



**INSTITUTO POLITÉCNICO NACIONAL  
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**

---

**“APLICACIÓN DE LA TEORÍA DE CONFUSIÓN EN  
BÚSQUEDA DE INFORMACIÓN DE  
OFERTA EDUCATIVA”**

**TESIS**

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**

PRESENTA:

**IYELIZ REYES DE LOS SANTOS**

DIRECTOR DE TESIS:

**DR. SERGUEI LEVACHKINE**



México, D.F., Mayo 2011.



# INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

## ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 12:00 horas del día 11 del mes de Febrero de 2011 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

**Centro de Investigación en Computación**

para examinar la tesis titulada:

### "APLICACIÓN DE LA TEORÍA DE CONFUSIÓN EN BÚSQUEDA DE INFORMACIÓN DE OFERTA EDUCATIVA"

Presentada por el alumno:

<b>REYES</b> Apellido paterno	<b>DE LOS SANTOS</b> Apellido materno	<b>IYELIZ</b> Nombre(s)
----------------------------------	--	----------------------------

Con registro: 

B	0	8	1	4	3	1
---	---	---	---	---	---	---

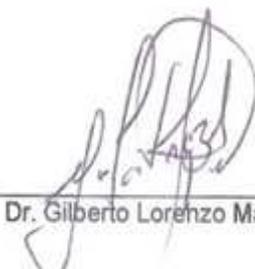
aspirante de: **MAESTRÍA EN CIENCIAS EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA

Directores de Tesis

  
Dr. Serguei Pavlovich Levashkine

  
Dr. Gilberto Lorenzo Martínez Luna

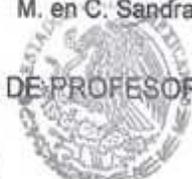
  
Dr. Marco Antonio Moreno Ibarra

  
Dr. Miguel Félix Mata Rivera

  
M. en C. Sandra Dinora Orantes Jiménez

PRESIDENTE DEL COLEGIO DE PROFESORES

  
Dr. Luis Alfonso Villalva

  
INSTITUTO POLITÉCNICO NACIONAL  
SECRETARÍA DE INVESTIGACIÓN  
EN COMPUTACIÓN  
DIRECCIÓN



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

**CARTA CESIÓN DE DERECHOS**

En la Ciudad de México, D.F. el día 04 del mes Mayo del año 2011, el (la) que suscribe Iyeliz Reyes de los Santos alumno (a) del Programa de Maestría en Ciencias de la Comp. con número de registro B081431, adscrito a Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. Serguei Pavlovich Levashkine y cede los derechos del trabajo intitulado "Aplicación de la teoría de Confusión en búsqueda de Información de Oferta Educativa", al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección iyelizrs@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Iyeliz Reyes de los Santos

Nombre y firma

# RESUMEN

En el presente trabajo se diseña una metodología para el procesamiento de consultas en una base de datos espacial dentro de un sistema orientado al usuario capaz de recomendar y recuperar información sobre un dominio geográfico, de acuerdo a preferencias; se pretende con esto personalizar Sistemas que permitan devolver resultados lo más exacto a lo que un usuario requiere, además de realizar recomendaciones a futuro después que el sistema haya sido actualizado, considerando siempre un perfil previamente definido. Para ello se requiere de un análisis tanto semántico como geográfico de los datos disponibles.

Las etapas de desarrollo propuestas para esta investigación son: Conceptualización, Modelado, Integración, Análisis, Resultados y Recomendaciones Futuras. La Conceptualización es la primera fase de la metodología en donde se investiga y entiende el problema, se hace una recopilación y análisis de datos necesarios, haciendo una búsqueda de fuentes que proporcionen información sobre el dominio de estudio y se examina la información para detectar heterogeneidad conceptual que impida la eficiente manipulación de la misma; Modelado se refiere al diseño de las estructuras que serán pilares de organización y almacenamiento de los datos; Integración es la fase en donde se conectan las estructuras del paso anterior y se comienzan a poblar con los datos recopilados; en el Análisis se procesa la información, consiste en el estudio de las preferencias del usuario, un análisis semántico y un análisis espacial a fin de ofrecer las mejores opciones al usuario, el caso de estudio sobre el que se aplicará la metodología será en la "Búsqueda del mejor lugar para tomar un curso de computación", posteriormente se presentan los resultados del procesamiento y por último se hacen las recomendaciones de cursos de computación al usuario.

# ABSTRACT

In this work I design a methodology for query processing in spatial database within a user-oriented information system, capable of recommending and retrieve searched information in a geographic domain, according to preferences. It is intended to customize the system to return results as accurate as to possible. In addition to do recommendations in the future after that the system has been updated, is necessary to consider a predefined profile. All those process requires a semantic analysis and a geographic analysis of available data.

The methodology I propose in this thesis is: Conceptualization, Modeling, Integration, Analysis, Results, Visualization and Future Recommendations. Conceptualization is the first step of the methodology where I search and understand the problem, in this step I recollect and analyze required data and I look for information about the study case, after that I examine the information for detecting heterogeneous data that impede an efficient manipulation of them. Modeling is referred to the design of structures which are pillars of organization and storing of data. Integration is the step where the structures are connected and the data are deposited in them. In Analysis's step the information is processed. Note that it is necessary to study the user preferences, carry out a semantic analysis and a spatial analysis in order to offer the better results to user. The study case is the "Searched about the best place where a user can take a computer course". Finally I present the results and the future recommendations to users which the system provides.

# AGRADECIMIENTOS

A ***Dios***, por darme la oportunidad y fuerza para conseguir este triunfo.

A ***Robert***, por su apoyo incondicional, amor y por ser mi fuente de inspiración.

A ***Mis Papás***, por su confianza y consejos.

Al ***Dr. Serguei Levachkine***, por compartir su conocimiento.

A los **Compañeros y Doctores del Laboratorio PIIG**, por su amistad y enseñanzas.

Al ***IPN y CONACYT***, por el apoyo administrativo, financiero y profesional.

Gracias a todos los que hicieron posible este logro en mi vida.

# ÍNDICE

---

RESUMEN .....	I
ABSTRACT .....	II
AGRADECIMIENTOS .....	III
ÍNDICE .....	IV
LISTA DE FIGURAS .....	VIII
LISTA DE TABLAS .....	IX
<b>Capítulo 1. Introducción .....</b>	<b>10</b>
1.1 Antecedentes .....	10
1.2 Planteamiento del Problema .....	11
1.3 Objetivos .....	12
1.3.1 Objetivo General .....	12
1.3.2 Objetivos Particulares .....	12
1.4 Alcances y Limitaciones .....	13
1.5 Motivación y Justificación .....	14
1.6 Organización de la Tesis .....	14
<b>Capítulo 2. Estado del Arte .....</b>	<b>16</b>
2.1 Sistemas de Información .....	16
2.1.1 Sistemas de Información Basados en Ontologías .....	17
2.1.2 Sistemas de Información Orientados al Usuario .....	18
2.1.2.1 Ejemplo de Sistemas de Información Orientados al Usuario .....	20
2.1.3 Sistemas Recomendadores de Información .....	20
2.1.3.1 ¿Cómo trabajan los Sistemas Recomendadores de Información? .....	21
2.1.3.2 Ejemplo de Sistemas Recomendadores de Información .....	22
2.1.4 Sistemas de Recuperación de Información .....	24
2.1.4.1 Modelos para la Recuperación de Información .....	25
2.1.4.2 Métodos de Recuperación de Información en la Web .....	26

2.1.5 Sistemas de Información Geográfica .....	27
2.1.5.1 Información Geoespacial y Objetos Geográficos .....	27
2.1.5.2 Métricas de distancia entre puntos geográficos .....	27
2.1.5.3 Sistemas de Recuperación de Información Geográfica .....	28
2.1.5.4 Sistemas de Información Geográfica basados en Ontologías .....	29
2.2 La Web Semántica .....	31
2.2.1 La Web Semántica Geoespacial .....	31
2.3 Sistemas actuales de Búsquedas de Cursos Computacionales .....	32
2.4 Resumen y Comentarios Finales .....	35
<b>Capítulo 3. Marco Teórico</b> .....	<b>36</b>
3.1 Formas tradicionales de representación del conocimiento especializado .....	36
3.2 Bases de Datos .....	36
3.3. Bases de Conocimiento .....	37
3.4 Diferencias entre Bases de datos y Bases de conocimiento .....	37
3.5 Bases de Datos Espaciales .....	39
3.6 Ontologías .....	40
3.6.1 Ejemplos de Ontologías .....	41
3.6.2 Tipos de Ontologías .....	41
3.6.3 Razones para desarrollar una ontología .....	42
3.6.4 Operaciones y Relaciones de una Ontología .....	43
3.6.5 Reglas fundamentales para el diseño de Ontologías .....	43
3.6.6 Enfoque iterativo para el diseño y mantenimiento de una Ontología .....	41
3.7 Jerarquías .....	46
3.7.1 Tipos de Jerarquías .....	46
3.7.1.1 Jerarquías Simples .....	46
3.7.1.2 Jerarquías Ordenadas .....	47
3.7.1.3 Jerarquías Porcentuales .....	48
3.7.1.4 Jerarquías Mixtas .....	48
3.8 Teoría de Confusión .....	48
3.8.1 Teoría de Confusión para Jerarquías Simples .....	49
3.8.2 Teoría de Confusión para Jerarquías Ordenadas .....	49
3.8.3 Teoría de Confusión para Jerarquías Porcentuales .....	50

3.8.4 Teoría de Confusión para Jerarquías Mixtas .....	50
3.9 Objetos idénticos, muy similares y algo similares .....	51
3.10 Resumen y Comentarios Finales .....	52
<b>Capítulo 4. Metodología</b> .....	<b>53</b>
4.1 Definiciones previas .....	54
4.2 Casos de Estudio .....	55
4.3 Conceptualización .....	56
4.4 Modelado .....	57
4.4.1 Diseño de la Ontología .....	57
4.4.1.1 Catálogo de áreas de conocimiento en Computación.....	57
4.4.2 Diseño de Base de Datos Espacial .....	63
4.4.2.1 Modelo Conceptual de la Base de Datos Espacial .....	63
4.4.3 Diseño de las Jerarquías de Propiedades .....	64
4.5 Integración.....	66
4.6 Análisis.....	69
4.6.1 Análisis Espacial.....	70
4.6.2 Análisis Semántico.....	70
4.6.2.1 Análisis Semántico de la Ontología .....	71
4.6.2.2 Análisis Semántico de las jerarquías de propiedades .....	71
4.7 Presentación de Resultados .....	72
4.8 Recomendaciones Futuras .....	72
4.9 Arquitectura de la Aplicación .....	72
4.10 Resumen y Comentarios Finales .....	75
<b>Capítulo 5. Pruebas y Resultados</b> .....	<b>76</b>
5.1 Interfaz principal de la aplicación .....	76
5.2 Pruebas a la ontología .....	77
5.3 Pruebas finales .....	79
<b>Capítulo 6. Conclusiones y Trabajo a Futuro</b> .....	<b>84</b>
6.1 Conclusiones .....	84
6.2 Propuestas de trabajo futuro .....	85

<b>Referencias .....</b>	<b>86</b>

# LISTA DE FIGURAS

---

Figura 2.1. Ejemplo 1 de Sistemas Recomendadores de Información .....	21
Figura 2.2. Ejemplo 2 de Sistemas Recomendadores de Información .....	21
Figura 2.3. Ejemplo 3 de Sistemas Recomendadores de Información .....	22
Figura 2.4. Expresión matemática de la distancia Euclidiana .....	26
Figura 2.5. Expresión matemática de la distancia Manhattan .....	26
Figura 2.6. Google. Buscador de Cursos Computacionales .....	31
Figura 2.7. EduPortal. Buscador de Cursos Computacionales .....	32
Figura 2.8. eMagister. Buscador de Cursos Computacionales .....	32
Figura 3.1. Jerarquía Top-Down .....	42
Figura 3.2. Jerarquía Button-Up .....	43
Figura 3.3. Ejemplo de una Jerarquía simple .....	44
Figura 3.4. Ejemplo de una Jerarquía ordenada .....	45
Figura 3.5. Ejemplo de una Jerarquía porcentual .....	46
Figura 4.1. Metodología propuesta .....	51
Figura 4.2. Parte I de la estructura de la Ontología (áreas y subáreas) .....	59
Figura 4.3. Parte II de la estructura de la Ontología .....	60
Figura 4.4. Parte III de la estructura de la Ontología .....	60
Figura 4.5. Modelo conceptual de la base de datos espacial .....	61
Figura 4.6. Jerarquía de Costos .....	63
Figura 4.7. Jerarquía de Modalidad .....	64
Figura 4.8. Jerarquía de Distancia .....	65
Figura 4.9. Jerarquía de Horario .....	66
Figura 4.10. Conexión Ontología-BDE .....	67
Figura 4.11. Arquitectura de la Aplicación .....	72
Figura 5.1 Pantalla principal del sistema .....	75

# LISTA DE TABLAS

---

Tabla 2.1. Modelos para la recuperación de Información .....	26
Tabla 5.1 Resultados de la Prueba a la ontología de algunas áreas y subáreas ...	78
Tabla 5.2 Resultados de la Prueba 1, al consultar un curso de Java.....	79
Tabla 5.3 Resultados de la Prueba 2, al consultar un curso de SQL Server .....	79
Tabla 5.4. Resultados de la prueba final .....	83

# CAPÍTULO 1.

# INTRODUCCIÓN

En esta sección se hace una presentación formal de la tesis, se expone en breve, los antecedentes del tema de investigación, el planteamiento del problema explica la problemática existente, se manifiesta el objetivo de la investigación, las razones que motivaron a realizarla y los fundamentos que la sustentan.

## 1.1. Antecedentes

La Ontología es una disciplina filosófica que se encarga de definir el ser de las cosas a partir del estudio de sus propiedades, estructuras y Sistemas; sin embargo, en las últimas décadas, éste término se ha utilizado en el área de los Sistemas de información para referirse a una entidad computacional creada para la especificación y estructuración del conocimiento, estas entidades resuelven los problemas de comunicación por falta de entendimiento compartido cuando los desarrollos de Sistemas de información se hacen en diferentes contextos y con distintos puntos de vista acerca de los datos, información o el dominio de estudio, además evitan respuestas nulas cuando se consulta bases de conocimientos.

Las ontologías se pueden encontrar en los Sistemas de información geográfica (GIS) representando, almacenando y organizando datos espaciales y geográficos, además de permitir la definición de todos los componentes del sistema.

Por otro lado, en los últimos años, el interés y estudio de los Sistemas recomendadores de información ha crecido dramáticamente, debido primordialmente a la demanda de aplicaciones personalizadas en Internet, la mayoría de los sitios Web que ofrecen comercio en línea usan Sistemas recomendadores para personalizar sus contenidos y objetivos de venta; este tipo de Sistemas los podemos usar también en

herramientas que permitan sugerir puntos geográficos de interés cultural y educativo, pocas son las aplicaciones de este tipo hoy en día.

La vinculación de ontologías con los Sistemas recomendadores es cada vez más estrecha ya que muchos han sido los beneficios que éstas aportan a este tipo de Sistemas.

## **1.2. Planteamiento del Problema**

Uno de los problemas a los que se enfrentan diariamente los usuarios de la Web como herramienta de búsqueda es con la inmensa cantidad de información disponible, lo que hace invertir demasiado tiempo para rastrear entre el conjunto de fuentes de información hasta que se consigue lo que se busca, si es que finalmente se llega a ella.

En las últimas décadas se han implementado Sistemas informáticos denominados motores de búsqueda que indexan archivos almacenados en servidores Web. Funcionan utilizando palabras claves y mediante programas llamados robots o arañas que rastrean la red y leen tantas páginas como puedan buscando coincidencias de las palabras con el contenido de la página. El resultado es un listado de direcciones de sitios en Internet en los que se mencionan temas relacionados con las palabras clave buscadas. Su rapidez y funcionalidad ha elevado el número de personas que utilizan esta tecnología, tanto como herramienta de búsqueda o como enlace para sus sitios.

Con el creciente uso de estos buscadores, el volumen de información que ofrecen ha aumentado sustancialmente y su organización es cada vez más compleja; esto genera en muchas ocasiones respuestas que son irrelevantes al tema buscado o en el peor de los casos respuestas vacías y entonces se procede a inspeccionar de uno en uno los sitios Web hasta encontrar la información que se precisa. Los principales afectados de esta problemática son los usuarios que buscan información sobre un tópico especializado o sobre información local, ya que existen muchos términos que tienen múltiples significados para diversos sectores o disciplinas y que al usarlos como palabras claves para una consulta, el resultado es una enorme cantidad de coincidencias sintácticas en diferentes sitios con temas distintos.

Organismos que ofrecen sus servicios en Internet o que disponen de ventas en línea requieren de un motor de búsqueda para mostrarle al cliente lo que tiene respecto a lo que está buscando. El mejor de los escenarios sería aquel en donde el proveedor pudiera darle al cliente lo que busca exactamente o los más parecido a ello.

Expuesto lo anterior para mejorar las búsquedas en Internet los sitios deben usar repositorios que permitan organizar la información, describir las relaciones de los datos con respecto a otros y definir sus propiedades, logrando con ello una integración eficiente del conocimiento en un dominio determinado. Siempre será más factible si la información presenta homogeneidad semántica entre sus conceptos.

## **1.3. Objetivos**

En este apartado se detallan los propósitos tanto generales como particulares del presente trabajo.

### **1.3.1. Objetivo General**

- Definir una metodología para el procesamiento de consultas en una base de datos espacial dentro de un sistema orientado al usuario capaz de recomendar y recuperar información sobre un dominio geográfico.

### **1.3.2. Objetivos Particulares**

- Desarrollar un sistema orientada al usuario para recuperar información en una base de datos espacial y que permita obtener resultados exactos o aproximados sobre un dominio determinado.

- Implementar la metodología de procesamiento de consultas en una base de datos espacial para un sistema que recomiende el mejor lugar para tomar un curso de computación y permita recomendaciones futuras.
- Definir y Diseñar estructuras que organicen y almacenen la información del dominio planteado.
- Referenciar la información para efectuar el análisis geográfico.
- Crear una ontología sobre las áreas de computación.
- Contar con un repositorio para que instituciones educativas puedan registrar sus ofertas de cursos computacionales.

## **1.4. Alcances y Limitaciones**

Dentro de los alcances de la tesis se encuentra crear una base de datos conceptual que jerarquice las áreas de computación para vincularlas con cursos computacionales, actualmente no existe un procedimiento que organice este conocimiento.

En las limitaciones se especifica que el caso de estudio está restringido a instituciones educativas pertenecientes a la delegación Gustavo A. Madero del Distrito Federal, México, para pruebas de la metodología.

El principal requerimiento de entrada para la consulta será el perfil del usuario definido con anterioridad, sin el cual la consulta no podrá ser ejecutada.

No se utilizan valores de ponderación para determinar el nivel de importancia para el usuario de cada una de las propiedades de los cursos (distancia, costo, modalidad, y horario), al usuario se le ofrecen las mejores opciones que el sistema determina.

Para la medida de similitud entre conceptos, parte esencial de este trabajo, estará determinada por la función *confusión* [Levachkine, et al., 2007].

## 1.5. Motivación y Justificación

En el presente apartado se expone las problemáticas que motivaron al desarrollo de este trabajo de investigación.

Hoy en día los Sistemas informáticos que soportan consultas por parte de un usuario están siendo orientados a la aplicación de mejores técnicas que satisfagan los requerimientos del mismo. Para lograr esto, muchas aplicaciones requieren que el usuario envíe al sistema un formulario con información sobre lo requerido y una vez que el sistema devuelve los resultados este borra de su memoria lo que el consultante le proporcionó, por lo que la próxima vez que haga otra consulta deberá volver a enviar el formulario para encontrar lo que busca. Dado esto, se hace necesario buscar métodos que además de ofrecer al usuario los mejores resultados permitan almacenar sus preferencias de búsqueda consiguiendo con esto Sistemas más personalizados en donde los usuarios tengan la menor interacción con el sistema y ahorren tiempo pero siempre satisfaciendo sus necesidades.

Por otro lado los organismos que ofrecen sus servicios y/o productos en sitios de Internet están siempre en la búsqueda de clientes potenciales, en este tema la componente geográfica puede jugar un papel importante para saber a quién y en donde vender u ofrecer. Pocos son los sitios Web donde se considera distancia y tiempo de traslado desde la ubicación del cliente hasta donde se ofrece el servicio o se vende un producto.

Expuesto lo anterior y la problemática planteada se justifica y motiva la creación de una metodología que proporcione mecanismos para crear un sistema de búsqueda considerando las preferencias de un usuario, una componente espacial para que además se consideren las ubicaciones geográficas con el fin de determinar distancias y ahorro de tiempo de traslado. El perfil del usuario permite hacer recomendaciones en el futuro, evitar repetir procesos durante la búsqueda y que los resultados devueltos sean lo más exacto o parecido a los que el solicitante requiere.

## 1.6. Organización de la tesis

**Capítulo 1.** En este capítulo se introduce a la problemática analizada y se da una visión general de lo que se presenta en la tesis; en este apartado también se plantean los objetivos, los cuales especifican lo que se desea lograr con la investigación y se exponen las razones que motivaron y justifican el presente trabajo, además de los alcances y limitaciones del mismo.

**Capítulo 2.** En el transcurso de este capítulo se da una panorámica del estado de arte referente a los Sistemas recomendadores de información, los Sistemas orientados al usuario y los Sistemas de información geográficos. Por último, se presentan una breve descripción de algunos artículos que tienen relación con la tesis.

**Capítulo 3.** En este capítulo se describe el marco teórico de la tesis, organizando los argumentos teóricos y referenciales del problema de investigación que dan sustento al trabajo. Se estudian las estructuras de representación del conocimiento, como son las bases de datos relacionales, las bases de datos espaciales y las bases de conocimiento, se hace una clara distinción entre ellas y los beneficios de vincularlas para cumplir con el objetivo de la tesis.

**Capítulo 4.** En este capítulo se describe la metodología propuesta, la cual da solución a la problemática planteada en el capítulo 1, se definen las estructuras que organizan el conocimiento y la información del área de aplicación de la tesis.

**Capítulo 5.** Este capítulo describe las pruebas y se exponen los resultados obtenidos con la metodología propuesta.

**Capítulo 6.** En este capítulo se describen las conclusiones de los resultados adquiridos con el presente trabajo de investigación, además se proponen temas y trabajos a futuro para darle seguimiento y optimización de ésta tesis.

Por último se presentan las referencias a los trabajos que sustentaron el trabajo de investigación.

## Capítulo 2.

# ESTADO DEL ARTE

En el siguiente capítulo se da una visión del estado de arte referente a los Sistemas recomendadores de información, los Sistemas orientados al usuario y los Sistemas de información geográficos, se presentan ejemplo ellos y por último, se exponen algunos artículos que tienen relación y sustentan la tesis.

## 2.1 Sistemas de Información

Un sistema de información es un conjunto de elementos (personas, datos, técnicas de trabajo y recursos materiales) orientados al tratamiento y administración de datos e información, organizada, almacenada y preparada para su uso posterior, generados para cubrir un objetivo.

El término Sistemas de Información hace referencia a un concepto genérico que tiene diferentes significados dependiendo del campo de conocimiento al que se aplique dicho concepto (Informática, Teoría de Sistemas, Seguridad Computacional, Geografía y Cartografía, Representación del Conocimiento, Sociología, etc.). Para el presente trabajo nos enfocamos en las áreas de Informática, Representación del Conocimiento, Geografía y Cartografía.

Para el área de **Informática** un sistema de información es cualquier sistema o subsistema de equipo de telecomunicaciones o computacional interconectados y que se utiliza para obtener, almacenar, manipular, administrar, mover, controlar, desplegar,

intercambiar, transmitir o recibir voz y/o datos, e incluye tanto los programas de computación como el equipo de cómputo.

Para el área de ***Representación del Conocimiento***, un sistema de información consiste de tres componentes: humano, tecnológico y organizacional. La información se define en términos de tres niveles de semiótica. Datos que pueden ser procesados automáticamente por un sistema de aplicaciones corresponden al nivel de sintaxis. En el contexto de un individuo que interpreta los datos, estos son convertidos en información, lo que corresponde al nivel semántico. La información se convierte en conocimiento cuando un individuo conoce, entiende y evalúa la información, esto corresponde al nivel pragmático.

Para el área de ***Geografía y Cartografía*** un Sistema de Información se utiliza para integrar, almacenar, editar, analizar, compartir y desplegar información georeferenciada.

Muchas son las aplicaciones de los Sistemas de Información en estas tres áreas, amplias son las técnicas y recursos que se utilizan para lograr sus objetivos y enfrentar las limitantes que se presenten. La categorización de estos Sistemas se puede definir de acuerdo a sus objetivos finales, aplicaciones y técnicas de desarrollo. En los siguientes apartados de este capítulo se describen algunos tipos de estos Sistemas que son de interés para el trabajo de investigación.

### **2.1.1 Sistemas de Información Basados en Ontologías**

Los Sistemas de Información se implementan en diferentes contextos, con distintos puntos de vista y suposiciones acerca del dominio de estudio. Esto provoca problemas de comunicación por falta de entendimiento compartido.

Los desarrolladores de estos Sistemas necesitan usar representaciones, tan generales como sea posible, para responder a los atributos de calidad del software y aumentar al

máximo la posibilidad de reusabilidad. Pero, al mismo tiempo, las representaciones deben corresponderse, lo más estrechamente posible, con las cosas y los procesos que ellos representan. Estas cuestiones relacionadas con la gestión de datos información–conocimiento son cuestiones efectivamente ontológicas.

La creación e implementación de ontologías, en el desarrollo y el uso de los Sistemas de Información, nos lleva al concepto de los Sistemas de Información Basados en Ontologías.

En [Barchini, et al., 2006] se explican los objetivos detallados de los Sistemas de Información Basados en Ontologías, dentro de los que destacan los siguientes:

- a) Alcanzar un mayor entendimiento de los modelos y paradigmas informáticos mediante la utilización de modelos ontológicos.
- b) Mejorar la calidad del proceso y del producto software con el uso de ontologías.
- c) Facilitar el modelado, el desarrollo, el mantenimiento y la reutilización de los Sistemas de Información.
- d) Conceptualizar y lograr un entendimiento compartido de distintas áreas del conocimiento de la Informática Teórica y de la Informática Aplicada.
- e) Reutilizar el conocimiento existente en dominios específicos.
- f) Reducir los problemas de comunicación entre los desarrolladores de software y los usuarios/clientes de los Sistemas de Información.

### **2.1.2 Sistemas de Información Orientados al Usuario**

Los Sistemas orientados al usuario son Sistemas de información que basan sus respuestas y resultados en las preferencias y necesidades de los usuarios finales del mismo. Para lograr sus objetivos requieren conocer acerca de los usuarios, sus

intereses, necesidades, ambiente donde se desenvuelven, etc.; capturando este conocimiento de forma directa o indirecta. Estas preferencias son registradas en perfiles, para tal fin existen dos formas o métodos en la recolección de características (implícitas o explícitas).

***Algunos ejemplos de recolección de datos de forma explícita son:***

- Solicitar al usuario que pondere en base a una escala proporcionada, algún tema en particular.
- Solicitar al usuario que pondere un conjunto de temas de una lista de temas favoritos.
- Presentar al usuario dos temas, y solicitarle que seleccione uno de ellos.
- Solicitar al usuario que cree una lista de temas de su preferencia.

***Algunos ejemplos de recolección de datos de forma implícita son:***

- Guardar un registro de los temas que el usuario ha visto en una tienda online.
- Analizar el número de visitas que recibe un artículo
- Guardar un registro de los artículos que el usuario ha seleccionado.
- Obtener un listado de los artículos que el usuario ha seleccionado o visto en su computadora.
- Analizar las redes sociales de las que el usuario forma parte y de esta manera conocer sus gustos y preferencias.

Capturar las preferencias de los usuarios es una tarea problemática. Simplemente preguntar al usuario qué es lo que quieren es indiscreto y propenso a error, por otro lado, monitorear el conocimiento discretamente y encontrar patrones significativos es difícil y conlleva un alto consumo de tiempo de procesamiento [Edward, 2003].

### **2.1.2.1 Ejemplo de Sistemas de Información Orientados al Usuario**

En [Castillo et al., 2008] se presenta SAMAP, proyecto cuyo objetivo es construir un software que ayude a gente a visitar diferentes ciudades. Esta herramienta integra módulos que dinámicamente captura modelos de usuarios, determina una lista de actividades que pueden ser de utilidad a los usuarios, mediante experiencias pasadas proporcionadas al sistema por usuarios similares y genera planes que pueden ser ejecutadas por él. Dentro de su arquitectura utiliza una ontología para almacenar toda la información que el sistema necesita.

En [Ibarra, 2009] se describe el desarrollo de una herramienta que sugiere los inmuebles potenciales para colocar un negocio considerando las propiedades del espacio geográfico así como los intereses y preferencias de los usuarios. Responde a preguntas como: ¿Dónde se encuentra el inmueble más adecuado para establecer un nuevo negocio o sucursal de acuerdo a mi perfil? y ¿Qué características descriptivas y geográficas debe tener ese inmueble?.

### **2.1.3 Sistemas Recomendadores de Información**

Los Sistemas recomendadores son un intento de modelar matemáticamente y reproducir técnicamente los procesos de recomendación del mundo real. Captan opiniones de usuarios, acerca de productos, sitios Web, personas, servicios etc. clasifican tales opiniones y las usan para sugerir nuevos elementos o para predecir la utilidad de un elemento para un usuario particular, basado en las preferencias previas de los usuarios y/o en las opiniones de otros usuarios que tiene preferencias similares.

Generalmente comparan el perfil del usuario con algunas características de referencias de los temas y busca predecir el valor o ponderación que el usuario le daría a un elemento que aún el sistema no ha considerado. Estos Sistemas pueden basarse también en la relación o acercamiento del usuario con el tema o el ambiente social del mismo.

En términos generales los Sistemas recomendadores actúan como un vendedor experto en línea que interactúa con los clientes, aprende de sus preferencias y responde de manera congruente con información relevante y recomendaciones personalizadas apoyadas con contenido rico, original y tradicional [Delgado et al, 2002].

### 2.1.3.1 ¿Cómo trabajan los Sistema Recomendadores de Información?

[Delgado et al, 2002] expone que los Sistemas Recomendadores trabajan con agentes de recomendación y los clasifica en términos de sus técnicas de filtro.

1. Filtro Basado en Características (también llamados Filtros Basados en Contenido): las recomendaciones dadas al usuario son a partir de la similitud de características de los elementos recomendados. Formalmente, se describe a  $u(c, s_i)$  como la utilidad de un elemento  $s$  para el usuario  $c$ , y se estima con base en las utilidades  $u(c, s_i)$  asignadas por el usuario  $c$  a los elementos  $s_i \in S$  que son "similares" al elemento  $s$ . Por ejemplo, en una aplicación que recomienda películas, para recomendar una película al usuario  $c$ , éste método evalúa las similitudes de características de las películas (actores, género, director, tema, etc.) con las preferidas de  $c$ , las cuales fueron proporcionadas por  $c$  con anterioridad. Este enfoque tiene su origen en las investigaciones sobre Recuperación de Información y Filtro de Información
2. Filtro de Colaboración Automática (ACF): Basa las recomendaciones en las preferencias de otros usuarios. Formalmente la utilidad  $u(c, s)$  de un elemento  $s$  para un usuario  $c$  es estimada con base a las utilidades  $u(c_j, s)$  asignadas al elemento  $s$  por otros usuarios  $c_j \in C$  quienes son similares al usuario  $c$ . Por ejemplo en una aplicación que recomienda películas éste método trata de encontrar un "par" al usuario  $c$ , esto es, otros usuarios que tengan gustos

similares en películas. Entonces sólo las películas que son preferidas de los usuarios "par" son recomendadas.

3. Filtro Basado en Restricciones (CSP). Basan sus recomendaciones en las restricciones o limitantes de los elementos.

Por último [Adomavicius, et al, 2005] detalla que puede existir un método híbrido, que combine los tres anteriores, que enfrente las limitaciones de uno y considere las ventajas del otro.

Tradicionalmente, los Sistemas recomendadores son referidos como Sistemas de filtro colaborativos quienes asisten y aumentan las transferencias de recomendaciones entre miembros de una comunidad. Un sistema común recolecta las preferencias y opiniones de usuarios particulares, después agrega y transfiere esas recomendaciones a otros miembros de la comunidad.

Estos Sistemas inicialmente acumulaban opciones o preferencias de los usuarios sobre productos, que eran dirigidas luego a los receptores adecuados, pero ahora además tienen en cuenta las características de quienes reciben la recomendación. Los Sistemas recomendadores nacieron de los Sistemas de recuperación de información y los motores de búsqueda pero se diferencian de estos en su personalización [Saboya, 2005].

### 2.1.3.2 Ejemplos de Sistemas Recomendadores de Información

- **Amazon.com** es una página Web de compra por Internet, incluye recomendaciones de productos de acuerdo a las preferencias de los usuarios.

Cuando se accede a la página de Amazon ofrece además de ofertas especiales y una amplia gama de productos, algunas recomendaciones que van dirigidas expresamente al que les visita. Una vez que uno se registra en Amazon con su

nombre y algunos datos, el visitante es tratado de forma personal como si tuviera su propio vendedor asignado.

El seguimiento de clientes es uno de los puntos fuertes de Amazon, el cliente puede recibir consejos, como por ejemplo recomendaciones basadas en anteriores compras y listas de revisiones y guías de usuarios que ya han comprado el producto que anda buscando.



Figura 2.1. Ejemplo 1 de Sistemas Recomendadores de Información

- En [Aqualimpia, 2009] se presenta **GEOSESYMU** (Gespatial Semantic System to Retrieve Information for Museums) cuya meta es la de recomendar diferentes Puntos de Interés Cultural, en particular los Museos en el centro histórico de la ciudad de México, considerando la ubicación geográfica del usuario y algunos criterios de preferencia como son: similitud, densidad cultural y distancia, permitiendo a los usuarios la ponderación de cada uno de ellos.
- **Daily me** es un sistema de recomendación de noticias fundado por Nicholas Negroponte en Estados Unidos, donde el usuario puede personalizar las noticias que desea recibir a través de perfiles. Sólo tiene que indicar qué temas le interesan y tendrá el periódico personalizado en la página. Un diario virtual personalizado basado en los gustos de cada individuo, donde también el usuario puede rastrear las noticias clasificadas en categorías como en los comunes portales de noticias. Sus artículos son publicados en inglés.



Figura 2.2. Ejemplo 2 de Sistemas Recomendadores de Información

- **Pandora** es un buscador y analizador de música donde el usuario puede encontrar canciones similares en características a las de su preferencia. Se estudia y recopila detalles musicales de una canción, reconociendo propiedades como melodía, armonía, instrumentación, voces, ritmos, letras, etc. Además de poder escuchar música se pueden encontrar artistas nuevos que se adapten a los gustos de los usuarios. Solo funciona en Estados Unidos.



Figura 2.3. Ejemplo 3 de Sistemas Recomendadores de Información

## 2.1.4 Sistemas de Recuperación de Información

Según el diccionario Mac Millan de Tecnología de la Información, Recuperación de Información se refiere a las técnicas empleadas para almacenar y buscar grandes cantidades de datos y ponerlos a disposición de los usuarios.

Un Sistema de Recuperación de Información puede ser descrito como un conjunto de ítem de información (documentos), un conjunto de peticiones y algún mecanismo que determine qué ítem satisface las necesidades de información expresadas por el usuario en la petición. Sus principales funciones son:

- 1) Identificar fuentes de información relevantes a las áreas de interés de las solicitudes de los usuarios.
- 2) Analizar los contenidos de los documentos.
- 3) Representar los contenidos de alguna forma que puedan ser comparables con las preguntas de los usuarios.
- 4) Analizar las preguntas de los usuarios y representarlas de alguna forma que sea adecuada para compararlas con las representaciones de los documentos contenidos en su base de datos.

- 5) Realizar la correspondencia entre la representación de la búsqueda y los documentos almacenados en la base de datos.
- 6) Recueperar la información relevante.
- 7) Realizar los ajustes necesarios en el sistema basados en la retroalimentación con los usuarios.

La evolución de estos Sistemas está orientada a obtener datos relevantes, precisos y exactos que satisfagan lo que un usuario u otro sistema requiere. Los Sistemas de recuperación involucran también un subsistema de búsqueda. La búsqueda de la información puede darse siguiendo dos procesos, el de interrogación (queryng) ó mediante el proceso de exploración (browsing); donde en el primero el usuario introduce palabras clave que representan sus necesidades de información, para posteriormente el sistema retorne una serie de respuestas, generalmente ordenadas por relevancia; y en el segundo, el usuario se encarga de explorar visualmente sin tener que expresar previamente cuáles son sus necesidades de información.

#### **2.1.4.1 Modelos para la Recuperación de Información**

El diseño de un sistema de recuperación de información se realiza bajo un modelo, quién establece cómo se obtienen las representaciones de los documentos y de la consulta, la estrategia para evaluar la relevancia de un documento respecto a una consulta, los métodos para establecer el orden de los documentos de salida y los mecanismos que permiten una realimentación por parte del usuario para mejorar la consulta.

Modelo	Descripción
Modelos clásicos	Se basan en el modelo booleano, espacio vectorial y probabilístico.
Modelos alternativos	Están basados en la lógica fuzzy.
Modelos lógicos	Basados en la lógica formal.
Modelos basados en la interactividad	Incluyen posibilidades de expansión del alcance de la búsqueda y hacen uso de la retroalimentación por la relevancia de los documentos recuperados.
Modelos basados en la Inteligencia Artificial	Se basan en redes neuronales, algoritmos genéticos, bases de conocimiento y procesamiento del lenguaje natural.

Tabla 2.1. Modelos para la recuperación de Información

### 2.1.4.2 Métodos de Recuperación de Información en la Web

Hoy en día existen tres mecanismos para recuperar información en la Web.

- 1) **Motores de búsqueda:** Indexan una porción de los documentos residentes en la globalidad de la Web y permiten localizar información a través de la formulación de una pregunta. Son grandes bases de datos de referencias a páginas Web recopiladas automáticamente, sin intervención humana.
- 2) **Directorios:** Sistemas que clasifican documentos Web seleccionados por materia y que nos permiten navegar por sus selecciones o buscar en sus índices. Son aplicaciones controladas por humanos que manejan grandes bases de datos con direcciones de páginas, títulos, descripciones, etc.
- 3) **Por explotación de su estructura hipertextual:** se refiere a la búsqueda por explotación de los enlaces recogidos en las páginas Web, incluye los leguajes de consulta a la Web y la búsqueda dinámica. La principal ventaja de este modelo es la de descubrir información relevante siguiendo los enlaces de las páginas recuperadas.

## 2.1.5 Sistemas de Información Geográfica

Es un sistema de información dedicado a la captura, modelado, almacenamiento, manipulación, integración, análisis y presentación de datos geográficamente referenciados. Existen muchas aplicaciones de estos Sistemas, desde ecología y geología, hasta las ciencias sociales.

### 2.1.5.1 Información geoespacial y objetos geográficos

La información geoespacial involucra datos de fenómenos ocurridos por encima, sobre y debajo de la superficie terrestre y la cual es representada en mapas. Un mapa contiene objetos geográficos, cualquier objeto geográfico tiene dos componentes, atributo espacial o geométrico y el atributo descriptivo, el primero se refiere a la componente que describe la ubicación, forma, orientación y tamaño del objeto, el segundo, a todos las demás propiedades que detallan dicho objeto [Rigaux, et al., 2002].

### 2.1.5.2 Métricas de distancia entre puntos geográficos

La distancia entre dos puntos geográficos puede ser medida de diversas formas, a continuación describo dos de ellas, la distancia Euclidiana y distancia Manhattan.

- ***Distancia Euclidiana:*** es una medida directa entre dos puntos ubicados en el mismo espacio. Considera que la distancia mas corta entre dos puntos es la recta que los une. Estos puntos pueden estar representados en un espacio de "n" dimensiones. Formalmente, la distancia Euclidiana puede ser matemáticamente expresada como sigue:

$$dE = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

Figura 2.4. Expresión matemática de la distancia Euclidiana.

Donde  $a$  y  $b$  son dos puntos en el mismo espacio de "n" dimensiones.

- **Distancia de Manhattan:** considera que la distancia entre dos puntos no es la recta entre dos puntos, sino el mínimo número de calles que se deben recorrer. Es descrita de la siguiente forma:

$$dM = |x_1 - x_2| + |y_1 - y_2|$$

Figura 2.5. Expresión matemática de la distancia Manhattan.

Donde los A y B son los siguientes puntos  $(x_1, y_1)$  y  $(x_2, y_2)$  respectivamente.

La distancia euclidiana es la componente espacial que se usará para el análisis geográfico de esta tesis, servirá para ofrecer mejores resultados.

### 2.1.5.3 Sistemas de Recuperación de Información Geográfica

La Recuperación de Información Geográfica (GIR, por su siglas en inglés) pertenece a la rama especializada de la recuperación de información. Incluye todas las áreas de investigación que tradicionalmente forman el núcleo de la Recuperación de Información, pero además con un énfasis en la información geográfica y espacial.

La GIR se preocupa de la recuperación de información que involucra algún tipo de percepción espacial [Perea, et al., 2008].

Un sistema de recuperación de información geográfica se encarga de recuperar documentos o datos relevantes en respuesta a una consulta con el formato <tema, localización>, donde la relación espacial puede implicar implícitamente contenido, o explícitamente seleccionado de un conjunto de posibles opciones topológicas, direccionales o de proximidad [Bucher, et al., 2005].

El éxito de la recuperación se basa en la integración de tres criterios de búsqueda para tres fuentes de datos heterogéneas: Ontologías Geográficas, diccionarios de datos y archivos vectoriales. Los cuales describen a los objetos utilizando la topología, los atributos geográficos y la semántica espacial [Mata, 2009].

La arquitectura de cualquier sistema GIR parte de un modelo básico de recuperación de información. Por lo tanto el elemento esencial es su motor de búsqueda.

En [Mata, 2009] expone una metodología para la recuperación de información geográfica y un método para ponderar los resultados de la recuperación. Para ejecutar consultas espaciales, esta metodología las define mediante una tripleta <O, R, L> donde O es un objeto geográfico, L es una referencia geográfica y R es la relación espacial entre O y L. El resultado de la búsqueda es el conjunto de documentos con datos geográficos, que tiene la mayor similitud, de acuerdo a las características evaluadas.

#### **2.1.5.4 Sistemas de Información Geográfica basados en Ontologías**

El diseño e implementación de las ontologías también se ha extendido a los Sistemas de Información Geográfica.

En [Fonseca, et al., 2002] se expone la creación de una metodología basada en ontologías para Sistemas de información geográfica basada en el universo físico y matemático, procediendo a la representación y a la implementación, también en este

trabajo se define el término ODGIS (Ontology Driven Geographic Information Systems). Este paradigma fue bien aceptada dentro de la comunidad de científicos ya que permite el manejo de mapas geográficos temáticos, además de tener descripciones de sensores remotos, representación de imágenes y la clasificación de imágenes.

En [Gómez H.F., 2003 ] se desarrolló una ontología para Sistemas de información geográfica que permite la representación de mapas, vías de comunicación, situación hidrológica, tipos de vegetación, además de probar su consistencia y taxonomía.

Crear una ontología para un Sistema de Información Geográfica es vital, ya que nos permite observar de forma clara sus componentes, la forma como se interrelacionan y como se dispara una regla de conocimiento que permite clasificar los objetos geográficos y espaciales [Gómez H.F., 2003].

## **2.2 La Web Semántica**

Se trata de una corriente, promovida por el inventor de la Web y presidente del consorcio W3C12, Tim Berners\_Lee, cuyo fin es lograr que las computadoras puedan entender, y por tanto utilizar, lo que la Web contiene. Esta nueva Web está poblada por agentes o representantes software capaces de navegar y realizar operaciones por el ser humano para ahorrar tiempo de trabajo y optimizar los resultados. Para conseguir esta meta, la Web semántica propone describir los recursos de la Web con representaciones procesables y entendibles no sólo por personas, sino por programas que puedan asistir, representar, o reemplazar a las personas en tareas rutinarias o inabarcables para un humano.

Las tecnologías de la Web semántica buscan desarrollar una Web más cohesionada, donde sea aún más fácil localizar, compartir e integrar información y servicios, para sacar un partido todavía mayor de los recursos disponibles en la Web [Castell, 2004].

La Web semántica mantiene los principios que han hecho un éxito de la Web actual, como son los principios de descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución, o la apertura al crecimiento y uso no previstos de antemano. En este contexto un problema clave es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la Web: usuarios, desarrolladores y programas de muy diverso perfil. La Web semántica rescata la noción de ontología del campo de la Inteligencia Artificial como herramienta para cumplir este objetivo.

### **2.2.1 La Web Semántica Geoespacial**

Extensión de la Web Semántica que intenta incorporar semántica Geoespacial a la Web tradicional. La idea principal es poder elaborar ambientes donde puedan realizarse búsquedas para ubicar lugares geográficos dejando atrás las comparaciones léxicas es decir, eliminando el uso etiquetas o palabras clave, dentro de los propósitos se describen los siguientes:

- Mejorar la representación de la información geográfica.
- Optimizar la integración de información.
- Añadir consultas que usen operadores espaciales.

Existen tres componentes que describen a la Web Semántica Geoespacial, los cuales se pueden agrupar de la siguiente manera:

- Componente geoespacial: Mapas, Objetos geográficos, Relaciones geográficas y Sistemas de referencia.
- Componente Web: Interoperabilidad entre tecnologías y Servicios compartidos.
- Componente semántica: Razonamiento automático e inferencia.

En función de estos componentes es posible definir algunas de las aplicaciones con mayor necesidad actualmente, estas son:

- Plataformas de búsqueda empleando componentes de tipo geográfico.
- Descubrimiento de conocimiento a través de relaciones semánticas, espaciales y temporales.
- Seguimiento al comportamiento de los usuarios de la Web Geoespacial.

El UCGIS (University Consortium for Geographic Information) está profundizando en la investigación de la Web semántica geoespacial. El UCGIS ha identificado a la Web Semántica Geoespacial como una prioridad en investigaciones recientes. El desarrollo de las ontologías espaciales ha sido identificado como un compromiso de investigación a largo plazo por el consorcio.

La creación de la Web Semántica Espacial necesita el desarrollo de Ontologías espaciales y terminológicas, cada una con una semántica formal, la representación de estas semánticas deben ser apropiadas para que estén disponibles en el procesamiento y para una fácil comprensión entre las personas [Egenhofer, 2002]. En este artículo, se proponen cuatro representaciones de la semántica espacial:

- 1) En lenguaje natural con el mínimo uso de marcadores, como HTML o XHTML.
- 2) Metadatos simples, lenguajes basados en XML.
- 3) Modelos de datos, RDF que permite una representación conceptual de entidades, atributos y relaciones.
- 4) Semánticas Lógicas. DAML+OIL donde las semánticas son expresadas por etiquetas.

## 2.3 Sistemas actuales de Búsquedas de Cursos Computacionales

**Google:** El buscador realiza una comparación sintáctica, devolviendo páginas que contengan palabras iguales a las que el usuario introdujo. La primer prueba fue hecha

con este buscador y nuestra solicitud fue la de encontrar “Cursos de Computación”, nos devolvió un listado de páginas Web que contuvieran esta frase, rastreamos la primer opción devuelta y para conseguir datos del curso como ubicación, horario y conocer qué tipos de cursos fue necesario abrir y explorar varias páginas.

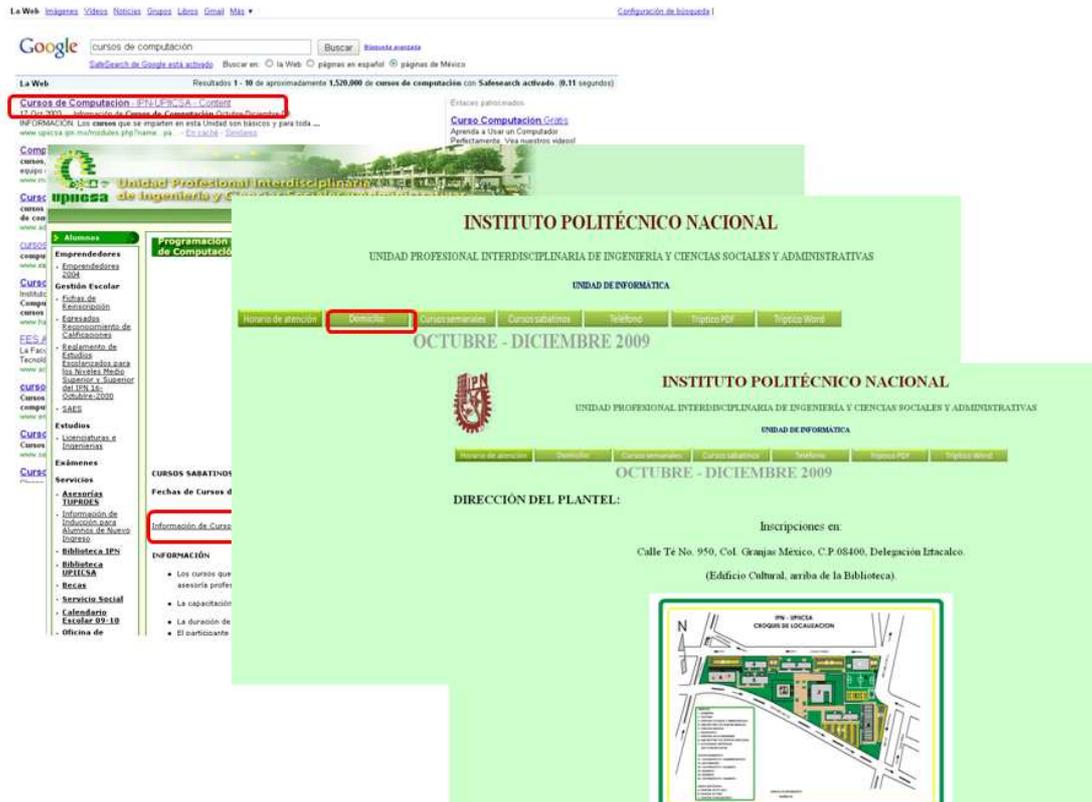


Figura 2.6. Google. Buscador de Cursos Computacionales.

**Eduportal:** Es un directorio público de escuelas en México, no incluye información acerca de cursos computacionales. Para consultar datos de alguna institución es necesario conocer el nivel académico que imparte y ubicación (Estado, Delegación o Municipio). La búsqueda se hace rastreando páginas a través de enlaces.



Figura 2.7. EduPortal. Buscador de Cursos Computacionales.

**Emagister:** Sitio Web para la búsqueda de cursos, la información se consigue seleccionando la temática de interés o tecleando lo que se desea (nombre del curso, ubicación, etc.) también hace comparaciones sintácticas, lo que puede producir resultados vacíos cuando no se encuentra alguna coincidencia.



Figura 2.8. eMagister. Buscador de Cursos Computacionales.

## 2.4 Resumen y Comentario finales

Los Sistemas Recomendadores son Sistemas de Información que ayudan a los usuarios a decidir qué artículos, productos o servicios están más apegados a lo que requiere o desea, existen mecanismos que facilitan esta tarea, entre ellos se encuentra el uso de ontologías o la programación de agentes. El principal elemento para estos Sistemas son sus motores de búsqueda.

Los Sistemas actuales para la Búsqueda de Cursos no consideran datos ni preferencias de los usuarios, como son ubicación, horario disponible o área computacional de interés del consultante. La búsqueda se logra haciendo solo comparaciones léxicas, lo que produce en ocasiones, resultados vacíos; para evitar ello, es necesario incluir análisis de los datos que permita reconocer los intereses de los usuarios y otro análisis de la información disponible para proporcionarle al cliente resultados similares a lo que desea, si es que no se encuentra exactamente lo que busca.

## Capítulo 3.

# MARCO TEÓRICO

En este capítulo se organizan los argumentos teóricos y referenciales del problema de investigación. Se estudian las estructuras de representación del conocimiento, como son las bases de datos relacionales, las bases de datos espaciales y las bases de conocimiento, se hace una clara distinción entre ellas y los beneficios de vincularlas para cumplir con el objetivo de la tesis.

### 3.1 Formas tradicionales de representación del conocimiento especializado

Los terminólogos han contado tradicionalmente con tres formas para representar el conocimiento especializado: la descripción de los conceptos por medio de sus características, la estructuración de las relaciones que existen entre conceptos y la formulación de definiciones que describen al concepto en el marco de una estructura conceptual, a partir de estas tres formas es como se originan las bases de datos y bases de conocimientos.

### 3.2 Bases de Datos

Una base de datos es un repositorio que almacena grandes cantidades de datos organizados con base a un esquema o modelo de datos que facilita su almacenamiento, recuperación, modificación y/o actualización.

### 3.3 Bases de Conocimiento

Las bases de conocimiento es un depósito de información creado a partir de una extensa investigación organizada en un árbol de conocimientos completo. Surgieron a partir de la investigación en Inteligencia Artificial como respuesta a las necesidades que las aplicaciones de esta disciplina planteaban. Las bases de conocimiento son una evolución lógica de los Sistemas de bases de datos tradicionales, con el fin de soportar no cantidades enormes de datos sino elementos de conocimiento, así como la manera en que estos elementos han de ser utilizados. También se les trata de dotar de conocimiento sobre sí mismas.

Algunas de las ventajas de las bases de conocimientos son:

- Ofrecen estructurar el conocimiento lo más parecido a la organización de la mente del ser humano.
- Facilitan el descubrimiento de nuevas relaciones y un mejor conocimiento del dominio.
- Pueden ser utilizadas en cualquier dominio del conocimiento
- Facilita la construcción de la estrategia de búsqueda.

### 3.4 Diferencias entre Bases de datos y Bases de conocimiento

Se ha estudiado mucho acerca de las diferencias concretas en cuanto a metodologías, técnicas de implementación, usos, contenidos y objetivos entre ambos Sistemas. La distinción entre base de datos y base de conocimiento, así como la distinción entre modelo de datos y esquema de representación se ha discutido durante mucho tiempo entre científicos. Un modelo de datos se asocia con una base de datos, mientras que su correspondiente en una base de conocimiento es el denominado esquema de representación.

Las bases de datos son más antiguas que las bases de conocimiento, nacieron a mediados de la década de los 50 y desde entonces han sido una de las principales herramientas que las computadoras ofrecen. A principios de los 60 se desarrollaron las primeras bases de datos estructuradas jerárquicamente y poco tiempo después, las bases de datos de red. A finales de los 60, Ted Codd, investigador de IBM, desarrolló un lenguaje de programación de propósito general que denominó programación relacional, basado en la teoría de conjuntos y la lógica, y que proporcionaría los fundamentos de las bases de datos relacionales.

Ejemplo de una diferencia entre estos dos modelos es, ante una consulta determinada, una base de datos accederá a los datos almacenados en ella y dará una respuesta afirmativa o negativa, independientemente de que tenga o no la información relevante; en cambio, una respondería "sí", "no" o "no lo sé", en el caso de carecer de todos los datos pertinentes a la consulta.

Se dice que una base de datos almacena únicamente hechos, que son un reflejo de universo, llamado Universo de Discurso determinado. Las funciones que el gestor de base de datos se limita a facilitar son, fundamentalmente, las de edición y consulta de los datos. Una base de conocimiento, por otra parte, puede almacenar, además de hechos, un conjunto de reglas que se sirven de esos hechos para obtener información que no se encuentra almacenada de forma explícita. El tipo de base de conocimiento al que se dota de una considerable capacidad de deducción a partir de la información que contiene se denomina sistema experto.

Ambos Sistemas de información cuentan con sus correspondientes gestores para simplificar al administrador las tareas comunes de mantenimiento: el sistema gestor de bases de datos (DBMS: Database Management System) y el sistema gestor de bases de conocimiento (KBMS: Knowledge Base Management System). También encontramos diferencias importantes en ellos. Básicamente los DBMSs actuales se encuentran estandarizados, ofreciendo un número de características y metodologías comunes que posibilitan la comunicación entre diversos tipos y productos comerciales. La comercialización de los KBMSs es prácticamente escasa y actualmente se está trabajando en su estandarización

La diferencia más importante entre ambos gestores es el modelo de datos en el que se basan. Los modelos de datos aportan la base conceptual para diseñar aplicaciones que hacen un uso intensivo de datos, así como la base formal para las herramientas y técnicas empleadas en el desarrollo y uso de Sistemas de información.

Con respecto al diseño de bases de datos, el modelado de datos puede ser descrito de la siguiente forma: dados los requerimientos de información y proceso de una aplicación cuyo uso de los datos es intensivo, construir una representación de la aplicación que capture las propiedades estáticas y dinámicas requeridas para dar soporte a los procesos deseados. Además de capturar las necesidades dadas en el momento de la etapa de diseño también debe ofrecer ejecutar cambios de requerimientos en el futuro.

En lo que respecta a las bases de conocimiento, la necesidad de una formalización para representar el conocimiento se hizo evidente en el área de la Inteligencia Artificial casi desde el principio de su estudio. Esta notación recibe el nombre de esquema de representación en el entorno de las bases de conocimiento. Resulta práctico considerar una base de conocimiento como un modelo de un mundo/empresa/sección de la realidad. Hemos de considerar el mundo/universo como una colección de individuos o entidades y una colección de relaciones que existen entre esos individuos. La colección de individuos que conforman el universo a representar y las relaciones que éstos mantienen constituye un estado, y puede haber transformaciones de estado que causan la creación o modificación de individuos o de las relaciones entre ellos.

## **3.4 Bases de Datos Espaciales**

Una base de datos espacial es un conjunto de datos referenciado espacialmente y que actúa como un modelo de la realidad, su construcción requiere de un proceso de abstracción para pasar de la complejidad del mundo real a una representación simplificada que pueda ser procesada por una computadora. En una base de datos espacial se describen objetos espaciales a partir de tres características: atributos, localización y topología. Los atributos permiten definir que es el objeto, la localización

proporciona el lugar que ocupa dicho objeto y por último la topología permite mejorar la interpretación semántica del contexto y establecer relaciones entre ellos.

## 3.6 Ontologías

Ontología (Onto: Ser. Logos: Estudios. "Estudio del Ser") a continuación se discuten algunas definiciones:

- Formulación de un exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados
- Forma de organizar el conocimiento con el fin de extraer información, procesarla y obtener conclusiones de una manera inteligente.
- Representa una visión del mundo con respecto a un dominio.

Actualmente coexisten dos usos diferenciados del término ontología, que corresponden a dos ramas del saber distintas y es por ello que se le atribuyen características y propiedades distintas. El término ontología se origina en el campo de la filosofía y la epistemología. Como ciencia, la Ontología es una rama de la metafísica que se ocupa del estudio de la naturaleza de la existencia, de los seres y de sus propiedades transcendentales; en filosofía, por tanto, una ontología se considera como una explicación sistemática de la Existencia.

Derivado de su significado original, el término ontología se usa en el ámbito de la Ingeniería del Conocimiento para referirse a un conjunto de conceptos organizados jerárquicamente, representados en algún sistema informático cuya utilidad es la de servir de soporte a diversas aplicaciones que requieren de conocimiento específico sobre la materia que la ontología representa. A éste segundo significado de ontología, entendida como un cuerpo estructurado de conocimiento, es al que queremos circunscribir nuestra discusión sobre ontologías y su aplicación y utilidad en la gestión terminológica.

El concepto de ontología se basa en la descripción del mundo real, por lo que nos permite representarlo por medio de clases, propiedades, ejemplos del mundo

(incidencias) y por lo tanto construir relaciones entre ellos, conocidos como reglas a través de la cual puede interactuar y por lo tanto funcionar.

Las ontologías son usadas para capturar conocimiento de algún dominio de interés, describen los conceptos en el dominio y las relaciones existentes entre ellos. Como se mencionó anteriormente una ontología está compuesta por instancias, propiedades y clases. Las instancias representan los objetos en el dominio que son de interés, las propiedades son relaciones que permiten unir dos instancias y las clases son interpretadas como conjuntos que contienen a las instancias, las clases pueden ser organizadas dentro de superclases.

### **3.6.1 Ejemplos de Ontologías**

WordNet: Base de datos léxica del idioma inglés, almacena las relaciones semánticas entre los conjuntos de sinónimos que la base de datos agrupa.

CYC: Proyecto de inteligencia artificial que intenta ensamblar un ontología comprensiva y una base de datos del conocimiento general, con el fin de permitir a las aplicaciones de inteligencia artificial realizar razonamiento del tipo humano.

### **3.6.2 Tipos de Ontologías**

Existen diversas clasificaciones, las principales son las que se basan en su grado de formalidad y las que se organizan de acuerdo a sus propósitos.

#### ***Dependiendo del grado de formalidad***

- Informales: Se expresan en cualquier lenguaje natural.
- SemiInformales: Se expresan en una forma estructurada y en lenguaje natural.
- SemiFormales: Se expresan en lenguajes estructurados.
- Formales: Se expresan utilizando lenguajes lógico-matemáticos.

***Dependiendo de su alcance y posibilidad de aplicación***

- De la aplicación: Usadas para la aplicación que la utiliza.
- Del dominio: Describen generalizaciones sobre tareas específicas en algún dominio.
- De técnicas básicas: Describen características generales de artefactos.
- Genéricas: Describe la categoría de más alto nivel. Describen conceptos muy generales.

[Guarino, 1998]. Propuso la siguiente clasificación de Ontologías:

- Ontología Universal (Ontología de Dios): Lo más general es cosa, materia, energía o universo.
- Ontología de dominio: Es un subconjunto de la Universal, en donde describe un solo dominio.
- Ontología de servicios: Es de la misma complejidad que la anterior. Describe conceptos de un tipo de servicio (mensajería, venta de libros, Autos, etc.)
- Ontología de Aplicación: Es un subconjunto de la de dominio. La cual se usa para una única aplicación.

**3.6.3 Razones para desarrollar una ontología**

1. Compartir el entendimiento común de la estructura de información entre agentes
2. Rehusó del dominio de conocimiento
3. Aclaraciones explícitas en el dominio
4. Separar el dominio de conocimiento del conocimiento operacional
5. Analizar el dominio de conocimiento

### 3.6.4 Operaciones y Relaciones de una Ontología

Las operaciones que se pueden aplicar a las ontologías se enumeran a continuación, estas operaciones requieren de una formalización (algoritmo).

- Integrar: Unir dos ontologías
- Alinear: A partir de dos ontologías eliminar conceptos iguales entre ellas
- Compartir: Las relaciones y conceptos participan entre ontologías

Las relaciones más comunes en el diseño de ontologías son:

- Es un (Is a)
- Tiene (Has)
- Hace (Do)

### 3.6.5 Reglas fundamentales para el diseño de Ontologías

1. No existe un solo camino correcto para modelar un dominio. Siempre hay varias alternativas.
2. La mejor solución siempre depende de una aplicación que se tiene en mente
3. El desarrollo de la ontología es un proceso iterativo
4. Los conceptos en la ontología deben ser cercanos a los objetos físicos y lógicos
5. Los objetos, conceptos y relaciones deben pertenecer al mismo dominio
6. Evitar el uso de palabras en plural

### 3.6.6 Enfoque iterativo para el diseño y mantenimiento de una Ontología

Describe el ciclo de vida de una ontología y pretende alargar la vida de una ontología.

1. Determinar el dominio y objetivo de la ontología. Los conceptos y relaciones debe pertenecer en el mismo dominio.

- a) Dominio que cubriría la ontología. ¿Qué dominio cubrirá la ontología?
  - b) ¿Para qué vamos a usar la ontología?
  - c) ¿Para qué tipo de consultas la información en la Ontología debe proveer las respuestas?
  - d) ¿Quién va a usar y mantener la ontología?
  - e) ¿De dónde provienen las fuentes del dominio?
2. Considerar el rehúso de otras ontologías existentes, interactuar con otras ontologías ya hechas.
  3. Enumerar términos importantes en la ontología. Dar una lista de estos términos
  4. Definir las clases y jerarquías de clases.
- Se pueden utilizar los siguientes enfoques:

- Top-Down. Se comienza de lo general a lo particular

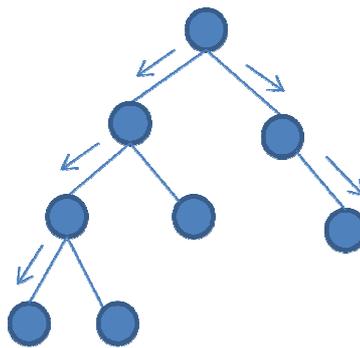


Figura 3.1. Jerarquía Top-Down.

- Botton-Up. Se parte de con cosas particulares

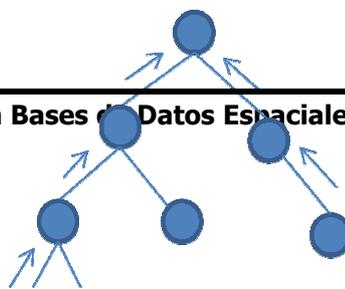


Figura 3.2. Jerarquía Button-Up.

- El tercer enfoque es una combinación del primero y el segundo
5. Definir las propiedades de las clases. Definir las propiedades tanto intrínsecas como extrínsecas.
  6. Definir los espectros de las propiedades, los cuales describen los siguientes criterios:
    - Tipo de valor
    - Valores permitidos
    - Numero de valores
    - Otras características de valores que pueden tomar los slots.
  7. Definir instancias. Se recomienda el siguiente algoritmo para crear las instancias:
    - 1) Elegir una clase
    - 2) Crear una instancia individual de esta clase
    - 3) Llenar con valores las propiedades

## 3.7 Jerarquías

Una Jerarquía es una colección de objetos arreglados de acuerdo a sus valores cualitativos, un árbol con ciertas propiedades, este concepto se usa para organizar el conocimiento de un cierto dominio. Son un tipo especial de Ontologías con ciertas restricciones. Estas estructuras se utilizan actualmente en áreas como minería de datos y data warehousing. El uso de jerarquías nos puede ayudar para estructurar información y en la búsqueda de un dato para obtener resultados precisos o lo más parecido a lo que estamos solicitando.

Una vez organizada la información en una jerarquía podemos cuantificar el error de obtener un objeto o concepto diferente al que estamos buscando. En este trabajo de investigación se utiliza la medida de similitud *Confusión* (Teoría de Confusión) [Levachkine, et al 2007] para determinar dicho error.

### 3.7.1 Tipos de Jerarquías

[Levachkine, et al 2007] define cuatro tipos de jerarquías:

- 1) Jerarquías Simples
- 2) Jerarquías Ordenadas
- 3) Jerarquías Porcentuales
- 3) Jerarquías Mixtas

#### 3.7.1.1 Jerarquías Simples

Una jerarquía simple es un árbol con raíz  $E$  y si un nodo tiene un hijo, éste forma una partición del padre. Una jerarquía simple describe una jerarquía donde  $E$  es un conjunto (así sus elementos no se repiten ni son ordenados).

Ejemplo:

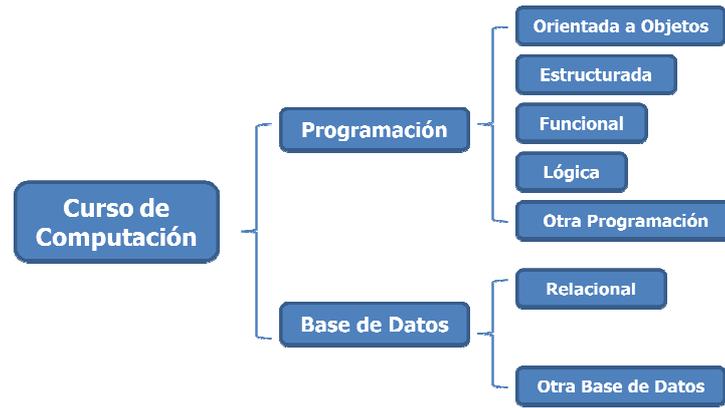


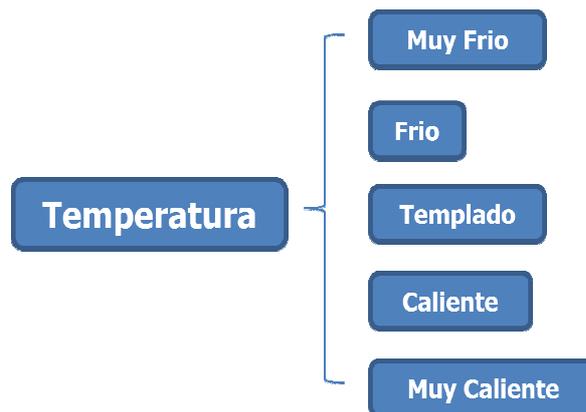
Figura 3.3. Ejemplo de una Jerarquía simple

Y se denota como: Curso de Computación {Programación {Orientada a Objetos, Estructurada, Funcional, Lógica, Otra Programación}, Base de Datos {Relacional, Otra Base de Datos}}

### 3.7.1.2 Jerarquía ordenada

Es una jerarquía en donde los nodos de una partición obedecen un orden.

Ejemplo:

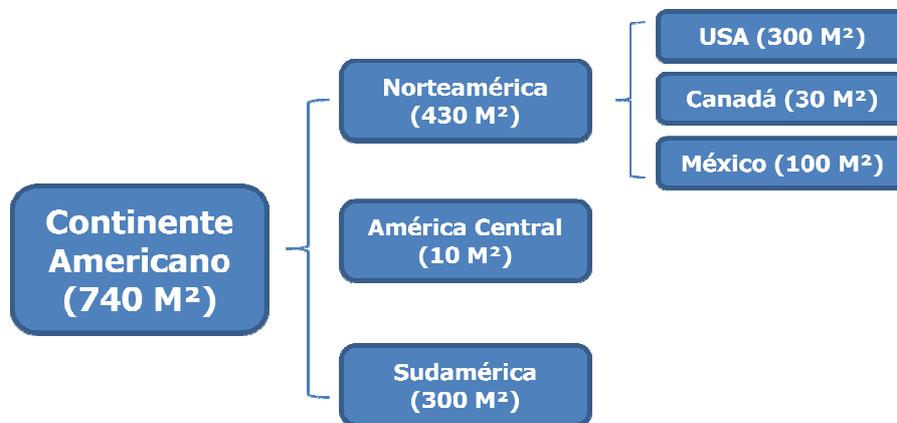


Y se denota como: Temperatura {Muy Frio, Frío, Templado, Caliente, Muy caliente}

### 3.7.1.3 Jerarquía porcentual

Es una Jerarquía en donde el tamaño de cada conjunto es conocido.

Ejemplo:



a 3.5. Ejemplo de una Jerarquía porcentual.

### 3.7.1.4 Jerarquía Mixta

Es una jerarquía que combina los tres tipos de jerarquías mencionadas anteriormente.

## 3.8 Teoría de Confusión

*Confusión* es una medida asimétrica y dependiente del contexto aplicable a las jerarquías para medir la similitud entre dos valores y el error cuando un valor  $r$  es usado en lugar de otro valor  $s$  [Levachkine, et al., 2007]. Este término se definió para diferenciar éste enfoque de otros que se orientan a medir distancias entre conceptos

(simetría, medidas independientes del contexto, cercanía, similitud) [Rentería, 2009]. La asimetría de *Confusión* se da por su definición y su dependencia al contexto por la estructura jerárquica.

### 3.8.1 Teoría de Confusión para Jerarquías Simples

#### Confusión de usar $r$ en vez de $s$ en jerarquías simples $H$ .

**Definición.** Si  $r, s \in H$ , entonces la confusión de usar  $r$  en lugar de  $s$ , que se escribe como  $\text{conf}(r, s)$ , es:

- $\text{conf}(r, r) = \text{conf}(r, s) = 0$ , donde  $s$  es cualquier ascendente de  $r$ ; (Regla 1)
- $\text{conf}(r, s) = 1 + \text{conf}(r, \text{padre\_de}(s))$ . (Regla 2)

Para medir  $\text{conf}(r,s)$ , se cuentan los enlaces descendentes de  $r$  a  $s$ , al valor reemplazado, este valor se suele normalizar al dividirlo por la profundidad de  $H$ .

### 3.8.2 Teoría de Confusión para Jerarquías Ordenadas

#### Confusión de usar $r$ en vez de $s$ , en jerarquías ordenadas

**Definición.** Para jerarquías simples compuestas por conjuntos ordenados, la confusión de usar  $r$  en vez de  $s$ ,  $\text{conf}'(r, s)$ , se define como:

- $\text{conf}'(r, r) = \text{conf}(r, \text{cualquier ascendente de } r) = 0$ ;
- Si  $r$  y  $s$  son hermanos distintos,  $\text{conf}'(r, s) = 1$  Si el padre no está en un conjunto ordenado; entonces,  $\text{conf}'(r, s)$  es igual a la distancia relativa de  $r$  a  $s$

es y a la vez es igual el número de pasos requeridos para llegar de  $r$  a  $s$  en el orden definido, dividido entre la cardinalidad-1 del padre; (Regla 3)

- $\text{conf}'(r, s) = 1 + \text{conf}'(r, \text{padre\_de}(s))$ .

### 3.8.3 Teoría de Confusión para Jerarquías Porcentuales

#### Confusión de usar $r$ en vez de $s$ , en jerarquías porcentuales

Considerando la jerarquía  $H$  (de un elemento del conjunto  $E$ ) pero compuesta por un conjunto desordenado en vez de un conjunto ordenado.

**Definición.** Para conjuntos desordenados, la confusión de usar  $r$  en lugar de  $s$ ,  $\text{conf}''(r, s)$ , es:

$\text{conf}''(r, r) = \text{conf}''(r, s) = 0$ , cuando  $s$  es cualquier ascendente de  $r$ ;

$\text{conf}''(r, s) = 1 - \text{proporción relativa de } s \text{ en } r$ . (Regla 4)

### 3.8.4 Teoría de Confusión para Jerarquías Mixtas

#### Confusión de usar $r$ en vez de $s$ , en jerarquías mixtas

**Definición.** Para calcular  $\text{conf}'''(r, s)$  en una jerarquía mixta:

- Aplicar (Regla 1) para la ruta ascendente de  $r$  a  $s$ ;
- En la ruta descendente, usar (Regla 3) en vez de (Regla 2), si  $p$  es un conjunto ordenado; o usar (Regla 4) en vez de (Regla 2), cuando los tamaños de  $p$  y  $q$  son conocidos. Es decir, usar (Regla 4) para las jerarquías porcentuales en lugar de (Regla 2). Esta definición es consistente y reduce las definiciones previas de jerarquías porcentuales, simples, ordenadas y mixtas.

### 3.9 Objetos idénticos, muy similares y algo similares.

Los objetos son entidades descritas por el par  $k$  (propiedad, valor), la cual en nuestra notación nos referimos como el par (variable, valor). Un objeto  $o$  con el par  $k$  (variable, valor) se define como  $(o(v_1, a_1)(v_2, a_2) \dots (v_k a_k))$ .

Se desea calcular el error de utilizar un objeto  $o'$  en lugar del objeto  $o$ .

Para un objeto  $o$  con  $k$  variables  $v_1, v_2, \dots, v_k$  y valores  $a_1, a_2, \dots, a_k$  se dice que otro objeto  $o'$  con las mismas variables  $v_1, v_2, \dots, v_k$  pero con valores  $a'_1, a'_2, \dots, a'_k$  se presentan las siguientes definiciones:

**Definición.**  $o'$  es **idéntico** a  $o$ , si  $a'_i = a_i$  para todas  $1 \leq i \leq k$ . Los valores correspondientes son idénticos.

Si todo lo que se sabe sobre  $o$  y  $o'$  son los valores de las variables  $v_1, v_2, \dots, v_k$  y ambos objetos tienen valores idénticos, se puede decir que "con todo lo que se sabe",  $o$  y  $o'$  son el mismo.

**Definición.**  $o'$  es **sustituto** para  $o$ , si  $\text{conf}(a'_i, a_i) = 0$  para todo  $1 \leq i \leq k$ . No se trata de una confusión entre el valor del atributo de  $o'$  y el valor correspondiente de  $o$ . Se puede usar  $o'$  en vez del  $o$  solicitado con  $\text{conf} = 0$ .

**Definición.**  $o'$  es **muy similar** a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = 1$

**Definición.**  $o'$  es **similar** a  $o$  si,  $\sum_i \text{conf}(a'_i, a_i) = 2$

**Definición.**  $o'$  es **algo similar a**  $o$  si,  $\sum_i \text{conf}(a'_i, a_i) = 3$

**Definición.** En general,  $o'$  es **Similar<sub>n</sub>** a  $o$ , si  $\sum_i \text{conf}(a'_i, a_i) = n$

## 3.10 Resumen y Comentarios Finales.

En este capítulo se analizaron las formas de representar el conocimiento especializado, todos estos métodos se crean de acuerdo a las necesidades de almacenamiento y procesamiento de la información, con el objetivo de permitir accesos más rápidos y análisis minuciosos para un eficaz aprovechamiento del conocimiento.

Cada vez es más frecuente el uso de las ontologías y jerarquías en los Sistemas de búsqueda de información y recomendación ya que facilitan la representación de los datos, los organiza y permiten crear relaciones entre ellos.

# Capítulo 4.

## METODOLOGÍA

En este capítulo se describe la metodología propuesta de la tesis para solución a la problemática planteada en el capítulo 1, se definen y describen el diseño de las estructuras que organizan el conocimiento y la información del área de aplicación de la tesis. La siguiente figura muestra los pasos generales de la metodología.



Figura 4.1. Metodología propuesta.

**Conceptualización:** es la primera fase de la metodología en donde se investiga y entiende el problema, se hace una recopilación y análisis de datos necesarios, haciendo una búsqueda de fuentes que proporcionen información sobre el dominio de estudio y se examina la información para detectar heterogeneidad conceptual que impida la eficiente manipulación de la misma.

**Modelado:** se refiere al diseño de las estructuras que serán pilares de organización y almacenamiento de los datos.

**Integración:** es la fase en donde se conectan las estructuras del paso anterior y se comienzan a poblar con los datos recopilados.

**Análisis:** se procesa la información, consiste en el estudio de las preferencias del usuario, un análisis semántico y un análisis espacial a fin de ofrecer las mejores opciones al usuario.

**Resultados:** se presenta y visualiza los resultados del procesamiento de la información y del análisis.

**Recomendaciones futuras:** de cursos computacionales al usuario vía correo electrónico de acuerdo al perfil y consultas registradas.

## 4.1 Definiciones previas

El caso de estudio está enfocado a la Búsqueda del mejor lugar para tomar un Curso de Computación, dado esto es necesario explicar la definición de "Curso" y hacer una distinción entre términos relacionados.

### **Curso:**

- Unidad educativa en la que se ofrece un conjunto estructurado de conocimientos teóricos y/o prácticos. (Definición dada por la Secretaría de Educación Pública).
- Término académico que denomina la duración de instrucción en una materia. (Wikipedia).

Requiere o no de conocimientos previos, no es un grado académico y puede no formar parte de un programa de estudios avalado por la Secretaría de Educación Pública.

***Materia:***

- Unidad académica que incluye una o varias asignaturas que pueden concebirse de manera integrada. (Wikipedia).

***Diplomado:***

- Programa curricular que estructura unidades de enseñanza-aprendizaje, sobre determinado tema y que tiene suficiente extensión y formalidad para garantizar la adquisición y desarrollo de un conocimiento teórico y/o práctico válido.

## 4.2 Casos de Estudio

En este apartado describiremos los casos de estudio que la metodología enfrentará.

Se clasifican en tres procesos.

**1) Consultas por el Usuario:** Para realizarlas es necesario definir el perfil del usuario. El cual consiste en solicitarle los siguientes datos.

- a) Nombre o clave de acceso: Es el identificador de cada uno de los usuarios.
- b) Dirección: Ubicación Geográfica del usuario.
- c) Horario: Días y Horas disponibles para la toma del curso.
- d) Modalidad: Modalidad del curso de interés.
- e) Costo: Cantidad monetaria que el usuario está dispuesto a pagar por el curso.
- f) Área Computacional de Interés: Área de Computación que es de interés para el usuario.
- g) Correo Electrónico: Correo Electrónico del usuario para recomendaciones futuras.

Una vez registrados los datos anteriores, el usuario podrá ejecutar una consulta de cursos. Se estudian las siguientes propiedades de los cursos: costo,

modalidad, horario y distancia con respecto a lo que el usuario desea para ofrecer la mejor alternativa de lo que se tiene disponible.

**2) Actualización de datos por Instituciones educativas:** Consiste en actualizar información de los cursos por parte de las instituciones que los imparte. Los datos requeridos son:

- a) Identificación de la Institución Educativa: Clave o Nombre de la Escuela.
- b) Cursos: Cursos que imparten.
- c) Horario: Días y Horas de los cursos.
- d) Modalidad: Modalidad de cada uno de los cursos.
- e) Costo: Costo de cada uno de los cursos.

**3) Recomendaciones Futuras:** Después de haber ejecutado los procesos anteriores el sistema podrá enviar sugerencias al usuario sobre cursos de interés.

## 4.3 Conceptualización.

La conceptualización es una perspectiva abstracta y simplificada del conocimiento que tenemos del mundo real, y que por alguna razón queremos representar. Para comenzar con esta primera fase de la metodología y entender el dominio de nuestro tema de estudio se recopilamos y analizamos los datos.

Esta información fue abstraída de artículos publicados, consulta bibliográfica, rastreo de páginas Web y entrevistas con algunas instituciones educativas que imparten cursos de computación.

## 4.4 Modelado.

En esta sección presento el diseño de la ontología, la base de datos espacial y jerarquías que representa las propiedades de los cursos.

### 4.4.1 Diseño de la Ontología.

Para el diseño de éste modelo se investigó acerca de la Asociación Nacional de Instituciones de Educación en Tecnologías de la Información, A.C. (ANIEI). La ANIEI nació en Guadalajara, Jal., el 8 de Octubre de 1982. Su objetivo es el de contribuir a la formación de profesionales en Informática y Computación sólidamente preparados, y de impulsar la difusión y la asimilación de una cultura computacional en la sociedad, acorde a lo que el mundo actual, cada vez más informatizado, y el futuro emanado de la revolución informática presente, exigen. Dentro sus fines principales se encuentra el de orientar, proponer y difundir las actividades que en materia de docencia, investigación y extensión educativa se realizan en el área de informática.

La ANIEI propone un catálogo de áreas de conocimiento en computación compuesta de subáreas y subsubáreas la cuales están estructuradas en grupos de temas de estudio, que no corresponden necesariamente a materias, con el propósito de establecer una norma de los tópicos que deben considerarse en la enseñanza y aprendizaje de las disciplinas computacionales. Este catálogo se utilizó como base para el diseño de la Ontología ODBCC (Ontología para la Difusión y Búsqueda de Cursos Computacionales).

#### 4.4.1.1 Catálogo de áreas de conocimiento en Computación

El catálogo se forma en primera instancia de ocho áreas, se describe en qué consiste cada una de ellas y las subáreas correspondientes:

- i) **Entorno Social:** Comprende conocimientos, normas, experiencias y motivaciones que hacen posible la buena integración de las unidades de informática y su personal en las organizaciones y en la sociedad en general. Se incluyen tópicos de administración, economía, contabilidad, derecho, sociología y psicología.
- a) Las Organizaciones
  - b) Las Unidades de Informática
  - c) Ética y Normatividad
  - d) Herramientas Computacionales
- ii) **Matemáticas:** Las matemáticas brindan una excelente e imprescindible base de tipo formativo para el desarrollo de habilidades de abstracción y la expresión de formalismos, además de proporcionar conocimientos específicos fundamentales para la informática y la computación.
- a) Matemáticas Básicas
  - b) Matemáticas Aplicadas
  - c) Matemáticas Discretas
  - d) Teoría Matemática de la Computación
  - e) Herramientas Computacionales
- iii) **Arquitectura de Computadoras:** Estudio de la teoría, técnicas, tecnologías y métodos para comprender el funcionamiento de los Sistemas digitales y las computadoras, así como de los principios físicos que los sustentan, con el objeto de formular algunas de sus especificaciones y saber integrar equipos diversos para fines particulares.
- a) Física
  - b) Sistemas Digitales
  - c) Tipos y Configuración de Computadoras
  - d) Instalaciones y Equipos
  - e) Herramientas Computacionales
- iv) **Redes:** Estudio de la fusión de los dominios tradicionalmente considerados como hardware y software, y formas de distribuir y compartir recursos computacionales, procesos e información.
- a) Transmisión y Comunicación de Datos

- b) Modelos
  - c) Protocolos
  - d) Intercomunicación de Redes
  - e) Seguridad e Integridad de la Información
  - f) Herramientas Computacionales
- v) **Software de Base:** Estudio, definición y construcción de las piezas de software que hacen posible el funcionamiento de las computadoras en diferentes niveles operativos. Por su importancia formativa y metodológica, esta área de conocimiento resulta fundamental para los desarrollos de la industria de los programas para computadoras.
- a) Traductores
  - b) Sistemas Operativos
  - c) Utilerías y Manejadores
  - d) Herramientas Computacionales
- vi) **Programación e Ingeniería de Software:** Cuerpo de conocimientos teóricos y prácticos, y conjunto de metodologías para la buena construcción de programas y Sistemas de software, considerando su análisis y diseño, confiabilidad, funcionalidad, costo, seguridad, facilidades de mantenimiento y otros aspectos relacionados.
- a) Algorítmica
  - b) Paradigmas de Programación y Lenguajes
  - c) Sistemas de Software
  - d) Industria del Software
  - e) Herramientas Computacionales
- vii) **Tratamiento de Información:** Área de conocimientos en la cual se conjuga una multiplicidad de tópicos computacionales de teoría, técnica y metodología, requeridos para la construcción de una amplia gama de soluciones de información, imprescindibles para el adecuado funcionamiento de todo tipo de organizaciones.
- a) Base de Datos
  - b) Recuperación de la Información
  - c) Sistemas de Información
  - d) Herramientas Computacionales

viii) ***Interacción Hombre-Máquina:*** Estudio de los dominios de aplicación conducentes a lograr formas superiores de expresión e interacción entre el hombre y la computadora, con el fin de buscar mejores y novedosas maneras de integración de la tecnología en la sociedad.

- a) Graficación
- b) Inteligencia Artificial
- c) Interfaces Humano-Máquina
- d) Herramientas Computacionales

A continuación se presenta la información organizada y detallada sobre las áreas computacionales, para la construcción formal de esta estructura se usó la herramienta Protegé, una plataforma que permite la creación de ontologías basada en el lenguaje Java.



Figura 4.2. Parte I de la estructura de la Ontología (áreas y subáreas).

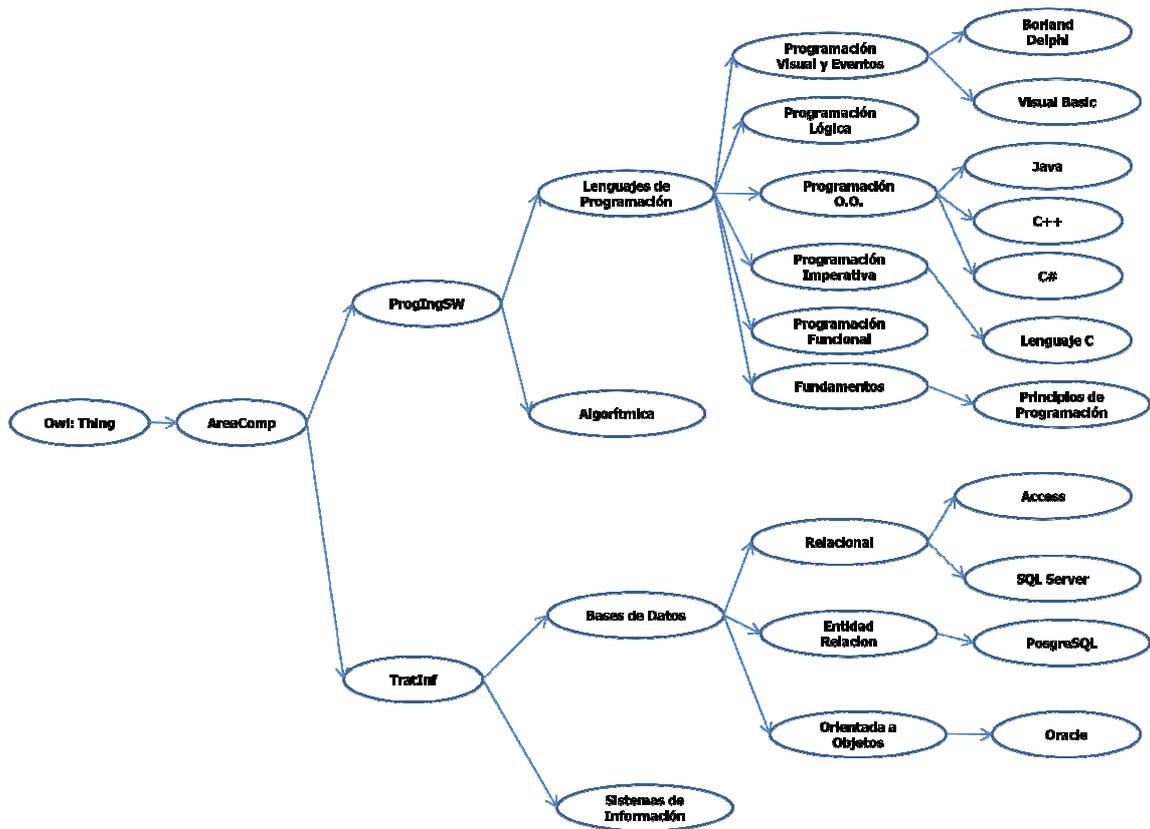


Figura 4.3. Parte II de la estructura de la Ontología.

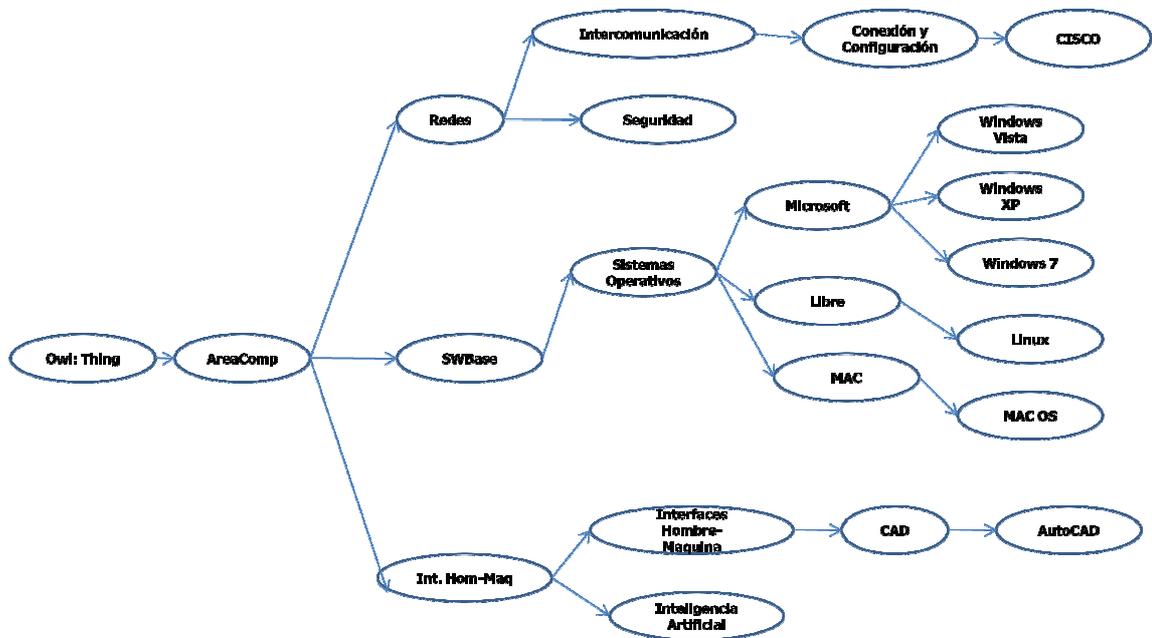


Figura 4.4. Parte III de la estructura de la Ontología.

### 4.4.2 Diseño de la Base de Datos Espacial.

Esta base de datos es la que organiza información de los usuarios (perfiles), datos de las instituciones educativas y cursos.

#### 4.4.2.1 Modelo Conceptual de la Base de Datos Espacial

Se compone de cinco entidades (Usuario, Perfil, Área de Computación, Curso y Escuela), las cuales dos de ellas (Usuario y Escuela) contiene una componente espacial que nos permite registrar la ubicación geográfica de sus elementos.

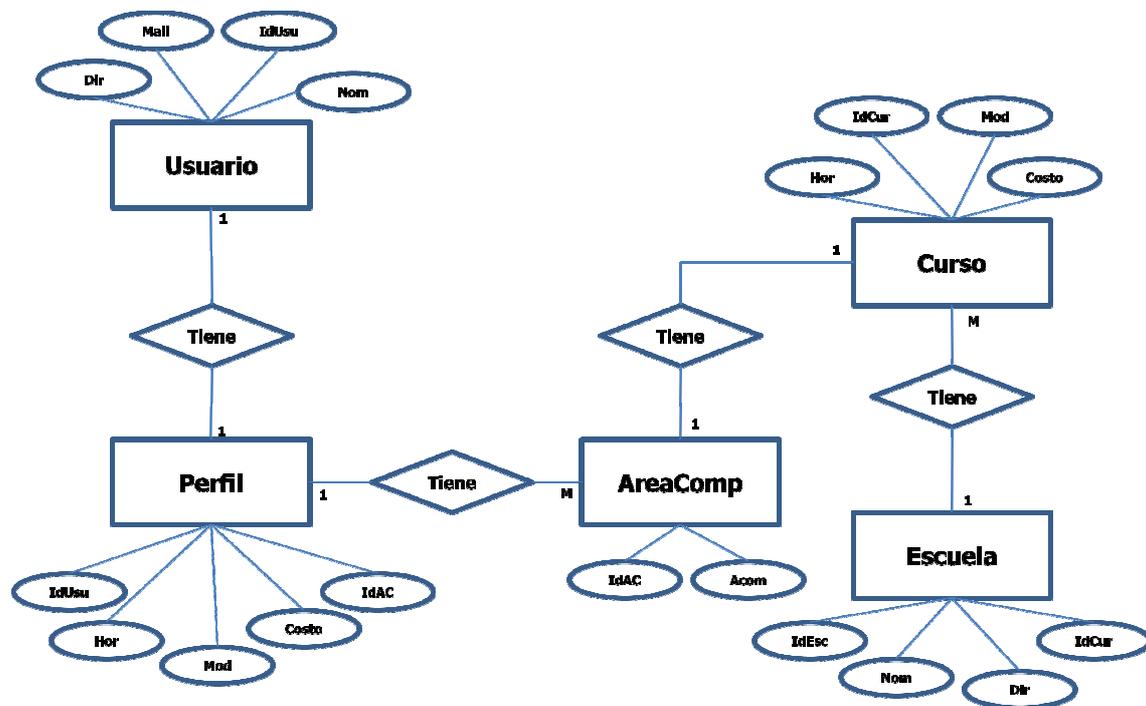


Figura 4.5. Modelo conceptual de la base de datos espacial

- Entidad **Usuario**: responsable de organizar los datos personales del usuario.
- Entidad **Perfil**: se conecta con la entidad de usuario, en ella se registran las preferencias de los usuarios respecto a los cursos computacionales.

- Entidad **Área Computación:** responsable de almacenar las áreas computacionales, las cuales están vinculadas con el perfil, cada usuario puede tener una o más área computacional de interés.
- Entidad **Escuela:** responsable de organizar los datos de las instituciones educativas que imparte los cursos.
- Entidad **Curso:** en esta entidad se registran las propiedades de cada uno de los cursos que imparten las escuelas, cada curso se vincula a un área de computación.

### 4.4.3 Diseño de las Jerarquías de Propiedades

Las propiedades de los cursos computacionales (costo, modalidad, distancia y horario) se organizan en jerarquías para poder ser evaluadas y ponderar los resultados devueltos a la consulta.

#### 4.4.3.1 Jerarquía de Costos

Es una jerarquía ordenada que organiza las diferencias de costos entre lo que está dispuesto a pagar el usuario con el costo real del curso. Se crearon cuatro categorías:

- **Muy barato:** cuando la diferencia de los costos sea menor a cero pesos.
- **Barato:** cuando la diferencia de los costos sea entre cero a mil pesos.
- **Caro:** cuando la diferencia de los costos sea entre mil dos mil pesos.
- **Muy Caro:** cuando la diferencia de los costos sea mayor a mil pesos.

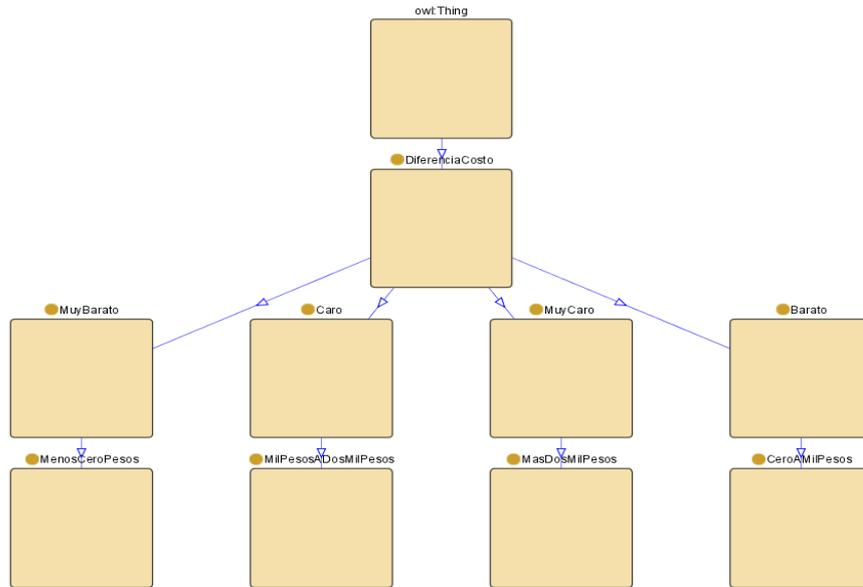


Figura 4.6. Jerarquía de Costos.

### 4.4.3.2 Jerarquía de Modalidad

Esta jerarquía organiza las modalidades de los cursos. Es una jerarquía simple. Se crearon tres categorías principales.

- **Intensivo:** cuyos cursos se imparten de lunes a viernes.
- **SemiIntensivo:** cuyos cursos de imparte tres días a la semana.
- **Sabatino:** corresponde a los cursos que se ofrecen los sábados.

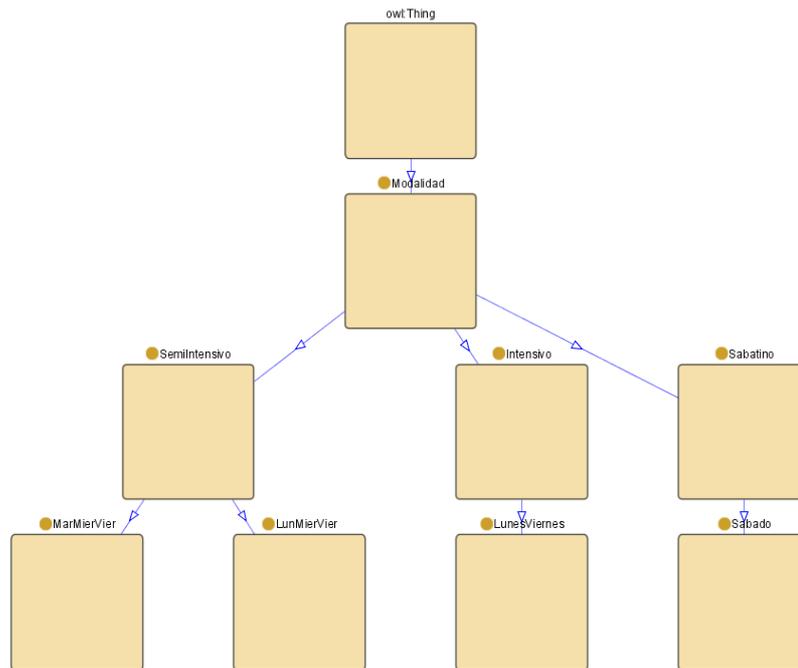


Figura 4.7. Jerarquía de Modalidad.

#### 4.4.3.3 Jerarquía de Distancia

Jerarquía ordenada que organiza las distancias entre la dirección que proporcionó el usuario y la dirección de la escuela donde se imparte el curso. Se crearon cuatro categorías:

- Muy cerca: cuando la distancia sea de cero a cinco kilómetros.
- Cerca: cuando la distancia sea de cinco a diez kilómetros.
- Muy Lejos: cuando la distancia sea de diez a quince kilómetros.
- Lejos: cuando la distancia sea mayor a quince kilómetros.

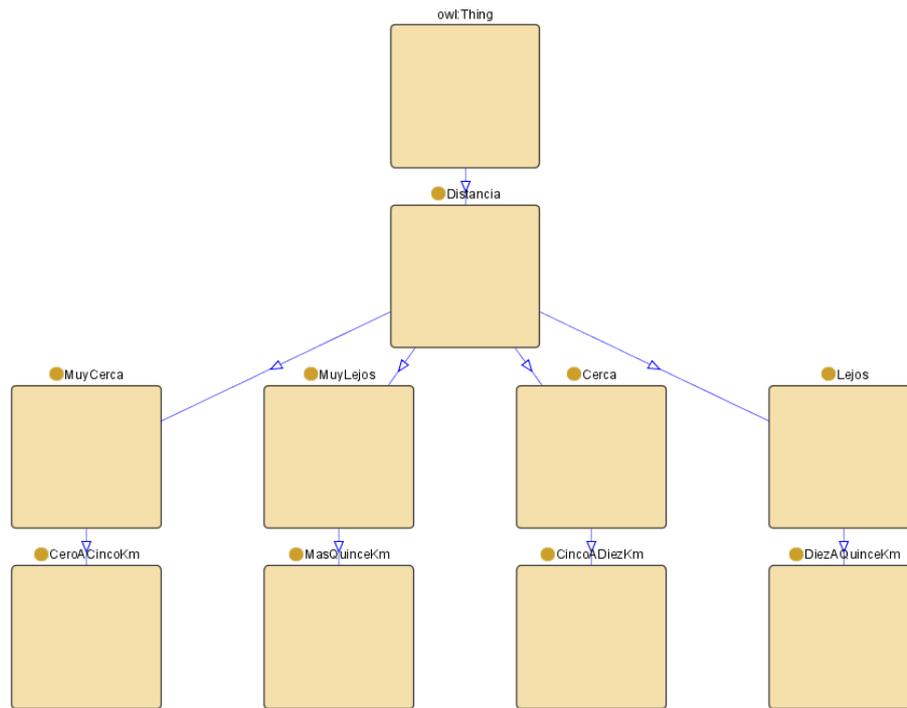


Figura 4.8. Jerarquía de Distancia.

#### 4.4.3.4 Jerarquía de Horario

Esta jerarquía es la responsable de organizar las horas del día, es también una jerarquía ordenada. El día se distribuye en cuatro turnos, cada hora se divide en dos, es decir se consideran las horas y medias:

- **Mañana:** Corresponde desde las seis horas (6:00 a.m.) del día hasta las once y media (11:00 a.m.).
- **Medio día:** Corresponde desde las doce horas (12:00 a.m.) del día hasta las catorce y media (2:30 p.m.).
- **Tarde:** Corresponde desde las quince (3:00 p.m.) horas hasta las diecinueve horas del día (7:00 p.m.).
- **Noche:** Corresponde desde diecinueve y media (7:30 p.m.) hasta las veinticuatro horas del día (12:00 p.m.).

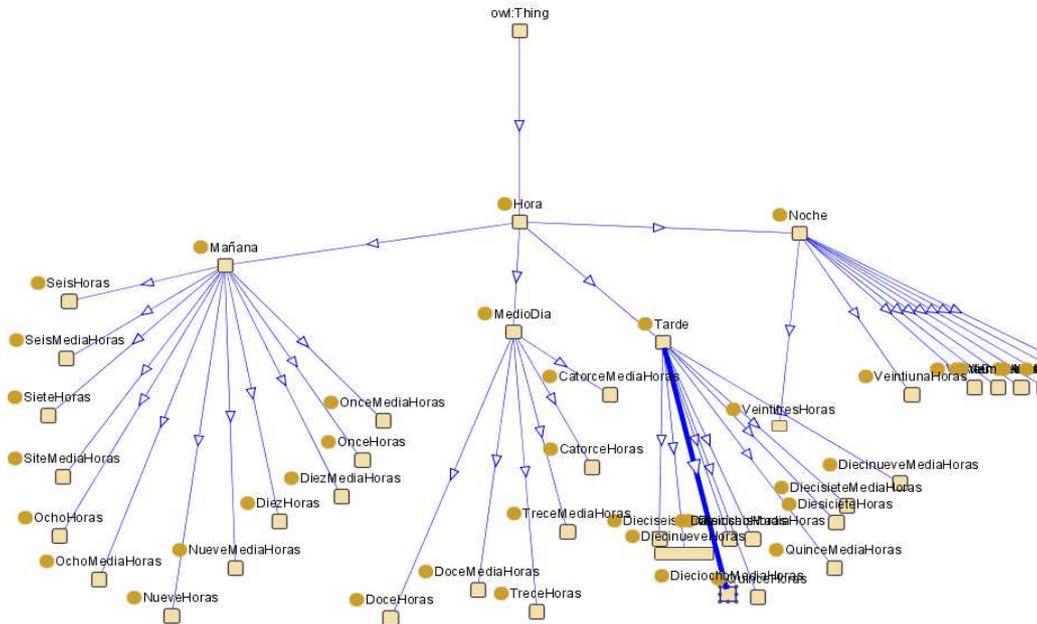


Figura 4.9. Jerarquía de Horario.

## 4.5 Integración

En esta fase se conecta la base de datos espacial con la ontología, es decir cada registro de cursos computacionales está relacionado directamente con un área computacional. La base de datos espacial también almacena datos tanto de las escuelas y los usuarios.

La base de datos espacial se encuentra sustentada sobre el manejador de bases de datos PostgreSQL y la ontología sobre Protegé. Cada curso en la base de datos espacial está vinculado a una subsubárea de la ontología como se muestra en la figura 4.10.



## 4.6 Análisis

En la fase de análisis se estudian las preferencias del usuario y la consulta ejecutada, se comparan con los datos capturados en la base de datos espacial y las jerarquías de propiedades. Se producen dos análisis, el semántico y el espacial.

### 4.6.1 Análisis Espacial

En esta fase se determina la distancia que existe entre la dirección que proporcionó el usuario con la de la escuela donde se imparten los cursos que son de interés al usuario. Se utiliza la distancia Euclidiana, para lograr este objetivo. Se adquirió un fragmento de "GeoPostcodes", una base de datos mundial de ciudades con códigos postales, regiones, latitud y longitud, para referenciar las direcciones. Una vez teniendo ambos códigos postales (el de la dirección del usuario y el de la escuela) se puede medir la distancia entre ellos, en el capítulo 2 se describió la distancia Euclidiana y la distancia de Manhattan, estas métricas permiten un análisis geográfico sobre la lejanía o cercanía de dos puntos de interés. En esta tesis se utiliza la distancia Euclidiana por su fácil implementación.

### 4.6.2 Análisis Semántico

En el análisis semántico se estudian las preferencias de los usuarios y la consulta ejecutada por el usuario, se comparan con las propiedades que se tienen organizadas en la ontología y en las jerarquías de propiedades. Se obtiene un valor de confusión por cada propiedad y al final se suman todos los valores de las propiedades (Ecuación para el cálculo de confusión de un Curso Computacional), el orden en que se presentan los resultados es en forma ascendente, de acuerdo al valor total de confusión, es decir,

iniciando la lista de resultados con aquel que tenga el menor valor y finalizando con el de mayor.

$$Conf_t = Conf_{ac}(\ ) + Conf_c(\ ) + Conf_m(\ ) + Conf_d(\ ) + Conf_h(\ )$$

Ecuación para el cálculo de confusión de un Curso Computacional.

Donde:

$Conf_t$  = Es el valor de similitud final de un curso de computación.

$Conf_{ac}(\ )$  = Valor de similitud obtenido de la ontología de áreas computacionales.

$Conf_c(\ )$  = Valor de similitud obtenido de la jerarquía de propiedad "Costo".

$Conf_m(\ )$  = Valor de similitud obtenido de la jerarquía de propiedad "Modalidad".

$Conf_d(\ )$  = Valor de similitud obtenido de la jerarquía de propiedad "Distancia".

$Conf_h(\ )$  = Valor de similitud obtenido de la jerarquía de propiedad "Horario".

#### 4.6.2.1 Análisis Semántico de la ontología

En este apartado se hacen los cálculos de similitud dentro de la ontología de áreas computacionales, para obtener cursos exactos o parecidos a los que el usuario solicita, los cursos computacionales se encuentran vinculados a las áreas y subáreas de la ontología.

#### 4.6.2.2 Análisis Semántico de las jerarquías de propiedades

En esta sección de la metodología se realizan los cálculos de similitud dentro de cada una de las jerarquías de propiedades, para determinar el valor de similitud cada curso

de acuerdo a sus características, se buscan aquellos que tengan propiedades más parecidas a las que el usuario desea.

No se utilizan valores de ponderación para determinar el nivel de importancia para el usuario de cada una de las propiedades, al usuario se le ofrecen las mejores opciones que el sistema determina.

Para la propiedad de costo, el mejor es el que se encuentra en la clase de "Muy Barato" y el peor para el de la clase "Muy caro", de la misma forma para la propiedad de distancia. Para las propiedades de horario, la mejor opción será aquella que se encuentra más cercana semánticamente a la hora que el usuario proporcionó, de la misma manera para la propiedad de modalidad.

## 4.7 Presentación de Resultados

En este paso de la metodología, se le presentan al usuario los resultados del procesamiento y análisis del paso anterior, los resultados se publican con un orden de relevancia, el cual consiste en los cursos cuyas propiedades sean más parecidas a los gustos del usuario.

## 4.8 Recomendaciones Futuras

Es el último paso de la metodología en donde al usuario se le proporciona sugerencias de cursos que pueden tomar en el futuro, de acuerdo a sus preferencias y consultas ejecutas. Las recomendaciones se les manda automáticamente vía correo electrónico.

## 4.9 Arquitectura de la Aplicación

La aplicación final está sustentada sobre las siguientes plataformas y herramientas.

- **Google Maps:** Es un servidor de aplicaciones de mapas en la Web. Ofrece imágenes de mapas desplazables, así como fotos satelitales del mundo entero e

incluso la ruta entre diferentes ubicaciones. Este servidor se utiliza en el sistema principalmente para ubicar en un mapa los resultados devueltos de la consulta.

- **PostgreSQL:** Es un sistema gestor de base de datos relacional orientada a objetos y soporta también organizar y registrar datos espaciales. Sobre este gestor se creó la base de datos espacial del sistema.
- **MS4W:** Es un servidor de mapas que permite interactuar con la información geográfica, entre sus propósitos para este sistema es la de visualizar, identificar y consultar atributos alfanuméricos de los objetos geográficos y la conexión de la base de datos espacial con el servidor de mapas.
- **Protégé:** Es una plataforma gratuita que permite la creación de ontologías, está basado en el lenguaje Java. Ofrece un ambiente fácil y sencillo para el diseño de las mismas.
- **Jena:** Es un framework para construir aplicaciones orientados a la Web Semántica. Proporciona un ambiente para lenguajes de consultas semánticas. Se utilizó en el sistema para consultar los conceptos de la ontología y de las jerarquías de propiedades.

El siguiente diagrama muestra la arquitectura general de la aplicación.

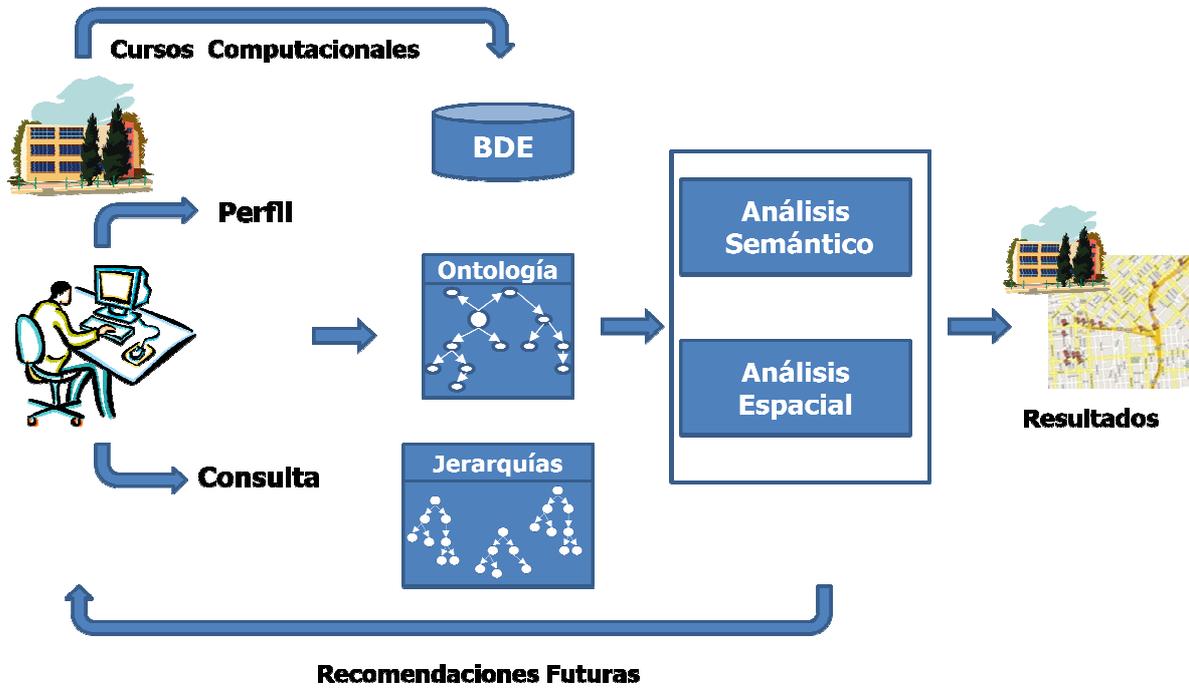


Figura 4.11. Arquitectura de la Aplicación.

El desempeño de la aplicación inicia cuando la base de datos espacial (BDE) es poblada con los datos sobre cursos computacionales que proporcionan las escuelas y cuando el usuario registra su perfil y ejecuta la consulta. Una vez poblada la BDE se comienza el procesamiento partiendo de la ontología y jerarquías de propiedades la cuales son predefinidas y su diseño permanece estable.

## 4.10 Resumen y Comentarios Finales.

En este capítulo se presentó la metodología propuesta en la tesis, la cual describe los pasos a seguir para la obtención de los mejores resultados en un sistema de búsqueda basado en las preferencias del usuario.

Dentro de la metodología se usa una ontología para organizar el conocimiento del área de interés, en este trabajo es sobre las disciplinas computacionales. Se usan también jerarquías para la representación de las propiedades de los elementos a buscar, para nuestro caso de estudio son las propiedades sobre los cursos computacionales. La base de datos espacial registra los perfiles de los usuarios (preferencias y datos de interés) así como las instancias finales de los cursos, es decir, datos reales proporcionados por las escuelas sobre los cursos que imparten. Una vez organizada la información se inicia el procesamiento con un análisis semántico y posteriormente con un análisis espacial para la obtención de los mejores resultados que se ofrecen a los usuarios.

## Capítulo 5.

# PRUEBAS Y RESULTADOS

En este capítulo se presentan los resultados obtenidos con el modelo propuesto en el capítulo anterior y se describen las pruebas realizadas al mismo.

### 5.1 Interfaz principal de la aplicación

La siguiente figura muestra la página principal del sistema, pantalla que podrá observar el usuario para realizar una búsqueda y mediante la cual deberá introducir las propiedades del curso solicitado:

- Curso: Nombre del curso
- Costo: Cantidad monetaria por el cual está dispuesto a pagar.
- Modalidad: Si está interesado por un curso Intensivo, SemiIntensivo o Sabatino
- Hora: Hora disponible para la toma del curso.

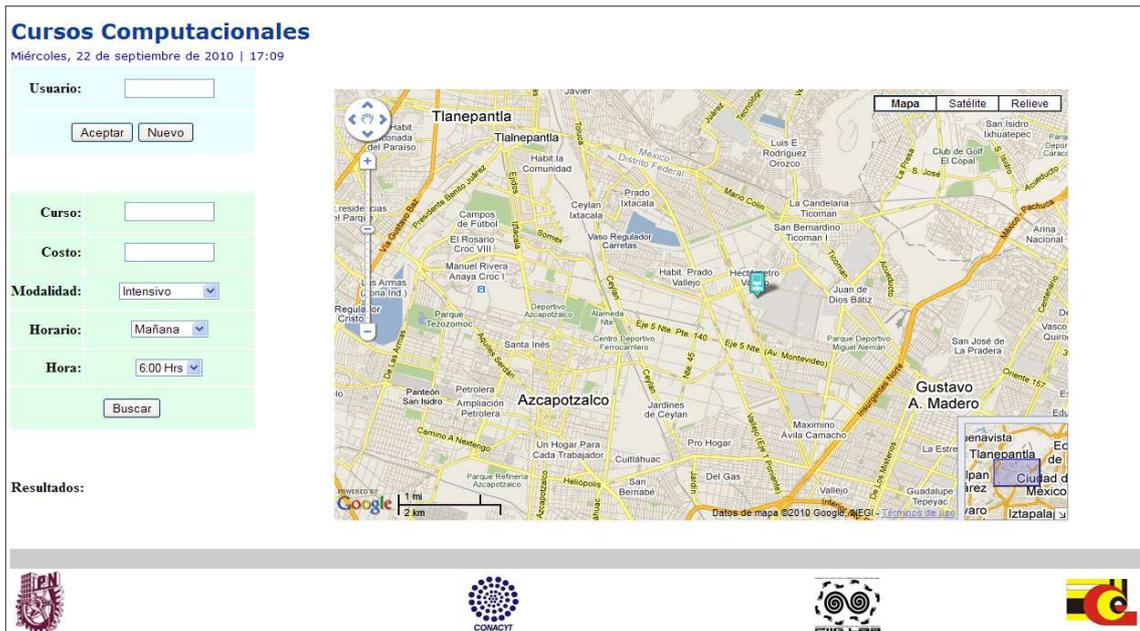


Figura 5.1 Pantalla principal del sistema.

## 5.2 Pruebas a la ontología

A continuación se muestra una prueba que se le hizo a la ontología, para conocer el valor de similitud (confusión) entre algunas áreas y subáreas de computación.

$r$	$s$	AreaComp	EntSocial	Matemáticas	ArqComp	Redes	SWBase	Etica y Normatividad	Matemáticas Discretas	Config. Computadoras	Sistemas Digitales	Seguridad	Sistemas Operativos	Lenguajes de Programación	Base de Datos	Graficación
		0	1	1	1	1	1	2	2	2	2	2	2	2	2	2
AreaComp	0	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
EntSocial	0	0	1	1	1	1	1	2	2	2	2	2	2	2	2	2
Matemáticas	0	1	0	1	1	1	2	1	2	2	2	2	2	2	2	2
ArqComp	0	1	1	0	1	1	2	2	1	1	2	2	2	2	2	2
Redes	0	1	1	1	0	1	2	2	2	2	1	2	2	2	2	2
SWBase	0	1	1	1	1	0	2	2	2	2	2	1	2	2	2	2
Etica y Normatividad	0	0	2	2	2	2	0	2	2	2	2	2	2	2	2	2
Matemáticas Discretas	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2
Config. Computadoras	0	2	2	0	2	2	2	2	0	2	2	2	2	2	2	2
Sistemas Digitales	0	2	2	0	2	2	2	2	2	0	2	2	2	2	2	2
Seguridad	0	2	2	2	0	2	2	2	2	2	0	2	2	2	2	2
Sistemas Operativos	0	2	2	2	2	0	2	2	2	2	2	0	2	2	2	2
Lenguajes de Programación	0	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2
Base de Datos	0	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2
Graficación	0	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2

Tabla 5.1 Resultados de la Prueba a la ontología de algunas áreas y subáreas.

A continuación se muestran una pruebas que se le hizo a la ontología, para conocer el valor de similitud (confusión) entre algunos cursos computacionales.

Curso	Similitud	Curso	Similitud
Java	0	Oracle	4
C++	1	PostgreSQL	4
C#	1	CISCO	4
Lenguaje C	2	Windows XP	4
Borland Delphi	2	Windows Vista	4
Visual Basic	2	Windows 7	4
Principios de Programación	2	Linux	4
Access	4	MAC OS	4
SQL Server	4	AutoCAD	4

Tabla 5.2 Resultados de la Prueba 1, al consultar un curso de Java.

Curso	Similitud	Curso	Similitud
Java	4	Oracle	2
C++	4	PostgreSQL	1
C#	4	CISCO	4
Lenguaje C	4	Windows XP	4
Borland Delphi	4	Windows Vista	4
Visual Basic	4	Windows 7	4
Principios de Programación	4	Linux	4
Access	2	MAC OS	4
SQL Server	0	AutoCAD	4

Tabla 5.3 Resultados de la Prueba 2, al consultar un curso de SQL Server.

### 5.3 Pruebas finales

Las siguientes pruebas muestran los resultados de una búsqueda de un curso computacional considerando sus cuatro propiedades (costo, modalidad, distancia y

horario). Se describen a fin de ejemplo, seis registros de cursos que se encuentran disponibles en el sistema.

### **Preferencias del Usuario**

---

**Curso computación:** Java

**Costo:** 2, 000 pesos

**Modalidad:** Sabatino

**Horario:** 9:00 a 13:00 hrs

**Dirección del usuario:** Joyas de Vallejo. A2. Depto. 203. Col. Santa Rosa. Gustavo A. Madero. Distrito Federal. C.P. : 07620

### **Disponibilidad en el sistema**

---

#### **Curso 1**

**Curso computación:** Java

**Costo:** 1, 500 pesos

**Modalidad:** Sabatino

**Horario:** 8:00 a 12:00 hrs.

**Dirección de la escuela:** Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal. Unidad Profesional Adolfo López Mateos. Col. Nueva Industrial Vallejo Gustavo A. Madero C.P. 07738.

#### **Curso 2**

**Curso computación:** C++

**Costo:** 2, 500 pesos

**Modalidad:** Sabatino

**Horario:** 9:00 a 12:00 hrs.

**Dirección de la escuela:** Av. Instituto Politécnico Nacional s/n, Unidad Profesional "Adolfo López Mateos", Edif. 1, 2, 3, 4 y 5, Col. Lindavista, Del. Gustavo A. Madero, México, D.F. C.P. 07740.

**Curso 3**

**Curso computación:** Java

**Costo:** 1, 500 pesos

**Modalidad:** Sabatino

**Horario:** 9:00 a 12:00 hrs.

**Dirección de la escuela:** Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal. Unidad Profesional Adolfo López Mateos. Col. Nueva Industrial Vallejo Gustavo A. Madero C.P. 07738.

**Curso 4**

**Curso computación:** C#

**Costo:** 2, 000 pesos

**Modalidad:** Intensivo

**Horario:** 8:00 a 10:00 hrs.

**Dirección de la escuela:** Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal. Unidad Profesional Adolfo López Mateos. Col. Nueva Industrial Vallejo Gustavo A. Madero C.P. 07738.

**Curso 5**

**Curso computación:** SQL Server

**Costo:** 3, 500 pesos

**Modalidad:** Sabatino

**Horario:** 9:00 a 12:00 hrs.

**Dirección de la escuela:** Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal. Unidad Profesional Adolfo López Mateos. Col. Nueva Industrial Vallejo Gustavo A. Madero C.P. 07738.

**Curso 6**

**Curso computación:** AutoCAD

**Costo:** 4, 200 pesos

**Modalidad:** SemiIntensivo

**Horario:** 11:00 a 13:00 hrs.

**Dirección de la escuela:** Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal. Unidad Profesional Adolfo López Mateos. Col. Nueva Industrial Vallejo Gustavo A. Madero C.P. 07738.

### Procesamiento del sistema

---

#### Curso 1

**Diferencia de costos:** -500 pesos

**Distancia entre la ubicación del usuario y la escuela:** 4.5 Kilómetros.

$Conf_{ac}(\ ) = 0$ ;  $Conf_c(\ ) = 0$ ;  $Conf_m(\ ) = 0$ ;  $Conf_d(\ ) = 0$ ;  $Conf_h(\ ) = 2/11$ .

#### Curso 2

**Diferencia de costos:** 500 pesos

**Distancia entre la ubicación del usuario y la escuela:** 5.5 Kilómetros.

$Conf_{ac}(\ ) = 1$ ;  $Conf_c(\ ) = 0$ ;  $Conf_m(\ ) = 0$   $Conf_d(\ ) = 2$   $Conf_h(\ ) = 0$ .

#### Curso 3

**Diferencia de costos:** -500 pesos

**Distancia entre la ubicación del usuario y la escuela:** 4.5 Kilómetros.

$Conf_{ac}(\ ) = 0$ ;  $Conf_c(\ ) = 0$ ;  $Conf_m(\ ) = 0$ ;  $Conf_d(\ ) = 0$ ;  $Conf_h(\ ) = 0$ .

#### Curso 4

**Diferencia de costos:** 0 pesos

**Distancia entre la ubicación del usuario y la escuela:** 4.5 Kilómetros.

$Conf_{ac}(\ ) = 1$ ;  $Conf_c(\ ) = 0$ ;  $Conf_m(\ ) = 2$ ;  $Conf_d(\ ) = 0$ ;  $Conf_h(\ ) = 2/11$ .

#### Curso 5

**Diferencia de costos:** 1,500 pesos

**Distancia entre la ubicación del usuario y la escuela:** 4.5 Kilómetros.

$Conf_{ac}(\ ) = 4$ ;  $Conf_c(\ ) = 2$ ;  $Conf_m(\ ) = 0$ ;  $Conf_d(\ ) = 0$ ;  $Conf_h(\ ) = 0$ .

#### Curso 6

**Diferencia de costos:** 2,200 pesos

**Distancia entre la ubicación del usuario y la escuela:** 4.5 Kilómetros.

$Conf_{ac}() = 4$ ;  $Conf_c() = 2$ ;  $Conf_m() = 2$ ;  $Conf_d() = 0$ ;  $Conf_h() = 4/11$ ;

De los seis registros disponibles en el sistema y considerando las preferencias del usuario los resultados obtenidos del procesamiento se muestran a continuación.

Curso	SimCurso
Curso 1. Java	0.1818
Curso 2. C++	2
<b>Curso 3. Java</b>	<b>0</b>
Curso 4. C#	3.1818
Curso 5. SQLServer	6
Curso 6. AutoCAD	8.3636

Tabla 5.4. Resultados de la prueba final.

La mejor opción para el usuario es el **Curso 3**, cuyo valor total de similitud es de cero.

## Capítulo 6.

# CONCLUSIONES Y TRABAJO A FUTURO

A continuación se describen las conclusiones de los resultados adquiridos con el presente trabajo de investigación, además se proponen temas y trabajos a futuro para dar seguimiento y optimización de ésta tesis.

## 6.1 Conclusiones

El conocimiento del mundo real es extenso y conceptualmente diverso para los humanos, lo que puede producir heterogeneidad en significados de conceptos y como consecuencia final, desentendimiento entre personas. Una adecuada organización del conocimiento facilita la comprensión del mundo que nos rodea, las ontologías son el recurso ideal para lograr este fin, además que contribuyen a optimizar la compartición de información en los Sistemas de información.

Los Sistemas orientados al usuario basados en ontologías permiten que los resultados sean más exacto a lo que se está buscando, ahorrando tiempo de búsqueda y rastreo.

Por otro lado, el procesamiento de consultas en las bases de datos espaciales es la actividad principal de los Sistemas de recuperación de información geográfica y los Sistemas orientados al usuario donde se involucren datos referenciados, este procesamiento además de generar resultados espaciales, los datos obtenidos pueden ser vinculados a otras variables que sean de interés al usuario y que ayudan a la toma de decisiones.

## 6.2 Propuestas de trabajo futuro

Las propuestas de trabajo futuro son:

- 1) El crecimiento del número de áreas computacionales es inevitable, ya que muchos son los campos del conocimiento que se adhieren a ellas con el fin de optimizar sus procesos, realizar pruebas y demostraciones más rápidas y eficientes. Este incremento de las áreas de computación no está considerado en el presente trabajo, por lo que se propone a futuro crear una interfaz que permita añadir nuevas áreas a la ontología y en el mejor de los casos que a través de sus características se puedan vincular automáticamente a las ya definidas.
- 2) Derivado del punto anterior, que la ontología pueda ser modificada respecto a los datos que las instituciones educativas proporcionen, en el caso, por ejemplo, cuando convoquen a un nuevo curso que corresponda a una nueva área de computación.
- 3) Las jerarquías de propiedades también puedan ser modificadas según las preferencias del usuario.
- 4) Crear grupos de personas con características similares. Esta propuesta servirá en el caso de estudio para que las escuelas tomen decisiones sobre que cursos impartir y a qué personas invitar.

---

# Referencias

[Abian, M., 2005]

**Abian, M.** *Ontologías: Qué son y para qué sirven*. <http://www.wshoy.sidar.org>. 2005.

[Adomavicius, et al, 2005]

**Adomavicius G., and Tuzhilin A.** *Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*. IEEE Internet Computing, Volumen 17, No. 6 pp. 734-749. 2005.

[Barchini, et al., 2006]

**Barchini G.E., Álvarez M.M., Diana Palliotto D., Herrera S., and Budán P.** *Ontologías en los Sistemas de Información / Conocimiento*. Universidad Nacional de Santiago del Estero. Argentina, 2006

[Böhm, et al., 2008]

**Böhm K. and Daub E.** *Geographical Analysis of hierarchical business structures by interactive drill down*. In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. ISBN: 978-1-60558-323-5. 2008.

[Bucher, et al., 2005]

**Bucher, B., P. Clough, H. Joho, R. Purves, and A. K. Syed.** 2005. *Geographic IR Systems: Requirements and Evaluation*. En *Proceedings of the 22nd International Cartographic Conference*.

[Castell, 2004]

**Castell P.** *La Web Semántica*. Universidad Autónoma de Madrid. España, 2004.

[Castillo, et al., 2008]

**Castillo L., Armengol E., Onaindía E., Sebastía L., González-Boticario J., Rodríguez J.A., Fernández S., Arias, J.D., and Borrajo, D.** *SAMAP. An user-oriented adaptative system for planning tourist visits.* Expert System with Applications, Volume 34, Number 34, p. 1328-1332. 2008.

[Delgado, et al., 2002]

**Delgado J. and Richard Davidson,** *Knowledge Bases and User Profiling in Travel and Hospitality Recomen der System.* In Proceedings of the ENTER 2002 Conference, pages 1–16, Innsbruck, Austria, 2002. Springer Verlag.

[Edward, 2003]

**Edward S .** *Capturing Knowledge of user preferences with recommenders system.* A thesis submitted for the degree of Doctor of Philosophy. University de Southmapton. Reino Unido. 2003.

[Egenhofer, 2002]

**Egenhofer M.** *Toward the Semantic Geospatial Web.* In GIS '02: Proceedings of the 10th ACM international symposium on Advances in geographic information systems. pp. 1-4, Virginia, USA, 2002.

[Fonseca, et al., 2002]

**Fonseca, Egenhofer, and Agouri.** Using Ontologies for Integrated Geographic. Information Systems, Transaction in GIS. 2002.

[Golbeck, et al., 2004]

**Golbeck J., Alford A. and Hendler J.** *Organization and Structure of Information using Semantic Web Technologies.* Handbook of Human Factors in Web Design , Robert W. Proctor and Kim-Phuong L. Vu (eds.). Lawrence Erlbaum Associates. 2004

[Gómez H.F., 2003 ]

**Gómez H.F.** *Ontología para Sistemas de Información Geográfica.* Universidad Nacional de Educación a Distancia. Universidad Técnica Particular de Loja.

[Gutierrez, 2006]

**Gutierrez M.** *El Rol de las Bases de Datos Espaciales en una Infraestructura de Datos.* Chile: Universidad Católica de la Santísima Concepción, 2006.

[Henrich, et al., 2007]

**Henrich A. and Lüdecke V.** *Characteristics of Geographic Information Needs.* In GIR '07 of the 4th ACM workshop on Geographical information retrieval. ISBN: 978-1-59593-828-2. 2007.

[Horridge, et al., 2004]

**Horridge M., Knublauch H., Rectos A., Stevens R. and Wroe C.** *A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools.* USA: The University of Manchester, Stanford University, 2004.

[Ibarra, 2009]

**Ibarra Vargas J.J.,** *Análisis Espacial en Espacios Conceptuales.* Tesis presentada en el Centro de Investigación en Computación CIC-IPN para obtención del grado de Maestro en Ciencias de la Computación. México, 2009.

[Kodama, et al.,2009]

**Kodama K., Iijima Y., Guo X. and Ishikawa Y.** *Skyline Queries Based on User Locations and Preferences for Making Location-Based Recommendations.* In Proceedings of the 2009 International Workshop on Location Based Social Networks. ISBN: 978-1-60558-860-5. 2009.

[Lausen, et al., 2005]

**Lausen H., Stollberg M., Lara R., Ding Y., Han S. and Fensel D.** *Semantic Web Portals – State of the Art Survey.* Journal of Knowledge Management, Vol. 9 Iss: 5, pp.40 – 49. 2005.

[Levachkine, et al., 2007]

**Levachkine, S. and Guzmán-Arenas, A.** *Hierarchy as a new data type for qualitative variables.* Expert Systems with Applications: An International Journal, 32(3): pp. 899-910, 2007.

[Mata, 2009]

**Mata Rivera M.F.**, *Recuperación y Ponderación de Información Geográfica desde Repositorios no estructurados conducidas por Ontologías*. Tesis presentada en el Centro de Investigación en Computación CIC-IPN para obtención del grado de Doctor en Ciencias de la Computación. México, 2009.

[Martins et al., 2005]

**Martins B., Silva J. and Andrade L.** *Indexing and Ranking in GeoIR Systems*. In Proceeding of the 2. International Workshop on Geo-IR, 2005.

[Papadia, et al., 2010]

**Papadias D. and Sellis T.** *Qualitative Representation of Spatial Knowledge in Two-dimensional Space*. The VLDB Journal. Volume 1 / 1992 - Volume 19 / 2010.

[Perea, et al., 2008]

**Perea J.M., García M., García M.A. and Ureña L.A.** *Sistemas de Recuperación de Información Geográfica multilingües en CLE*. Procesamiento del Lenguaje Natural, Revista nº 40, marzo de 2008, pp. 129-136.

[Radilla, 2008]

**Radilla Flor.** *Modelado de datos para bases de datos espaciales. Caso de estudio: Sistemas de información geográfica*. . Tesis presentada en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. Departamento de Computación. Para obtención del grado de Maestro en Ciencias en la especialidad de Ingeniería Eléctrica. México, 2008.

[Ray, et al., 2004]

**Ray R., and Frontiera A.** *Spatial Ranking Methods for Geographic Information Retrieval (GIR) in Digital Libraries*. Lecture Notes in Computer Science, 2004, Volume 3232/2004, 45-56. DOI: 10.1007/978-3-540-30230-8\_5.

[Rentería, 2009]

**Rentería Agualimpia W.**, *Recuperación Controlada de Información Cualitativa desde Repositorios de Datos*. Tesis presentada en el Centro de Investigación en Computación

CIC-IPN para obtención del grado de Maestro en Ciencias de la Computación. México, 2009.

[Rigaux, et al., 2002]

**Rigaux, P., Scholl M. and Voisard A.** *Spatial Databases. With application to GIS.* USA: Morgan Kaufmann publishers, 2002.

[Saboya, 2005]

**Saboya A.** *Uso de Recomendadores, Asistentes y Ayudantes en Sistemas Tutores.* Universidad Politécnica de Catalunya. España, 2005.

[Shekha S., et al, 2008]

**Encyclopedia of GIS.** *Shekhar Shashi and Xiong Hui.* Springer. USA, 2008.

[Torres, et al., 2008]

**Torres M., Levachkine S., Quintero R., Guzmán G. and Moreno M.** *Geospatial information integration based on the conceptualization of geographic domain.* In GIS '08. Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. 2008.

[Wang, et al., 2005]

**Wang Ch., Xie X., Wang L., Lu Y. and Ma. W.** *Detecting Geographic Locations from Web Resources.* In GIR'05 Proceedings of the 2005 workshop on Geographic information retrieval. 2005.

**Weakliam J., Bertolotto M. and Wilson D.** *Implicit Interaction Profiling for Recommending Spatial Content.* In GIS'05. Proceedings of the 13th annual ACM international workshop on Geographic information systems. ISBN:1-59593-146-5. USA. 2005.

[Wenchen, et al., 2009]

**Wenchen V., Zheng Y. and Yang Q.** *Joint Learning Users Activities and Profiles from GPS Data.* In Proceedings of the 2009 International Workshop on Location Based Social Networks. ISBN: 978-1-60558-860-5. 2009.