada. Dicha estructura es definida por el diseñador del documento XML, el cual tiene completa libertad para definir propios datos y organización.

Debido a que XML es un estándar, permite la comunicación entre máquinas que lo requieran siempre y cuando cada máquina tenga el correspondiente validador - DTD o XML Schema- (ya se hablará de ellos más adelante) para así poder conocer la definición de los tipos de datos, la organización y luego poder interpretarlos.

Un documento XML, se puede diseñar en cualquier editor de texto y desarrollar con cualquier aplicación sin problemas. Sin embargo, por lo general se hace necesario, realizar un proceso de validación del archivo para estar seguros de que está bien construido. Dicha validación se hace de dos formas: una es mediante un archivo DTD – actualmente esta validación ya esta un poco obsoleta - y la otra mediante la utilización de XML-Schema.

DTD: (Definición de Tipos de Documentos). Como su nombre lo indica, son definiciones de elementos que se pueden incluir en un documento XML y que da la forma en que se debe hacer (qué elementos van dentro de otros) y los atributos que se pueden tomar. En un DTD está definida la forma adecuada de construir los bloques de datos, se pueden encontrar una lista de elementos que pueden ir dentro del archivo XML y los DTD´s pueden estar embebidos en el archivo XML o en un archivo externo.

XML Schema [14]: Al igual que los DTD´s, el XML Schema describe la estructura de la información. El motivo de la creación de este nuevo estándar para realizar la labor de los DTD es, básicamente, la utilidad; Es un método de validación que surgió como alternativa debido a que los anteriores métodos de validación, no eran archivos XML y por lo tanto no prestaban mayores ventajas. Por tal motivo, un XML Schema es un archivo XML que contiene la misma información que los DTD´s con algunos complementos, como por ejemplo, definir qué elementos son permitidos dentro del documento XML, qué tipos de datos, qué restricciones sobre los datos, qué atributos son los que pueden aparecer, cuáles elementos son hijos de cuáles etc.

Entre las ventajas más importantes de este tipo de validador está que se pueden definir restricciones sobre los datos y darles formato, sin embargo, a estos esquemas se les debe aplicar las mismas reglas ya descritas para documentos XML, y además, para poder utilizar este validador se debe manejar namespace en XML.

RDF

RDF [15] (Resource Description Framework) o Infraestructura de Descripción de Recursos, es un modelo de metadatos, normalmente implementado como una aplicación de XML, desarrollada por el consorcio de la W3C. En otros términos, es un lenguaje multipropósitos general para representar información en la Web.

El objetivo de RDF [15] es definir un mecanismo para describir recursos que no creen ningún supuesto sobre algún dominio de aplicación, ni tampoco que defina a priori la su semántica; La definición del mecanismo debe ser neutral con respecto al dominio pero adecuado para describir la información sobre cualquier dominio.

Entre las características más sobresalientes de RDF se pueden destacar las siguientes:

- Está destinado a la representación de la meta-información sobre recursos Web tales como:
  - Título.
  - Autor.
  - Fecha de modificación de la página.
  - Derechos de autor.
  - Disponibilidad de un recurso compartido.
  - Descripción de las preferencias de usuario.
- Es una base para procesar metadatos.
- Proporciona interoperabilidad entre aplicaciones que intercambian información entendible por una máquina en la Web.
- Destaca la facilidad para habilitar el procesamiento automatizado de los recursos en la Web.
- RDF define recursos por medio de los URI (Uniforme Resource Identifiers), para que esté disponible en forma comprensible para la máquina.
- RDF en conjunto con las firmas Digitales, son la clave para construir el Web de confianza.
- La sintaxis RDF que se extiende de XML, es importante para codificar y transmitir metadatos de forma que maximicen la interoperabilidad de servidores y clientes Web desarrollados independientemente.
- RDF y XML son complementarios. Aunque debe ser claro que XML es sólo una sintaxis posible para RDF y que pueden surgir formas alternativas para representar el mismo modelo de datos RDF. Se puede utilizar en distintas áreas como:
  - Recuperación de recursos, para proporcionar mejores prestaciones a los motores de búsqueda.
  - Catalogación, para describir el contenido y las relaciones de contenido disponibles en un sitio Web, página Web, Biblioteca Digital particular, entre otras.
  - Agentes de software inteligentes, para facilitar el intercambio y para compartir conocimiento.
  - Calificación de contenido.
  - Descripción de colecciones de páginas.
  - Descripción de los derechos de privacidad.
  - Políticas de privacidad.

Es importante destacar que el modelo RDF, describe los recursos en términos de propiedades simples y valores de propiedad. Por ejemplo, en la figura 2, el concepto básico es que un recurso está descrito a través de una colección de propiedades llamadas RDF-Description. Cada una de estas propiedades tiene un tipo y un valor de la propiedad, y de este modo, el recurso es identificado con un URI.

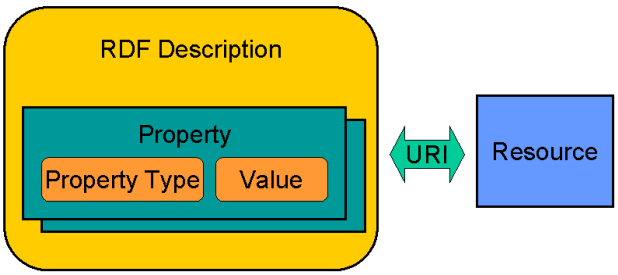

Figura 2. Modelo RDF

El modelo RDF es muy útil para representar metadatos en RDF y consiste de tres tipos:

- Recurso. Son todas las cosas descritas por expresiones RDF. Siempre se designan por un URI más un identificador especial opcional. Un recurso puede ser una página Web, una parte de una página Web, una colección completa de páginas Web o sitios Web, un libro, entre otros.
- Propiedades. Son un aspecto específico, característica, atributo o relación utilizada para describir un recurso. Tiene un significado específico, por lo que define sus valores permitidos, los tipos de recurso que puede describir y la relación con otras propiedades.
- Sentencias, declaraciones o enunciados. Es un recurso específico junto con una propiedad denominada, más el valor de dicha propiedad. Estas tres partes individuales de la sentencia se denominan respectivamente sujeto, predicado, objeto. Éste último puede ser un valor de la propiedad, un recurso especificado por un URI o puede ser un literal, es decir, una cadena de caracteres o string.

La Figura 3 muestra un ejemplo de RDF que describe la información de una persona y está representado como un grafo. Se observa que RDF usa un URI para identificar diferentes tipos de recursos y cosas como se describe a continuación:

- Individuos, ejemplo *Eric Miller*, identificado por http://www.w3.org/People/EM/contact#me.
- Tipos de cosas, ejemplo *Person* identificado por http://www.w3.org/2000/10/swap/pim/contact#Person.
- Propiedades de estas cosas, ejemplo *mailbox* identificado por http://www.w3.org/2000/10/swap/pim/contact#mailbox
- Valores de éstas propiedades, ejemplo mailto:em@w3.org

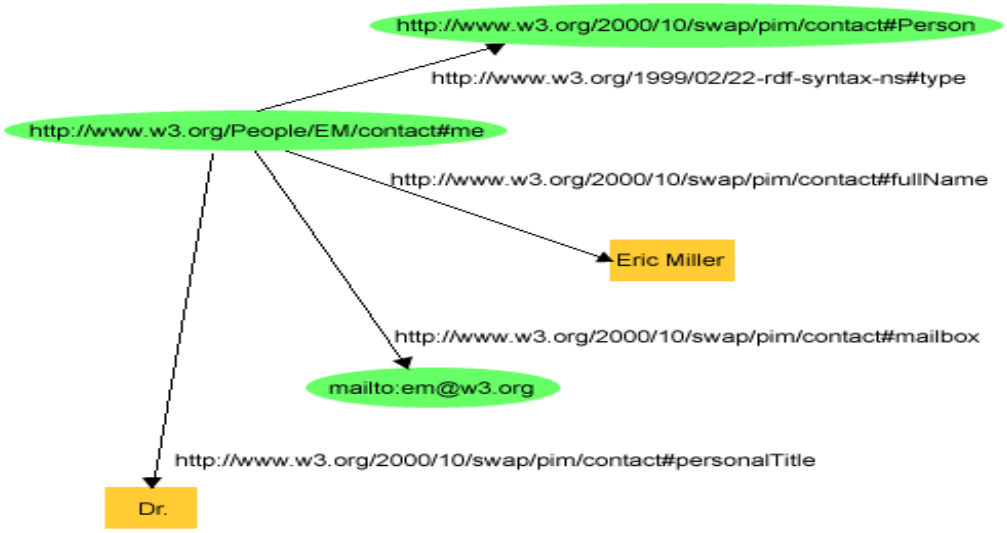

Figura 3. Un grafo RDF

RDF Schema es un conjunto simple de recursos y propiedades RDF estándar que permite a las personas crear sus propios vocabularios RDF. El modelo de datos expresado por RDF Schema es el mismo modelo de datos usado por los lenguajes de programación orientados a objetos como Java. Este modelo de datos permite la creación de clases de datos donde una clase está definida como un grupo de cosas con características comunes.

Algunas de las características más sobresalientes de RDF Schema son:

- Un esquema define no sólo las propiedades de un recurso (Ej. título, autor, materia, tamaño, color, etc.) sino que puede también definir los tipos de recursos que se describirán (libros, páginas Web, personas, empresas, etc.).
- El lenguaje de especificación del RDF Schema es un lenguaje de representación influenciado por las ideas de la representación del conocimiento tales como las redes semánticas, marcos, lógica de predicados entre otros. Sin embargo, es mucho menos expresivo, aunque mucho más sencillo de implementar, que lenguajes de cálculo de predicado completo tales como *CycL* [16] y *KIF* [17].
- El RDF Schema proporciona información sobre la interpretación de una sentencia dada en un modelo de datos RDF.

RDF Schema tiene un conjunto de términos los cuales permitirán construir las expresiones RDF Schema válidas. Así, se tienen los términos *Class*, *subClassOf* y *Property* que permitirán formular expresiones acerca de los recursos (clases), representar sus propiedades y a la vez representar una jerarquía de recursos. Los objetos podrán ser instanciados a partir de las clases usando la propiedad *type*. Las restricciones de las propiedades podrán ser especificadas usando *domain* y *range*.

En la figura 4 se muestra un ejemplo de RDF Schema.

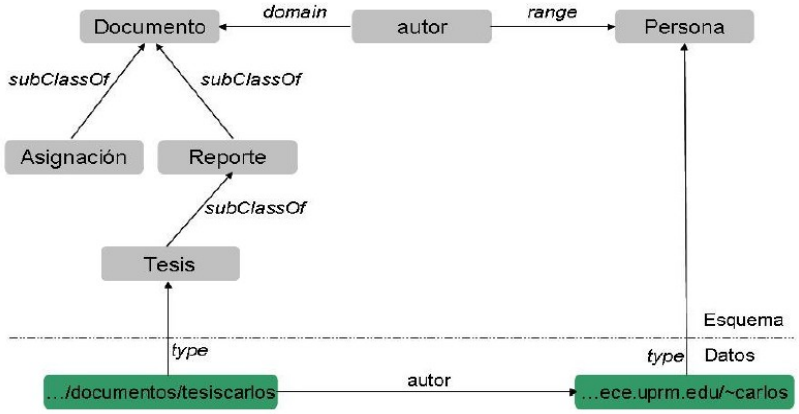

Figura 4. Ejemplo de un RDF Schema [10]

ONTOLOGÍAS

Para que las máquinas puedan intercambiar información de forma efectiva y eficiente, necesitan acceder a colecciones estructuradas de información y a formalismos que les permitan tener un cierto grado de razonamiento automático. Estas necesidades se pueden cubrir utilizando ontologías para anotar los recursos Web.

El término ontología, utilizado en filosofía para hablar acerca de una '*teoría sobre la existencia*' [18] ha sido adoptado por la comunidad de investigadores de inteligencia artificial para definir una categorización y las relaciones entre sus términos. El término categorización ha sido últimamente muy usado como sinónimo de ontología y en algunos casos, es usado indistintamente para significar la misma idea.

De acuerdo con *Gruber (1993*), una ontología es "*una especificación explicita y formal de una conceptualización compartida*"; Con esto la definición de *Gruber* quiere decir que [18]:

- Conceptualización se refiere a un modelo abstracto de algún fenómeno del mundo del que se identifican los conceptos que son relevantes.
- Explícito hace referencia a la necesidad de especificar de forma consciente los distintos conceptos que conforman una ontología.
- Formal indica que la especificación se debe representar por medio de un lenguaje de representación formalizado.
- Compartida refleja que una ontología debe, en el mejor de los casos, dar cuenta de conocimiento aceptado, como mínimo, por el grupo de personas que deben usarla.

Más específicamente, una ontología está formada por una taxonomía relacional de conceptos y por un conjunto de axiomas o reglas de inferencia mediante los cuales se podrá inferir conocimiento nuevo. Las principales motivaciones para el desarrollo de una ontología tienen que ver con los siguientes aspectos:

- Compartir la comprensión común de la estructura de información entre personas y/o agentes de software.
- Permitir la reutilización del conocimiento perteneciente a un dominio.
- Hacer explícitos los supuestos de un dominio.
- Tratar de mostrar la distinción entre el conocimiento del dominio y el conocimiento operacional.
- Analizar el conocimiento del dominio.

Por tanto, las ventajas en el desarrollo de las ontologías son innumerables, algunas de ellas son:

- Presentan un nivel más alto de concepción y de descripción del vocabulario.
- Potencian la comunicación entre usuarios y máquinas.
- Promueven la normalización y reutilización de la representación de la información.
- El desarrollo semántico es más profundo, es decir, niveles de jerarquía más profundos, fuertes relaciones entre clases y conceptos y capacidad de formular reglas de inferencia.
- Las ontologías destacan el uso de la lógica descriptiva empleada para la descripción de situaciones.

La Web Semántica ha sido el resultado del esfuerzo y el desarrollo de varias tecnologías - las que se han mencionado anteriormente - que ha requerido, para cumplir las metas que se ha propuesto y dejar de ser tan sólo la ilusión que visionan muchos investigadores en el área. Es así como en los últimos años, se ha comenzado a experimentar con el uso de técnicas de representación del conocimiento, entre ellas las ontologías, retomadas del área de la IA, para extender la actual Web actual con metadatos. Así, las máquinas no sólo serían capaces de presentar toda la información contenida en ella, sino que además, podrían entenderla y gestionarla de forma inteligente.

En un dominio específico se pueden identificar los siguientes pasos para la creación de una ontología [19]:

- Lo primero es determinar el dominio y el alcance de la ontología, es decir, sus limitaciones y medir la calidad de su desarrollo y si vale la pena o no.
- Una vez el punto anterior esté claro, lo segundo es verificar que la ontología que se desea desarrollar ya exista y ver la posibilidad de reutilizarla de acuerdo a los requerimientos propios que se tengan.

- En el dominio del problema, tratar de identificar los conceptos más relevantes por medio de un proceso completo de abstracción de información en conocimiento.
- De acuerdo con los conceptos, definir las clases y la jerarquía de tales clases así como en el modelamiento que se realiza en el paradigma orientado a objetos.
- Una vez se establezca la jerarquía de clases, definir las propiedades de las clases, es decir los atributos y características mas sobresalientes.
- Posteriormente definir los aspectos de interés relacionados con los atributos. Esto es, el tipo, la cardinalidad etc.
- Y para finalizar, la creación de instancias, es decir, de casos particulares de los conceptos.

Es mucho lo que se puede decir acerca de las ontologías, pues son muchas las áreas de aplicación en las que se pueden emplear. Por ejemplo, en la ingeniería y gestión del conocimiento, en el procesamiento del lenguaje natural, en la recuperación de información, en el comercio electrónico, en la integración inteligente de servicios, entre otros, y para cada área, una ontología tiene diversas características como por ejemplo el tipo de ontología, la clasificación a la que pertenece la ontología, etc. Pero lo más importante es que una ontología siempre debe mantener el objetivo de ser consensuada y compartida por quienes hagan uso de ella en el dominio especificado.

LENGUAJES ONTOLÓGICOS

A través de la historia de las ontologías, se han creado y estandarizado muchos lenguajes que han cumplido un rol muy importante, puesto que en el proceso de explotar la Web Semántica se han requerido lenguajes de marcado apropiados que representen el conocimiento de las mismas.

Actualmente, mediante anotaciones RDF y RDF Schema se pueden representar algunos aspectos sobre los conceptos de un dominio y permite, mediante relaciones taxonómicas, crear una jerarquía de conceptos. Pero se necesitan lenguajes de marcado, basados en RDF con mayor expresividad y capacidad de razonamiento para representar los conocimientos que contienen las ontologías [20].

Son muchos los lenguajes para ontologías existentes, se pueden destacar lenguajes ontológicos tradicionales como *Ontolingua* [21], *KIF* [17], *OCML* [22], *Flogic* [23], *LOOM [*24]. También lenguajes para la Web estandarizados por el consorcio de la W3C como *XML* [13], *XOL* [25], *SHOE* [26], *DAML* + *OIL* [27] e indudablemente *OWL* [28].

Es importante destacar a OWL, pues es el lenguaje que más importancia tiene actualmente para el desarrollo de las ontologías, por lo tanto OWL (*Web Ontology Language)*, o el Lenguaje de Ontologías para la Web es el estándar que se ha estado desarrollando como una versión de DAML+OIL por el W3C. Fue desarrollado para ser compatible con la arquitectura de la World Wide Web en general, y la Web Semántica en particular. Por lo tanto proporciona un lenguaje que utiliza la conexión suministrada por RDF para añadir las siguientes capacidades a las ontologías [29]:
- Capacidad de ser distribuidas a través de varios sistemas.
- Escalable a las necesidades de la Web.
- Compatible con los estándares Web de accesibilidad e internacionalización.
- Abierto y extensible.

OWL extiende RDF Schema para permitir la expresión de relaciones complejas entre diferentes clases RDF Schema [29], y una mayor precisión en las restricciones de

clases y propiedades específicas. Esto es, es un lenguaje de marcado para la publicación de ontologías en la WWW y tiene como objetivo facilitar un modelo de marcado, construido sobre RDF y codificado en XML que permita representar ontologías a partir de un vocabulario más amplio y una sintaxis más fuerte que la que permite RDF. Es el lenguaje que actualmente más se esta usando en la descripción de ontologías pues supone un nuevo horizonte en el marcado de sistemas de organización del conocimiento que tiene como punto de partida las experiencias previas realizadas con DAML-OIL en las cuales se inspiraron los creadores de OWL para desarrollar el lenguaje. OWL puede ser utilizado para representar de forma explícita el significado de términos pertenecientes a un vocabulario y definir las relaciones que existen entre ellos. OWL se divide en tres sublenguajes OWL-Lite, OWL-DL y OWL-Full, cada uno de los cuales proporciona un conjunto definido sobre el que trabajar, siendo el más sencillo OWL-Lite y el más completo OWL-Full. Cada uno de ellos varía en la expresividad que permite así como también en el nivel de inferencia de conocimiento que se puede obtener.

HERRAMIENTAS PARA ONTOLOGÍAS [30]

Las herramientas de anotación son las que permiten por una parte estructurar la información que ahora esta publicada de forma desestructurada, mediante una clasificación con base en conceptos semánticos; y por otro lado, añadir información adicional, estructurada o no, a los contenidos actuales de la Web y a los de nueva creación.

Existe un proyecto llamado RODA (colocar la referencia) que consiste en evaluar herramientas de anotación, tanto de las que llamadas de "anotación externa", que facilitan la gestión del conocimiento adquirido a través de la navegación por la Web; como de las de "autor", que facilitan la estructuración de la información publicada. También evalúa todas aquellas herramientas que son complementarias a las herramientas de anotación.

Por lo que se refiere a las herramientas de "anotación externa", éstas harán posible la anotación de las páginas Web visitadas por los miembros de una comunidad de usuarios; la meta información (introducida a través de las anotaciones) relacionada con la página Web se almacenará en un repositorio común, al cual podrán acceder todos los miembros de la comunidad.

Las herramientas de "autor", facilitarán la estructuración de la información publicada en la Web, ya sea en el ámbito de Intranet, Extranet o Internet, haciendo posible la inclusión de meta información en los documentos publicados.

Las herramientas de creación de ontologías o editores de ontologías, son las herramientas que permiten la codificación de una determinada ontología en base a un determinado lenguaje. Se estudia este tipo de herramientas ya que permiten definir la estructura en base a la cual se clasificará la información realizada a través de anotaciones. Algunas herramientas de este tipo son:

- Apollo: Aplicación amigable de modelado de conocimiento. El modelado está basado en torno a los principios básicos tales como clases, instancias, funciones, relaciones, etc. La interface de usuario tiene una arquitectura abierta y está escrito en lenguaje de programación JAVA.
- LinkFactory: Se trata de una herramienta utilizada para construir completos sistemas de terminología corporativa capaz de extraer valor significativo de gran

cantidad de datos no estructurados almacenados en bases de datos de contenido corporativo.

- OILEd: Es un editor de ontologías que permite al usuario construir ontologías utilizando DAML+OIL.
- OntoEdit Free and Professional versions: Permite crear y gestionar ontologías. Confía en los estándares del W3C y ofrece muchas interfaces exportables a la mayor parte de lenguajes de representación de ontologías. Esta herramienta permite crear, navegar y modificar ontologías.
- Ontolingua Server: Provee de un entorno de colaboración distribuido para navegar, crear, editar, modificar y utilizar ontologías.
- OntoSaurus: Es un navegador Web para las bases de conocimiento de LOOM. Proporciona una interfaz gráfica hyperenlazada hacia varias de las bases de conocimiento.
- OpenKnoME: Es la piedra angular de la aplicación utilizada por los motores de conocimiento topThing. Es un sistema de gestión del conocimiento y un motor de ontologías. Desde el 2001 el código fuente está abierto para la comunidad académica y la clínica sin ánimo de lucro.
- Protégé-2000: Editor de ontologías y editor de bases de conocimiento. Es también de código abierto, herramienta JAVA que proporciona una arquitectura extensible para la creación de aplicaciones de bases de conocimiento personalizadas.
- SymOntoX: Es un software que almacena y gestiona un dominio de ontología.
- WebODE: Herramienta para modelar el conocimiento utilizando ontologías. Facilita la máxima flexibilidad e interoperabilidad con otras aplicaciones de negocios necesarias para las empresas de hoy en día.
- WebOnto: JAVA applet con un servidor Web personalizado que permite a los usuarios navegar y editar modelos de conocimiento sobre la Web.

Las herramientas de anotación externas permiten asociar meta información a páginas Web que ya están presentes en la Web. La meta información no se almacena dentro de la misma página, sino que se almacena de forma externa en un repositorio destinado específicamente a mantener las anotaciones. Estos repositorios suelen ser bases de datos RDF (que actualmente se implementan sobre sistemas gestores de base de datos relacionales). Esta meta información también puede estar basada, o no, en una ontología. Dentro de esta categoría se encuentran tanto editores como servidores de anotaciones. Ejemplos de este tipo de herramientas son:

- COHSE (The Conceptual Open Hypermedia Project): Proyecto de investigación sobre métodos que mejoren significativamente la calidad, consistencia y la amplitud de documentos Web enlazados mientras se recuperan (cuando los lectores navegan sobre los documentos) y al mismo tiempo que se crean (cuando los autores crean los documentos). Se utilizan tres tecnologías: un servicio de razonamiento de ontologías que se utiliza para representar un sofisticado modelo conceptual de términos documentales y sus relaciones, un servicio abierto de enlaces hypermedia basado en Web, que sea escalable, y la integración de estos dos para poder enlazar documentos vía metadata describiendo sus contenidos.
- Annotea: Annotea es un proyecto que pretende mejorar el ambiente colaborativo de la W3C (el "Consorcio de la World Wide Web") a través del uso de anotaciones compartidas; se entiende que una anotación puede ser un comentario, una nota, una explicación o cualquier texto que se pueda adjuntar a un documento Web externamente, es decir, sin necesidad de tocar el documento mismo. El usuario recibe las notas de un servidor o varios adjuntas a un documento junto con éste y puede ver qué piensan sus colegas del documento.
- Annozilla: Proyecto diseñado para ver y crear anotaciones asociadas a una página Web, como define el proyecto Annotea del W3C.

- Yawas (Yet Another Web Annotation System): Programa que permite destacar partes de un documento Web.
- Annotation System: Sistema basado en tres arquitecturas vinculadas: un cliente que puede realizar / buscar anotaciones, un servidor de aplicaciones que gestiona todas las peticiones del cliente, un servidor de base de datos para almacenar las anotaciones realizadas por los usuarios.
- Trellis Web: Entorno interactiva que permite a los usuarios añadir sus observaciones, puntos de vista y conclusiones como la información que analizan realizando anotaciones semánticas a los documentos y a otros recursos on-line.

Las herramientas de autor, que parten de una ontología definida previamente, permiten incluir la información estructurada, la "meta información", en la misma página Web. Esta información se incluye normalmente en alguno de los nuevos lenguajes de marcado promovidos por el W3C, como son XML, o RDF, este último es un lenguaje definido específicamente para representar meta información (RDF, se suele implementar a través de XML). Para la definición de las ontologías se suela usar DAML, aunque actualmente el W3C está trabajando en la definición de un lenguaje estándar para definir ontologías, OWL. Algunos ejemplos de este tipo de herramientas son:

- MnM: MnM es una herramienta de anotación basada en ontologías que permite anotar páginas Web con contenidos semánticos de forma automática y semi-automática. MnM integra un navegador Web con un editor de ontologías y proporciona unas APIs abiertas para enlazar MnM con servidores de ontologías y para integrar MnM con herramientas de extracción de información. MnM trabaja con diferentes lenguajes de ontologías como RDF, DAML+OIL y OCML.
- OntoAnnotate: Las ontologías funcionan como base conceptual para el intercambio de conocimientos con diferentes propósitos. Con OntoAnnotate se puede mediante la navegación habitual en la Web y en documentos internos ampliar su base de conocimientos, al instante. La herramienta permite aportar el conocimiento de cada uno. El usuario puede aprehender hechos individuales que corren directamente en la base de conocimientos, con ellos sus asociaciones, relaciones, sus contextos en las informaciones serán capaces de ser transmitidas, interpretadas y valoradas para sistemas capaces de procesar como la máquina de inferencias Ontobroker. OntoAnnotate puede ser utilizado junto a las anotaciones de los documentos para proveer de contenidos semántico, los cuales y de forma compatible con los estándares son directamente accesibles e interpretables como contenidos en la Web Semántica. Permite a cada usuario crear documentos con contenidos semánticos capaces de ser entendidos por las máquinas.
- OntoMat-Annotizer: Herramienta de anotación de páginas Web interactivas amigables. Ayuda al usuario en la tarea de crear y mantener ontologías basadas en DAML+OIL, por ejemplo, para crear instancias, atributos y relaciones. Incluye un navegador de ontologías para explorar la ontología y las instancias y un navegador HTML que visualiza las partes anotadas del texto. Está basado en JAVA.
- SHOE Knowledge Annotator: Lenguaje de representación del conocimiento basado en HTML. Añade información semántica a las páginas Web a través de etiquetas. Las etiquetas están divididas en dos categorías: para la construcción de ontologías y para anotar documentos Webs para suscribirlos a una o más ontologías, declarar entidades de datos y realizar afirmaciones sobre las entidades según las normas proscritas por las ontologías.
- SMORE: Herramienta que permite a los usuarios etiquetar sus documentos en RDF utilizando ontologías Web asociadas a los elementos y términos específicos.
- Melita: Melita es una herramienta de anotación de texto semi-automática basada en ontologías que tiene integrada una máquina de extracción de información

(Amilcare). Implementa una metodología con el propósito de gestionar para los usuarios todo el proceso de anotación. Algunos de los pasos del proceso, que hasta ahora se realizan manualmente, el sistema los puede automatizar y gestionar fácilmente. Las principales capacidades de Melita se pueden resumir en cuatro grupos, la tarea de gestión, la extracción, el aprendizaje, y el etiquetado de la Información de forma autónoma.

- GATE (General Architecture for Text Engineering): es una infraestructura de componentes de software reutilizables para dar soporte al procesamiento del lenguaje humano.

METODOLOGÍAS PARA LA CONSTRUCCIÓN DE ONTOLOGÍAS [31].

Como en cualquier otro campo tecnológico, para construir ontologías de una manera fiable es aconsejable emplear una adecuada metodología, siguiendo los procesos de desarrollo y las actividades y técnicas indicadas. En el campo de la ingeniería del conocimiento así como en el naciente campo de la ingeniería ontológica existe un sin numero de metodologías de muchos tipos que sirven de apoyo a la hora de desarrollar una ontología. Por lo tanto, en el campo de la ingeniería del conocimiento se pueden destacar la metodología de Waterman[Wat86], CommonKADS[Wie92], IDEAL[Gom97] entre otras y en el campo de la ingeniería ontológica se pueden destacar metodologías tales como CYC[Len90], Uschold y King[Usc95], Gruninger y Fox[Gru95b], la propuesta de Bernaras y colegas[Ber96], la metodología de SENSUS[Swa97], METHONTOLOGY[Fer97] y [Gom98], Co4 methodology [Euz95] y $(KA)^2$ methodology[Ben99]

La siguiente tabla fue tomada de un resumen del estudio de Fernández [Fer99b] sobre metodologías de desarrollo de ontologías.

	Uschold & King	**Gruninger & Fox**	**Bernaras et al.**	**METHONTOLOGY**	**Basado en SENSUS**
Herencia de Metodologías de IC	Parcial	Poca	Grande	Grande	Ninguna
Detalle de actividades y técnicas	Muy poco	Nada	Muy poco	Muy detallado la mayor parte	Medio
Recomendaciones para la formalización del conocimiento	Ninguno	Lógica	Ninguno	Ninguno	Redes semánticas
Estrategia para construir ontologías	Independiente de la aplicación	Semidependiente de la aplicación	Dependiente de la aplicación	Independiente de la aplicación	Semidependiente de la aplicación
Estrategia para identificar conceptos	Media	Media	De arriba a abajo	Media	De abajo a arriba
Ciclo de vida recomendado	Ninguno	No definido totalmente	Ninguno	Prototipos evolutivos	No definido totalmente
Diferencias con IEEE 1074-1995	Faltan procesos y actividades	Faltan procesos y actividades	Faltan procesos y actividades	Faltan procesos, predesarrollos y actividades	Faltan procesos y actividades
Técnicas recomendadas	Desconocido	Desconocido	Desconocido	Faltan algunas técnicas	Desconocido

Ontologías desarrolladas	Enterprise Ontology	TOVE	Ontologías para redes eléctricas	CHEMICALS	SENSUS

AGENTES DE SOFTWARE INTELIGENTE

Un agente es una entidad de software que funciona continua y autónomamente en un medio particular a menudo habitado por otros agentes y procesos sin requerir de guía constante o intervención humana. Es decir, un asistente personal que está dentro del computador y que cumple varios roles en representación de una función específica o de un usuario. En términos de la Web semántica, son los encargados de realizar la búsqueda de servicios, para lo cual la semántica será la encargada de facultarlos para describir unos a otros la función exacta que realizan y qué datos han de recibir para ello, es decir, los que recogen, filtran y procesan información contenida en la Web, realizando inferencias sobre dicha información e interactuando con el entorno. Las características que debe tener un agente inteligente son:

- Comunicativo. El agente debe entender las necesidades, objetivos y preferencias del usuario, ya que de dicha comunicación o interacción usuario-agente depende que el agente pueda llevar a cabo su función de forma eficiente. Así mismo debe poder comunicarse con el entorno mediante representaciones compartidas de conocimiento u ontologías.
- Autónomo. El agente, además de comunicarse, debe poder interactuar con el entorno, tomando decisiones y actuando por sí solo, limitando sus acciones según el nivel de autonomía permitida por el usuario.
- Adaptable. El agente debe ser capaz de aprender del entorno: usuarios (preferencias), fuentes de información y otros agentes.

Los agentes inteligentes también requieren de un ambiente en el cual se cumplan las siguientes características:

- Una red donde los recursos estén bien identificados.
- Una reserva de todo el conocimiento disponible, es decir, una ontología.
- Una descripción de las preferencias del usuario.
- Capacidad de inferencia.

Estas últimas características son ofrecidas por las ontologías que, como ya se ha dicho, son las encargadas de representar el conocimiento, por medio de anotaciones en un formato que sea entendido por los agentes Web. Tal formato está representado de acuerdo a las especificaciones del RDF sobre XML. En conclusión, el papel principal de un agente de software es permitir, a partir de la información contenida en las descripciones, que se pueda descubrir servicios automáticamente, como también invocarlos y componerlos para así monitorizar su ejecución.

Con el tema de agentes se concluye la descripción de las tecnologías más importantes de la Web Semántica. En términos muy generales, la siguiente figura trata de ilustrar los aspectos, las tecnologías y el desempeño que se desea llevar a cabo en esta nueva Web.

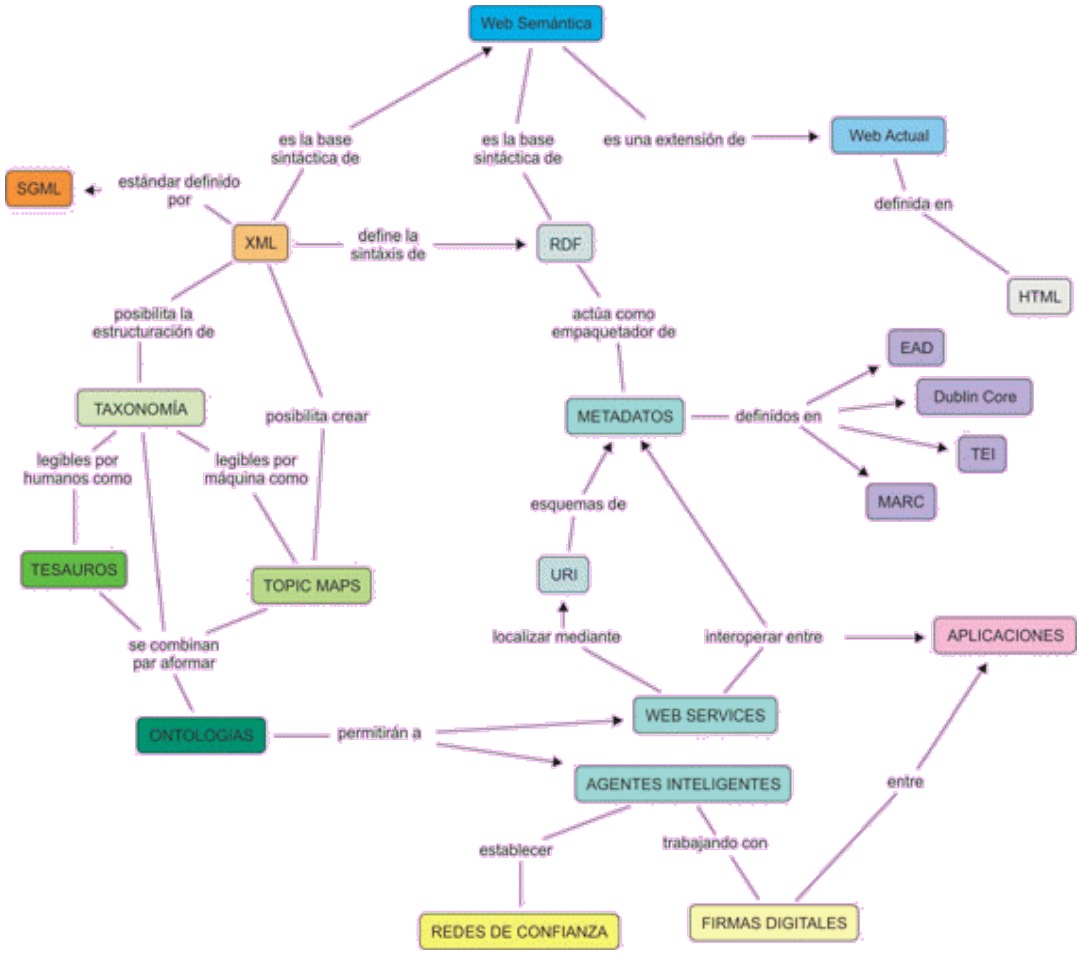

Figura 5. Mapa Conceptual de la Web Semántica [32].

PROCESO DE DESARROLLO DE SOFTWARE

El proceso de desarrollo de software es el proceso mediante el cual se desarrollan una serie de etapas que se componen, a su vez, de tareas para la realización de un proceso en particular. [Pressman]

En este proyecto se tomará como base la metodología RUP la cual consta de dos dimensiones:

- Tiempo, el cual es la división del ciclo de vida en fases e iteraciones.
- Componentes del Proceso, el cual es la producción de un conjunto específico de artefactos con sus actividades bien definidas.

Estructurando el proyecto a lo largo de la dimensión del tiempo, se ven involucradas las siguientes fases:

- Inicio. Es la especificación inicial del proyecto.
- Elaboración. Es la especificación de las características y diseño de la arquitectura.
- Construcción. Es como tal, la construcción del producto en una serie de iteraciones incrementales.
- Transición. Es la fase en la que se suministra el producto a quien lo va a usar.

Si se estructura el proyecto a lo largo de la dimensión de los componentes del proceso, se ven involucradas las siguientes actividades:

- Análisis de requerimientos. Es la descripción de lo que el sistema debe hacer.
- Diseño. Cómo el sistema lleva a cabo la fase de implementación.
- Implementación. Es la producción del código como tal.
- Prueba. Es la verificación del sistema.

Cada actividad de la dimensión de los componentes del proceso es aplicada a cada una de las fases de la dimensión basada en el tiempo. Ver la figura 6.

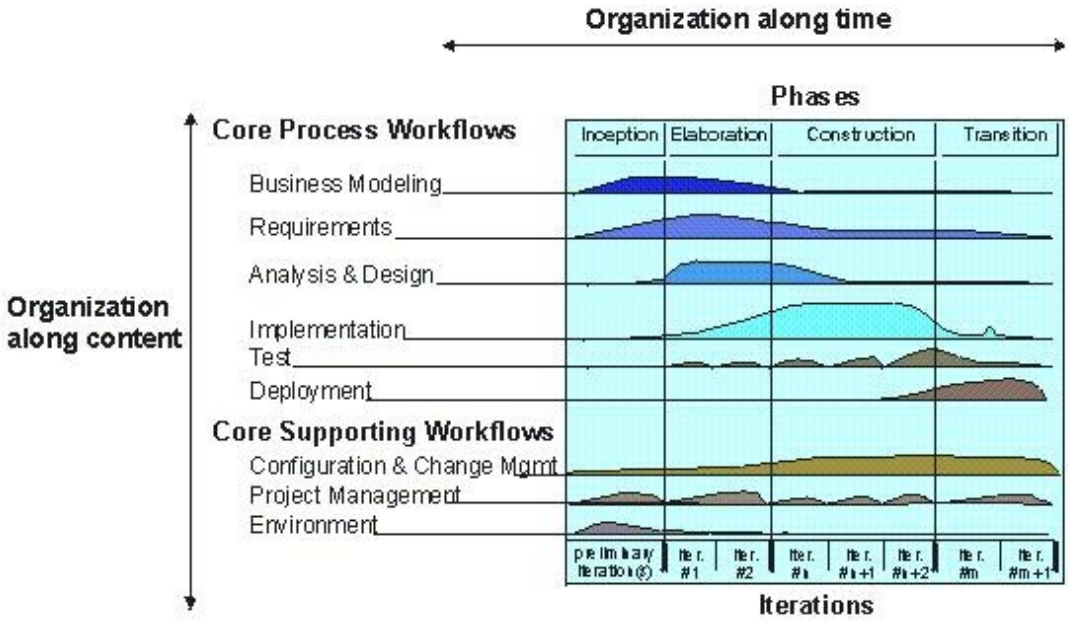

Figura 6. El proceso de Desarrollo de Software

ASPECTOS ADMINISTRATIVOS

PLAN DE ACTIVIDADES Y CONTENIDO.

ACTIVIDAD	PARTICIPANTES	SEMANA
Definición del problema.	Diana Paola Hurtado Bustamante	1
Mostrar avances e investigación en el tema.	Diana Paola Hurtado Bustamante	2
La especificación de los avances realizados e investigaciones que se puedan extraer para el desarrollo de la aplicación.	Diana Paola Hurtado Bustamante	3
Especificación de requerimientos funcionales y no funcionales.	Diana Paola Hurtado Bustamante	4
Análisis y estudio de la ontología para Ciencias de la Computación.	Diana Paola Hurtado Bustamante	5
Análisis y estudio del lenguaje de consulta.	Diana Paola Hurtado Bustamante	6
Análisis y estudio de la herramienta para el tratamiento de la ontología.	Diana Paola Hurtado Bustamante	6
Análisis y estudio de la metodología para la ontología.	Diana Paola Hurtado Bustamante	6
Búsquedas de herramientas disponibles para cumplir con la parte gráfica y con el requerimiento gráfico del proyecto.	Diana Paola Hurtado Bustamante	7 – 10
Búsqueda de los antecedentes teóricos, tecnológicos y prácticos que presenta desarrollar la aplicación.	Diana Paola Hurtado Bustamante	10 – 11
Buscar Información sobre R2O.	Diana Paola Hurtado Bustamante	12
Probar R2O.	Diana Paola Hurtado Bustamante	13
Se analizará y se escogerá de todas las alternativas posibles tanto en metodologías como en herramientas ya diseñadas y se empezara con gran parte de la implementación.	Diana Paola Hurtado Bustamante	14
Concordancia e integración de la ontología con la BD y las instancias respectivas.	Diana Paola Hurtado Bustamante	15 – 17
Definición de los casos de uso.	Diana Paola Hurtado Bustamante	18

Definición preliminar del modelo conceptual.	Diana Paola Hurtado Bustamante	19
Mejoramiento del modelo conceptual.	Diana Paola Hurtado Bustamante	20
Definición de los diagramas de secuencia del sistema.	Diana Paola Hurtado Bustamante	21
Definición de los casos de uso reales.	Diana Paola Hurtado Bustamante	22 – 23
Definición de los diagramas de clases.	Diana Paola Hurtado Bustamante	24
Definición del plan de pruebas.	Diana Paola Hurtado Bustamante	25 – 26
Puesta en marcha del plan de pruebas	Diana Paola Hurtado Bustamante	27 – 29
Terminación de la implementación y entrega	Diana Paola Hurtado Bustamante Diana Paola Hurtado Bustamante	30 – 36

CRONOGRAMA DE ACTIVIDADES

PARTE 1

SEMANA DE LA 1 A LA 18

ACTIVIDADES	TIEMPO EN SEMANAS																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Definición del problema.																		
Mostrar avances e investigación en el tema.																		
La especificación de los avances realizados e investigaciones que se puedan extraer para el desarrollo de la aplicación.																		
Especificación de requerimientos funcionales y no funcionales.																		
Análisis y estudio de la ontología para Ciencias de la Computación.																		
Análisis y estudio del lenguaje de consulta.																		
Análisis y estudio de la herramienta para el tratamiento de la ontología.																		
Análisis y estudio de la metodología para la ontología.																		
Búsquedas de herramientas disponibles para cumplir con la parte gráfica y con el requerimiento gráfico del proyecto.																		
Búsqueda de los antecedentes teóricos, tecnológicos y prácticos que presenta desarrollar la aplicación.																		
Buscar Información sobre R2O.																		
Probar R2O.																		
Se analizará y se escogerá de todas las alternativas posibles tanto en metodologías como en herramientas ya diseñadas y se empezara con gran parte de la implementación.																		
Concordancia e integración de la ontología con la BD y las instancias respectivas.																		
Definición de los casos de uso.																		

CRONOGRAMA DE ACTIVIDADES

PARTE 2

SEMANA DE LA 19 A LA 36

ACTIVIDADES	**TIEMPO EN SEMANAS**																	
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Definición preliminar del modelo conceptual.																		
Mejoramiento del modelo conceptual.																		
Definición de los diagramas de secuencia del sistema.																		
Definición de los casos de uso reales.																		
Definición de los diagramas de clases.																		
Definición del plan de pruebas.																		
Puesta en marcha del plan de pruebas																		
Terminación de la implementación y entrega																		

PRESUPUESTO

CONCEPTO	**APORTE DE UNIVALLE**	**APORTE PERSONAL**
PERSONAL		
Investigadores Principales		**$ 8.250.000**
Otros Investigadores		
Consultores y Asesores	**$ 1.350.000**	
Técnicos	**$ 1.000.000**	
Auxiliares		
Becarios y Monitores		
EQUIPO PERMANENTE Y DEVOLUTIVO		
Infraestructura Física		
Equipos de Laboratorio		
Equipos de Oficina		
Equipos de Cómputo	**$ 1.000.000**	
Libros y Revistas		**$ 100.000**

Fotocopias		**$ 80.000**
Otros		**$ 40.000**
ELEMENTOS DE CONSUMO		
Útiles de Escritorio y Papelería		**$150.00**
Formas Continuas		
Disquetes - CD's		**$ 30.000**
Material de Laboratorio		
Material de Impresión		**$ 50.000**
Otros		**$ 30000**
GASTOS GENERALES		
Viáticos		**$ 1.400.000**
Transporte		**$ 1.056.000**
Mantenimiento		
Suministros		
Gastos de Instalación		

Gastos de Capacitación		
Otros		
GASTOS GENERALES DE ADMINISTRACIÓN		
IMPREVISTOS		**$ 150.000**
SUB TOTAL	**$ 3.350.000**	**$ 11.336.000**
TOTAL	**$ 14.686.000**	

REFERENCIAS BIBLIOGRÁFICAS

[1**]** www.google.com
[2] www.w3c.org
[3] http://www.w3.org/People/Berners-Lee/
[4] S. Biagioni, J.L.Borbinha, R. Ferber, P. Hansen, S. Kapidakis, L. Kovacs, F. Roos, y A. M. Vercoustre. The ERCIM Technical Reference Digital Library. En *Second European Conference on Research and Advanced Technology for Digital Libraries*, September 1998.
[5] James Powell y Edward A. Fox. Multilingual Federated Searching Across Heterogeneous Collections. *D-Lib Magazine*, Septiembre 1998.
[6] Andreas Paepcke, Steve B. Cousins, Scott W. Hassan, Ketchpel Steven P, Martin Roscheisen, Héctor García Molina, y Terry Winograd. Using distributed objects for digital library interoperability. *Computer*, Mayo 1996.
[7] Gerard Rodríguez i Mulà. *Federation Mechanisms for Large Scale Cooperative Systems*. PhD thesis, Department of Computer Architecture. Universitat Politècnica de Catalunya.
[8] AQUARELLE. The Information Network on the Cultural Heritage. http://aqua.inria.fr/Aquarelle/.
[9] Nieves R. Brisaboa, M. José Durán, Charlo Lalín, Juan Ramón López, José R. Paramá, Miguel R. Penabad, y Ángeles S. Places. Propuesta de Arquitectura para un Sistema de Bases de Datos Documentales. En *IV Jornadas de Ingeniería del Software y Bases de Datos*, Noviembre 1999.
[10] Dinos Rojas J. "Arquitectura de un Sistema basado en Agentes para la Recuperación de Metadatos RDF en base a una Ontología de Documentos". Tesis maestría.
[11] http://www.interacciones.i-p.com/doc24.xhtml
[12] Peis Redondo E., Hassan Montero Y. "Ontologías, metadatos y agentes: recuperación de semántica de la información". Pág. 2-5.
[13] http://www.w3.org/XML/
[14] http://www.w3.org/XML/Schema
[15] http://www.w3.org/RDF/
[16] http://www.cyc.com/cycdoc/ref/cycl-syntax.html
[17] http://logic.stanford.edu/kif/kif.html
[18] http://elies.rediris.es/elies18/531.html
[19] Noy N. McGuinness D. "Ontology Development 101: A Guide to Creating your first Ontology".
[20] Corcho O., Gómez-Pérez A. "A Roadmap to Ontology Specification Languages". EKAW 2000 pp. 80-96
[21] Farquhar, A., Fikes, R., Rice, J. *The Ontolingua Server: A Tool for Collaborative Ontology Construction*. Proceedings of KAW96. Banff, Canada, 1996.
[22] http://kmi.open.ac.uk/projects/ocml/
[23] Kifer, M., Lausen, G., Wu, J. *Logical Foundations of Object-Oriented and Frame-Based Languages*. Journal of the ACM. 1995
[24] MacGregor, R. *Inside the LOOM clasifier*. SIGART bulletin. #2(3):70-76. June, 1991.
[25] Karp, R., Chaudhri, V., Thomere, J. *XOL: An XML-Based Ontology Exchange Language*. July, 1999.
[26] 15.Luke S., Heflin J. *SHOE 1.01. Proposed Specification*. SHOE Project. February, 2000.
[27] http://www.daml.org/
[28] http://www.w3.org/TR/2002/WD-owl-ref-20020729/

[29] http://www.w3c.es/Traducciones/es/SW/2005/owlfaq
[30] http://roda.ibit.org/herramientas.cfm
[31] Lozano Tello A. "Tesis Doctoral: Métrica de idoneidad de Ontologías". Universidad de Extremadura. Depto de Informatica. Escuela Politécnica de Caceres. Febrero 2002 pp. 31- 34
[32] http://www.hipertexto.info/documentos/web_semantica.htm
[Ben99] Benjamins R., Fensel D., Decker S. y Gomez Perez A., "(KA)2: Building Ontologies for the Internet: a Mit-Term Report", *International Journal of Human-Computer Studies*. Vol. 51.
[33] http://www.ndltd.org/
[34] Ryszard Kruk S., Decker S. y Zieborak L. "JeromeDL – Adding Semantic Web Technologies to Digital Libraries". Digital Enterprise Research Institute, National University of Ireland, Galway Ireland y Main Library, Gdansk University of Technology, Gdansk, Poland.
[Ber96] Bernaras A., Laresgoti I. y Corera J. "Building and Reusing Ontologies por Electrical Network Applications", *European Conference of Artificial Intelligence(ECAI ´96)* Budapest, Hungary, 1996
[Euz95] Euzenat J. "Building Consensual Knowledge Bases: Context and Architecture". *Toward Very Large Knowledge Bases. 2nd Conference on Building and Sharing Very Large-scale Knowledge Bases (KBKS)* The Netherlands. Abril 1995.
[Fer99b] Fernandez Lopez M. "Overview of Methodologies for Building Ontologies" *Proceedings of the Workshop on Ontologies and PSMs*. Sweden, Agosto 1999.
[Fer97] Fernandez M., Gomez Perez A. y Juristo N. "METHONTOLOGY: From Ontological Art Toward Ontological Engineering", *Spring Symposium Series on Ontological Engineering.* Stanford, California. Marzo 1997
[Gom97] Gomez-Perez A., Juristo N., Montes C. y Pazos J. "Ingenieria del Conocimiento". 1997.
[Gru95b] Gruninger M. y Fox M. "Methodology for the Design and Evaluation of Ontology". *Proceedings of the workshop on Basic Ontological Issues in Knowledge Sharing.* Montreal, Canada, 1995
[Gom98] Gomez-Perez A. "Knowledge Sharing and reuse" *the Handbook of Applied Expert Systems*. Londres 1998.
[Lem90] Lenat D. y Guha R. "Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project" USA 1990.
[Pressman]Ingeniería del software: un enfoque práctico. - 5ed.
[Swa97] Swartout B., Patil R., Knight K. y Russ T."Toward Distributed Use of Large-Scale Ontologies" *Spring Symposium Series on Ontological Engineering* 1997.
[Usc95] Uschold M. y King M."Towards and Methodology for Building Ontologies". *Proceeding of the workshop on Basic Ontological Issues in Knowledge Sharing.* Montreal, Canada, 1995
[Wat86] Waterman D. "A Guide to Experts Systems" USA 1986.
[Wie92] Wielinga B., Schreiber A. y Breuker J. "KADS: A modelling Approach to Knowledge Engineering " *Knowledge Acquisition* 1992