



**INSTITUTO POLITÉCNICO NACIONAL**

---

---

**CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**

**LABORATORIO DE PROCESAMIENTO INTELIGENTE DE  
INFORMACIÓN GEOESPACIAL**

**“USO DE MEDIDAS DE SIMILITUD SEMÁNTICA  
PARA PROCESAMIENTO DE  
INFORMACIÓN NO ESTRUCTURADA”**

**TESIS**

**QUE PARA OBTENER EL GRADO DE  
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA:**

**NAHUN ENRIQUE MONTOYA IRIBE**

**DIRECTOR DE TESIS:**

**DR. SERGUEI LEVACHKINE**



**México D. F.**

**Julio, 2010**



---

## Agradecimientos

Quiero agradecer a mi madre María Magdalena Iribe Rubio, a mis hermanas Christi Dayana e Isela Marlen y demás miembros de mi familia por su apoyo en todo momento, su cariño incondicional, sus enseñanzas, sus valores y principios.

Agradezco a mi esposa Lucia Lizeth Delgado Covarrubias por su paciencia y apoyo todo este tiempo, por haberme regalado, lo más valioso que tengo en la vida, mi pequeño hijo (Nahun Fernando Montoya Delgado).

Tuve el privilegio de haber estudiado este posgrado, en el Instituto Politécnico Nacional (IPN) y quedo agradecido ante esta institución y específicamente ante el Centro de Investigación en Computación (CIC), por la formación recibida a lo largo de este periodo de estudios. Pero muy particularmente, agradezco a todos los miembros del Laboratorio de Procesamiento Inteligente de Información Geoespacial (PIIG-Lab). A mi asesor el Dr. Serguei Levachkine y a los demás doctores del laboratorio: Dr. Marco A. Moreno, Dr. Miguel Torres, Dr. Rolando Quintero, Dr. Giovanni Guzmán. A todos ellos gracias por sus comentarios, opiniones y bellas tardes de seminario donde compartimos y aprendimos juntos, sobre diversos temas.

Agradezco a dios, haber tenido la dicha de conocer y hacer grandes compañeros y amigos, como: Walter, Roberto, Gerardo, Iyeliz, Child (Alejandro Martínez), Lina, John, Fabiola, Olga, Fer, Manuel, Dorian, Jimmy, Alejandro, Víctor, Julio, entre otros más. A todos ellos gracias, por las experiencias y buenos ratos que pasamos juntos, sus amabilidades y atenciones.

Agradezco al Dr. Jesús Figueroa Nazuno, porque gracias a aquel congreso donde nos conocimos y coincidimos con una gran amiga (la Dra. Ma. Lucia Barrón Estrada), tuve la inquietud de hacer un posgrado en este centro de investigación. De igual forma, agradezco al Dr. Santiago Suarez haberme brindando la oportunidad de participar en el proyecto SIGAPO en la ESCOM, proyecto del cual han surgido nuevos y mejores trabajos para mí.

Finalmente, doy las gracias a todas las personas que me han apoyado de alguna forma, para dar este nuevo paso en la carrera de la vida.

# ÍNDICE

---

AGRADECIMIENTOS.....	I
ÍNDICE .....	II
LISTA DE FIGURAS .....	VI
LISTA DE TABLAS.....	VIII
RESUMEN.....	IX
ABSTRACT .....	X
<b>CAPÍTULO 1. INTRODUCCIÓN .....</b>	<b>1</b>
1.1 Antecedentes .....	1
1.2 Planteamiento del problema.....	2
1.3 Objetivos .....	3
1.3.1 Objetivo general .....	3
1.3.2 Objetivos específicos .....	3
1.4. Motivación y justificación .....	4
1.5. Beneficios esperados .....	5
1.6. Alcances y limites.....	5
1.7. Hipótesis .....	5
1.8. Organización del documento .....	6
<b>CAPÍTULO 2. ESTADO DEL ARTE .....</b>	<b>8</b>
2.1 Recuperación de información .....	8
2.1.1 Recuperación de información geográfica .....	9
2.2. Similitud.....	9
2.2.1. Relación entre similitud y disimilitud.....	10
2.2.2. Medidas de similitud.....	11
2.2.3. Un enfoque basado en clases para relaciones Léxicas.....	11

2.2.4. Medidas de similitud en taxonomías.....	12
2.2.5. Medidas de similitud en WordNet.....	13
2.2.6. Un enfoque no-probabilístico.....	13
2.2.7. Similitud semántica en jerarquías.....	16
2.3. Búsquedas semánticas.....	18
2.3.1. Web semántica geoespacial.....	19
2.4. Aplicaciones relacionadas.....	20
2.4.1. Búsqueda de productos o servicios con herramientas GIS .....	20
2.4.1.1. Google maps.....	21
2.4.1.2. Nestoria.....	21
2.4.2. Aplicaciones GIR.....	22
2.4.2.1. iRank.....	22
2.4.2.2. Análisis espacial en espacios conceptuales .....	22
2.4.2.3. GEOSYSMU.....	23
2.4.3. Sistemas de compra-venta de vehículos vía Web.....	23
2.5. Comentarios finales.....	24
<b>CAPÍTULO 3. MARCO TEÓRICO.....</b>	<b>25</b>
3.1. Recuperación de información geográfica.....	25
3.2. Información georeferenciada.....	26
3.3. Bases de datos espaciales.....	26
3.4. Análisis espacial y procesamiento de datos.....	27
3.5. Ontologías como una alternativa para estructurar y procesar información .....	28
3.5.1. Las ontologías como una conceptualización.....	28
3.6. Jerarquías.....	31
3.6.1. Tipos de jerarquías.....	33
3.6.1.1. Jerarquía simple.....	33
3.6.1.2. Jerarquía ordenada.....	34

3.6.1.3. Jerarquía porcentual.....	34
3.6.1.4. Jerarquía mixta.....	34
3.7. Teoría de confusión.....	34
3.7.1. Cálculo de confusión en jerarquías simples.....	35
3.7.2. Cálculo de confusión en jerarquías ordenadas.....	35
3.7.3. Cálculo de confusión en jerarquías porcentuales.....	36
3.7.4. Cálculo de confusión en jerarquías mixtas.....	36
3.8. Comentarios finales.....	36
<b>CAPÍTULO 4. METODOLOGÍA.....</b>	<b>38</b>
4.1. Conceptualización del dominio.....	39
4.1.1. Selección de características.....	40
4.1.2. Construcción de la base de datos conceptual.....	41
4.1.3. Construcción de la base de datos espacial.....	45
4.1.3.1. Diseño lógico de la base de datos espacial.....	46
4.1.3.2. Diseño físico de la base de datos espacial.....	47
4.2. Búsqueda y recuperación de información.....	47
4.2.1. Análisis semántico.....	49
4.2.2. Análisis geoespacial.....	51
4.2.3. Ranking de resultados.....	53
4.3. Arquitectura de la aplicación.....	56
4.4. Comentarios finales.....	57
<b>CAPÍTULO 5. PRUEBAS Y RESULTADOS.....</b>	<b>59</b>
5.1. Evaluación de las componentes del modelo.....	59
5.1.1. Análisis semántico.....	61
5.1.2. Análisis geoespacial.....	60
5.1.3. Asignación y evaluación de los valores de ponderación de cada criterio.....	61
5.2. Evaluación del sistema de búsqueda y recuperación de información....	65

5.2.1. Interfaz del sistema.....65

- 5.2.1.1. Ubicación del usuario sobre el mapa..... 66
- 5.2.1.2. Criterios de búsqueda.....67
- 5.2.1.3. Semáforo de colores..... 72
- 5.2.1.4. Realizar búsqueda..... 72
- 5.2.1.5. Resultados.....72
- 5.2.1.6. Información del vendedor..... 75
- 5.2.1.7. Detalles del producto.....75

5.2.2. Análisis de resultados..... 75

5.2. Comentarios finales..... 82

**CAPÍTULO 6. CONCLUSIONES Y TRABAJO FUTURO..... 84**

- 6.1. Conclusiones.....84
- 6.2. Aportaciones.....85
- 6.3 Trabajo futuro.....86

**REFERENCIAS.....87**

**ANEXOS.....92**

- ANEXO A. Jerarquías..... 92
- ANEXO B. Construcción de la base de datos.....106
- ANEXO C. Poblando la base de datos.....108
- ANEXO D. Códigos fuente.....121



---

# LISTA DE FIGURAS

---

2.1. Ejemplo de similitud en figuras geométricas.....	9
2.2. Jerarquía de CD's utilizada en [18].....	16
2.3. Evaluación de las medidas de similitud según Genesan y Molina .....	18
2.4. Google maps.....	21
2.5. Nestoria.....	21
2.6. Error: "no match found" en Yahoo! Autos.....	23
4.1. Conceptualización del dominio.....	40
4.2. Tipos de transmisión.....	41
4.3. Tipos de dirección.....	41
4.4. Clasificación de automóviles.....	42
4.5. Tipos de tracción.....	43
4.6. Cilindrada del motor.....	43
4.7. Tipos de combustible.....	43
4.8. Aspiración del motor.....	44
4.9. Configuración del motor.....	44
4.10. Tipos de interiores.....	44
4.11. Nivel de equipamiento.....	45
4.12. Diseño lógico de la base de datos espacial.....	46
4.13. Proceso de búsqueda y recuperación de información.....	48
4.14. Seudocódigo de los pasos que se deben realizar durante el proceso de búsqueda de información.....	48
4.15. Seudocódigo para el proceso de ranking.....	53
4.16. Arquitectura de la aplicación.....	56
5.1. Resultados del cálculo de confusión.....	60
5.2. Segmento de jerarquía sobre los seres vivos.....	61

---

5.3. Barra para indicar la importancia de cada criterio y semáforo de colores.....	65
5.4. Interfaz del sistema.....	65
5.5. Modalidades de navegación.....	66
5.6. Controles de navegación.....	66
5.7. Reubicación del usuario.....	67
5.8. Criterio de clasificación del vehículo.....	68
5.9. Listado de automóviles por marca.....	69
5.10. Tipo de búsqueda.....	69
5.11. Selección de los criterios de búsqueda.....	71
5.12. Listado completo de criterios para la búsqueda avanzada.....	71
5.13. Semáforo de colores.....	72
5.14. Botón para realizar la búsqueda.....	72
5.15. Listado de resultados – Interfaz gráfica.....	73
5.16. Ubicación de un vendedor.....	74
5.17. Información sobre un vendedor.....	74
5.18. Descripción del producto.....	75
5.19. Búsqueda de un Nissan 350z.....	76
5.20. Consulta con el criterio de distancia.....	77
5.21. Ubicación de la agencia Honda Visión.....	78
5.22. Combinación de los criterios de distancia y clasificación de los automóviles.....	79
5.23. Búsqueda de un automóvil de 10 cilindros.....	81



---

## LISTA DE TABLAS

4.1. Conjunto de datos de prueba.....	49
4.2. Resultados del análisis semántico.....	51
4.3. Resultados del cálculo de la distancia.....	52
4.3. Resultados del cálculo de la distancia normalizados.....	52
4.5. Ponderación de criterios.....	52
4.6. Criterios ponderados.....	55
4.7. Resultados ordenados de acuerdo a su relevancia.....	55
5.1. Resultados de la consulta simple de un Nissan 350Z.....	76
5.2. Resultados utilizando únicamente la distancia.....	78
5.3. Resultados utilizando únicamente de los criterios de distancia y clasificación.....	79
5.4. Resultados de una consulta compleja.....	81
5.5. Resultados obtenidos a partir de una consulta donde no existe correspondencia en la BD.....	82



---

## **RESUMEN**

La Web Semántica requiere la mejora de mecanismos para la búsqueda y recuperación de información. En esta tesis, se presenta un modelo de recuperación de información enfocado en la búsqueda de productos o servicios considerando su ubicación geográfica, el cual se basa en los atributos cualitativos para realizar un análisis semántico y obtener resultados semánticamente similares a la información buscada por los usuarios.

Adicionalmente, se realiza un análisis geoespacial para obtener resultados que estén cercanos a la ubicación de los usuarios. Los resultados son recuperados y ordenados según su relevancia y evaluados por una expresión que permite a los usuarios ponderar los criterios semánticos de la búsqueda, así como, la distancia desde la ubicación actual del usuario hasta los objetos geográficos buscados con la finalidad de personalizar los resultados considerando las preferencias del usuario.

Finalmente, se presenta un sistema que implementa el modelo propuesto, usando la venta de vehículos por internet como caso de estudio. El cual se centra en la recuperación de automóviles similares a los que el usuario está buscando. Sin embargo, el sistema recupera y ordena las agencias con los automóviles deseados que están localizados cerca de la posición actual del usuario. Los resultados muestran las ventajas de utilizar el análisis semántico en conjunto con el análisis geoespacial.



---

## **ABSTRACT**

The Semantic Web requires improving mechanisms for search and retrieval of information. In this thesis, we describe an information retrieval model focused on products and services considering their geographic locations, which is based on qualitative attributes to make a semantic analysis and obtain results that are semantically similar to the information searched by the users.

Additionally, we make geospatial analysis to obtain results that are closer to the users' location. The results are retrieved and ranked by its relevance and evaluated by an expression that let the users weigh up both the semantic search criteria as well as the distance from their actual location to the location of the searched geographic object in order to personalize the results considering the user preferences.

Finally, we present a system that implements the proposed model, using internet-supplied car sales as a study case. It is focused on information retrieval of similar cars that the user is looking for. Moreover, the system retrieves and ranks at first places the agencies with desired cars that are closer located to the user actual position. The results show the advantages in using semantic analysis in conjunction with geospatial analysis.

# CAPÍTULO 1.

## INTRODUCCIÓN

En este capítulo se mencionan algunos antecedentes, que ayudan a ubicar e introducir al lector en el contexto de esta investigación. Posteriormente, se menciona el planteamiento del problema que se aborda, cuales son los objetivos que se pretenden alcanzar, así como, la motivación y justificación de este trabajo. Adicionalmente se mencionan cuales son los beneficios esperados y las hipótesis o premisas que rigen esta investigación. Finalmente se menciona la organización de este documento.

### 1.1. Antecedentes

Los sistemas dedicados a la búsqueda y recuperación de información, son utilizados por los usuarios para recuperar e interactuar con la información almacenada por los sistemas. Clásicamente, este tipo de sistemas sólo pueden procesar consultas donde la información solicitada, se encuentre contenida en las fuentes de datos que utilizan y actualmente, es deseable que se enfoquen en evitar respuestas vacías. Es decir, que su prioridad sea proporcionar en la mayoría de los casos algún resultado que responda a las consultas hechas por el usuario y que dichas respuestas le sean útiles a este.

Uno de los principales problemas en los sistemas de búsqueda de información, es la precisión con la que se recupera la información después de una consulta y esto se debe a la falta de semántica en los datos. En este sentido, el presente trabajo de tesis se centra en la recuperación de información, haciendo uso de la similitud semántica para proporcionar respuestas aproximadas (las más

similares), con el fin de evitar respuestas vacías en caso de no encontrar la respuesta exacta a las consultas hechas por los usuarios.

En la actualidad el uso de bases de datos conceptuales u ontologías para representar, almacenar y organizar la información, es una fuerte tendencia en algunas áreas de investigación y en la industria, ya que permiten formalizar la información en conocimiento.

Las nuevas técnicas de recuperación de información, han permitido ir más allá y manejar conceptos en lugar de términos, agregando así un componente semántico a la información. Las medidas de similitud semántica permiten evaluar la distancia existente entre dichos conceptos (distancia semántica); siendo esto de gran utilidad en el proceso de recuperación (ya que dicha distancia representa la similitud existente entre la información solicitada por el usuario y las respuestas emitidas por el sistema), permitiendo controlar la precisión con la que se recupera información.

## **1.2. Planteamiento del problema**

El presente trabajo de investigación aborda el problema de la búsqueda y recuperación de información cualitativa y se orienta hacia la búsqueda de productos o servicios, considerando sus posiciones geográficas.

De forma más específica, la problemática a tratar consiste en recuperar todos aquellos productos o servicios similares conceptualmente a lo que un usuario está buscando, tomando en cuenta las diferentes características o cualidades de dichos productos (como criterios de búsqueda). Además de considerar la ubicación de cada producto con respecto al usuario, con la finalidad de ofrecerle resultados que se encuentren cercanos (menos distantes) a él.

Bajo este contexto, surgen las siguientes preguntas:

- ¿Cómo determinamos que productos o servicios son similares semánticamente a otros?
- ¿Cómo podemos ofrecer un servicio personalizado (a la medida) para cada usuario en particular?
- ¿Qué criterios de búsqueda son más importantes y bajo qué condiciones?

Como caso de estudio particular, se trabajara en el desarrollo de un modelo de recuperación de información orientado a la búsqueda de automóviles, considerando las características cualitativas de estos para realizar la recuperación.

## **1.3. Objetivos**

### **1.3.1. Objetivo general**

El objetivo general de esta investigación es implementar mecanismos que refinen y mejoren la búsqueda y recuperación de productos o servicios considerando su ubicación geográfica. Evitando en lo posible, retornar resultados vacíos, ofreciendo como alternativa respuestas similares o aproximadas a la información solicitada por los usuarios. Permitiendo ajustar los resultados según las preferencias de cada usuario.

### **1.3.2. Objetivos específicos**

Los objetivos específicos para cumplir con el objetivo general de este trabajo son los siguientes:

- Identificar los criterios de búsqueda involucrados con el caso de estudio.
- Construir una base de datos conceptual compuesta por jerarquías de los diferentes criterios de búsqueda identificados para el caso de estudio.
- Construir una base de datos espacial que permita almacenar y georeferenciar los distintos productos o servicios almacenados dentro del sistema de información.
- Diseñar e implementar una herramienta de recuperación de información geográfica para obtener resultados similares a lo que el usuario este buscando y que dichos resultados estén cercanos a este.
- Realizar análisis espacial sobre la base de datos espacial para determinar la distancia entre los productos buscados y la posición actual del usuario.

- Implementar un análisis semántico sobre la base de datos conceptual utilizando la medida de similitud denominada *confusión*.
- Controlar la precisión con la que se recupera información.
- Permitir a los usuarios ajustar los resultados según sus preferencias.
- Implementar un mecanismo de visualización de resultados sobre mapas digitales.

## 1.4. Motivación y justificación

Cuando una persona busca información en un sistema y no obtiene resultados (resultados vacíos), debido a que no encuentra exactamente lo que está buscando, el sistema deja de ser útil para dicha persona.

Es por eso que uno de los principales problemas en los sistemas de búsqueda y recuperación de información, consiste en desarrollar sistemas que den flexibilidad a las búsquedas y que ofrezcan resultados que van más allá de una coincidencia sintáctica, sino que se basen en comparaciones semánticas para evitar obtener resultados vacíos.

En particular las estructuras conceptuales permiten representar el conocimiento, y evaluar bajo una medida de similitud que tan cercano (similar) es un concepto a otro de acuerdo a sus significados (semántica). En este sentido, cuando una persona busca un producto o servicio es posible utilizar diferentes características cualitativas que describan a dicho producto para recuperarlo.

En la actualidad con el creciente uso de sistemas de información geográfica, la búsqueda de productos o servicios basados en la ubicación geográfica cada vez toma mayor importancia. Es por eso que ofrecer resultados que se encuentren cercanos (menos distantes) a los usuarios puede resultar de gran utilidad para estos.

El caso de estudio de este trabajo se basa en los sistemas de compra-venta de vehículos vía Web y la justificación de porque utilizar este caso de estudio, está dada por un estudio denominado Cars Online [1] realizado desde hace casi una década, donde se demuestra un constante crecimiento en este tipo de sistemas, donde en una década ha aumentado su uso en un 80% en países desarrollados.

## 1.5. Beneficios esperados

Con el presente trabajo de investigación se espera flexibilizar la búsqueda y recuperación de información haciendo uso de variables cualitativas, considerando además las componentes geográficas de la información de tal forma que los resultados arrojados por el sistema sean personalizados en función de las necesidades específicas de cada usuario.

El uso de medidas de similitud semántica para recuperar información, va más allá de la coincidencia sintáctica entre palabras, permitiendo así, incrementar el número de respuestas recuperadas después de una consulta. Lo que incrementa la probabilidad de que los usuarios queden satisfechos con las respuestas obtenidas del sistema después de una búsqueda.

## 1.6. Alcances y límites

Para la implementación del sistema sólo se consideran vehículos nuevos, vendidos por las agencias de autos en la ciudad de Culiacán, Sinaloa.

La información de los vehículos es contenida en una base de datos espacial para georeferenciar dicha información y poder realizar un análisis geoespacial, mientras que las diferentes características de los automóviles se estructuran en jerarquías con el fin de realizar un análisis semántico, lo que permite flexibilizar las consultas evitando así, obtener resultados vacíos y permitiendo controlar la precisión (similitud) con la que se recupera dicha información.

Aunque existen trabajos previos acerca de la construcción automática o semiautomática de ontologías (por ejemplo [2]), el presente trabajo de tesis no intenta realizar esta etapa de forma automatizada, pues ese no es el objetivo de esta investigación.

## 1.7. Hipótesis

El presente trabajo de tesis considera las siguientes hipótesis:

- El uso de medidas de similitud semántica en la recuperación de información geográfica permite flexibilizar las consultas realizadas por

los usuarios ofreciendo resultados que podrían ser útiles y que en otro caso no podrían ser recuperados.

- Al realizar un análisis espacial sobre la información de la base de datos es posible considerar la ubicación de los usuarios y la oferta existente de productos o servicios (automóviles en el caso de estudio particular) basados en su ubicación geográfica.
- La combinación del análisis espacial y la similitud semántica para recuperar información permite abordar este problema bajo dos perspectivas: por un lado la semántica implícita en cada búsqueda y por otro lado las componentes geográficas en la información enriqueciendo el modelo de recuperación de información.
- La ponderación de los distintos criterios de búsqueda, permite a los usuarios la personalización de resultados de acuerdo a sus preferencias.
- Por último, al implementar el modelo de recuperación de información propuesto, es posible recuperar información con un cierto grado de error y visualizar dichos resultados sobre una aplicación de Web Mapping.

## **1.8. Organización del documento**

**Capítulo 1:** En este capítulo se presenta una introducción general al trabajo de tesis, así como la motivación y justificación por los cuales se desarrolla este trabajo, también se describe el planteamiento del problema, los objetivos que se persiguen y las hipótesis que guían esta investigación.

**Capítulo 2:** En el capítulo dos se aborda el estado del estado del arte, haciendo una revisión sobre los trabajos que tienen relación con el presente trabajo de tesis con el fin de situar la presente investigación. Se hace una breve introducción al problema de semejanza, se menciona en que consiste la similitud semántica, cual es la importancia del uso de jerarquías en diversas áreas de investigación, así como, algunas medidas de similitud semántica que operan sobre jerarquías.

**Capítulo 3:** Este capítulo describe el marco teórico de la tesis, abordando los conceptos y fundamentos acerca de las ontologías y las jerarquías, como es que estas pueden ser utilizadas para buscar y recuperar información. Además se

describen las herramientas necesarias para el despliegue de resultados en herramientas GIS.

**Capítulo 4:** El capítulo cuatro detalla la metodología propuesta describiendo cada componente y mencionando como son utilizadas las herramientas descritas en el marco teórico; también se presenta la arquitectura del sistema de recuperación de información que implementa dicha metodología como propuesta de solución al problema.

**Capítulo 5:** En este capítulo se presentan y se discuten los resultados obtenidos con el sistema de recuperación de información.

**Capítulo 6:** En el capítulo seis se exponen las conclusiones obtenidas durante el presente trabajo de investigación y se mencionan algunas propuestas como trabajo futuro.

**Anexos:** Finalmente, se incluyen los códigos fuentes, los archivos OWL con las jerarquías y la información de la base de datos en los anexos del documento.

## **CAPÍTULO 2.**

# **ESTADO DEL ARTE**

En este capítulo se mencionan algunos de los trabajos previos en el marco de esta investigación, tales como: algunas definiciones de similitud, cual es la relación existente entre similitud y disimilitud. Se describe como son utilizadas las estructuras conceptuales u ontologías para representar y organizar información y se presentan algunas propuestas para el cálculo de la similitud semántica sobre estructuras conceptuales. Finalmente, se presentan algunas aplicaciones relacionadas con la búsqueda y recuperación de productos o servicios considerando sus componentes geográficas.

### **2.1. Recuperación de información**

Los sistemas de recuperación automática de información (IR por sus siglas en ingles) fueron originalmente desarrollados para manejar los grandes volúmenes de literatura científica generados a partir de la década de los cuarentas (1940) [3]. Actualmente, el uso de estos sistemas se ha extendido a diversas áreas como lo son: universidades, compañías, librerías públicas por citar algunos ejemplos. Sin embargo, en años recientes ha existido un gran incremento en los volúmenes de información manejados por los sistemas de información y las capacidades que brinda el uso de Internet y la Web han propiciado la búsqueda de mecanismos donde los usuarios puedan encontrar rápidamente información que sea relevante y útil para sus necesidades. La presente investigación está orientada a la recuperación de productos o servicios que satisfagan en el mayor grado posible las necesidades de los usuarios, considerando las componentes geográficas de la información que se desea recuperar [4].

### 2.1.1. Recuperación de información geográfica

El desarrollo y el uso creciente de los Sistemas de Información Geográfica (GIS por sus siglas en ingles), han propiciado un nuevo tipo de sistemas de Recuperación de Información, denominados sistemas de Recuperación de Información Geográfica (GIR por sus siglas en ingles).

La Recuperación de Información Geográfica puede ser vista como un área especializada de la Recuperación de Información. La cual incluye todas las áreas de investigación que tradicionalmente han constituido el núcleo de la investigación en el área de Recuperación de Información, pero con un énfasis especial en la organización, almacenamiento y recuperación de información del dominio geográfico [5].

### 2.2. Similitud

Considere que se tienen las figuras geométricas de la siguiente imagen (*Figura 2.1*), ¿Cuáles son similares? ¿Cuáles de son diferentes?

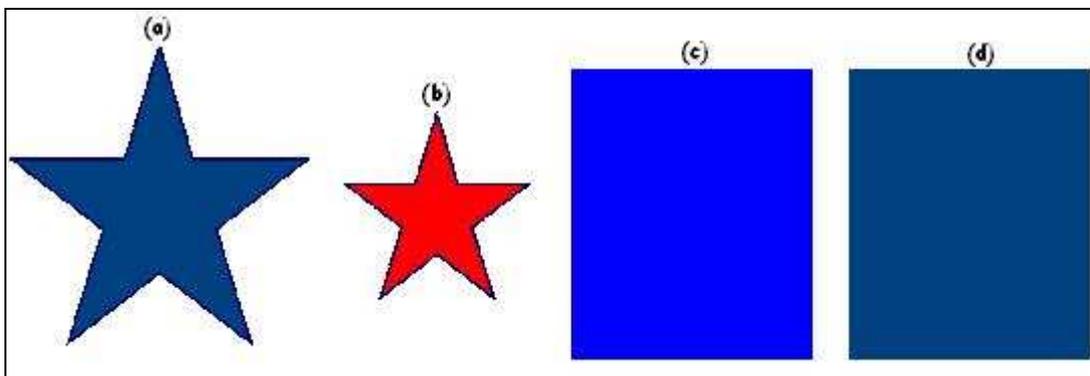


Fig. 2.1 Ejemplo de similitud en figuras geométricas.

Usted podría decir que:

- Las figuras a y b son similares pues las dos son estrellas ó diferentes pues no tienen el mismo tamaño y color.
- Que a y d son similares pues tienen el mismo color ó diferentes ya que tienen distinta forma.

Tanto la forma, el color y el tamaño son características que se pueden medir, la similitud es la cantidad que refleja la relación existente entre dos objetos o características [6]. Esta cantidad generalmente se encuentra en un rango de -1 a 1 o normalizada entre 0 y 1.

Como se puede notar con el ejemplo anterior la medir la similitud puede resultar una tarea no trivial y algunos de los primeros trabajos que abordan la problemática de cómo determinar la similitud existente entre dos objetos o entidades de información, fueron los trabajos realizados por Shepard en 1962 [7] y Tversky en 1977 [8]; dichos trabajos se basan en el uso de las características similares y diferentes que tienen los objetos a comparar.

Las entidades de información a comparar pueden ser objetos más abstractos que una figura geométrica, como lo son los conceptos y a su vez dichos conceptos pueden pertenecer a un concepto más general, formando así, una estructura de conceptos o estructura conceptual. Existen distintas estructuras que permiten almacenar y organizar conceptos, como lo son: grafos, taxonomías y árboles, por mencionar algunas.

Una de las estructuras conceptuales más utilizadas en la actualidad, para representar de manera formal un dominio son las ontologías. Específicamente en el dominio geográfico son muy utilizadas las jerarquías (caso particular de las ontologías), pues estas permiten representar y organizar de manera natural la información geográfica.

### 2.2.1. Relación entre similitud y disimilitud

Cuando la expresión para el cálculo de la similitud entre dos objetos  $a$  y  $b$  es una función  $S(a,b)$  en un intervalo de valores  $[0,1]$ , se puede decir que la expresión para el cálculo de la disimilitud es una función  $D(a,b)$  tal que:

$$D(a,b) = 1 - S(a,b)$$

De lo anterior, se puede decir que cuando:

- La similitud  $S(a,b) \rightarrow 1$  la disimilitud  $D(a,b) \rightarrow 0$
- La similitud  $S(a,b) \rightarrow 0$  la disimilitud  $D(a,b) \rightarrow 1$

### 2.2.2. Medidas de similitud

La investigación en el uso de medidas de similitud no es algo nuevo, pues se ha hecho desde hace ya varios años. En 1965 Rubenstein y Goodenough realizaron un trabajo sobre la similitud entre términos del idioma inglés, dicho trabajo se basa en la semejanza del contexto donde aparecen los términos [9]. Para validar el trabajo se realizó un experimento donde dos grupos de personas (uno de 15 y otro de 36 personas, respectivamente) evaluaban la similitud existente entre 65 pares de sustantivos. La evaluación consistía de la asignación de un valor numérico, el cual iba de 0.0 (indicando que no existía similitud) hasta 4.0 (indicando que existía similitud entre los sustantivos). Dicho experimento fue la base para que se siguieran realizando más experimentos y pruebas en el campo de las medidas de similitud.

Posteriormente en 1991, Miller y Charles [10] realizaron un experimento donde seleccionaron 38 estudiantes de grado para estimar la semejanza entre 30 pares de nombres. Esta semejanza, al igual que en el trabajo de Rubenstein y Goodenough, también se expresaba con un valor entre 0.0 y 4.0. A partir de este experimento se propusieron una serie de trabajos para medir similitud entre elementos de estructuras taxonómicas. Dichos trabajos están dirigidos hacia el uso de WordNet.

### 2.2.3. Un enfoque basado en clases para relaciones léxicas

En 1993 Resnik [11], propuso uno de los algoritmos más destacados de este apartado, el cual propone que la similitud entre dos conceptos  $c_1$  y  $c_2$  de una taxonomía, puede ser obtenida por la expresión:

$$sim(c_1, c_2) = \max_{c \in S(c_1, c_2)} (-\log p(c))$$

Donde:

- $S(c_1, c_2)$  representa el conjunto de conceptos de los que  $c_1$  y  $c_2$  descienden.

- $p(c)$  es la probabilidad del concepto  $c$ . El termino concepto hace referencia al conjunto de términos que hacen referencia a una misma idea.

Para estimar la probabilidad de un concepto  $c$ , Resnik utiliza la frecuencia de aparición de los términos de ese concepto en el Brown Corpus of American English.

#### 2.2.4. Medidas de similitud en taxonomías

En 1998, Lin [12] propone uno de los trabajos más citados en la literatura. Dicho trabajo también se basa en la probabilidad de un concepto y partiendo de la teoría de la información, se obtiene de forma teórica una expresión para el cálculo de la similitud entre dos conceptos  $c_1$  y  $c_2$  pertenecientes a una taxonomía como la de WordNet. Dicha expresión es la siguiente:

$$sim(c_1, c_2) = \frac{2 * \log p(c_3)}{\log p(c_1) + \log p(c_2)}$$

Donde  $c_3$  es el padre común de  $c_1$  y  $c_2$ .

Con la expresión anterior Lin, pretende dar una definición universal de similitud en términos de teoría de la información. Para llegar a dicha definición Lin parte de tres premisas intuitivas:

1. La similitud entre A y B se relaciona con su entorno. Entre más características comparten, más similares son.
2. La similitud entre A y B está relacionada con las diferencias que existen entre ellos. Entre más diferencias hay, más diferentes son.
3. La máxima similitud entre A y B se alcanza cuando A y B son idénticos.

### 2.2.5. Medidas de similitud en WordNet

En [13] Leacock y Chodorow proponen otra medida de similitud, donde consideran que la similitud entre  $c_1$  y  $c_2$  puede ser obtenida con la siguiente expresión:

$$sim(c_1, c_2) = -\log\left(\frac{len(c_1, c_2)}{2 * MAX}\right)$$

Donde:

- $len(c_1, c_2)$  es el número de saltos para a dar entre  $c_1$  y  $c_2$ .
- $Max$  es el valor de los saltos que se pueden dar entre el nodo raíz y los nodos hoja de la taxonomía.

En [14] Hirts y St-Onge, emplean cadenas léxicas como representación del contexto para la detección y correlación de nodos entre conceptos. En este trabajo Hirts y St-Onge se alejan del enfoque probabilístico y cuantitativo (conteo de nodos), empleando cadenas léxicas para poder calcular la similitud.

Es notable que WordNet ha sido objeto de un gran número de propuestas que existen en la actualidad y que van dirigidas al uso de estructuras taxonómicas.

### 2.2.6. Un enfoque no-probabilístico

En [15], se plantea el problema que existe entre las medidas de similitud semántica, cuando los algoritmos encargados del cálculo de la similitud no pueden ser aplicados en taxonomías donde no exista alguna forma de estimar la probabilidad de un concepto  $c$ . Si se observa el ejemplo de Resnik, en el cual se estima la probabilidad de un concepto por medio de la frecuencia de aparición de los términos asociados a este concepto dentro de un corpus significativo, como lo es el Brown Corpus of American English. Dicha probabilidad puede estimarse utilizando la siguiente expresión:

$$\tilde{p}(c) = \frac{freq(c)}{N}$$

Donde  $freq(c)$  es el número de veces que aparecen en el corpus los nombres corresponden al concepto  $c$  o a sus subconceptos, y  $N$  es el número total de nombres presentes.

Si se desea utilizar la expresión anterior, para estimar la probabilidad en taxonomías diferentes a WordNet, es necesario encontrar otra forma de estimar la frecuencia de aparición de cada concepto. Pero lo anterior genera dos problemas, en primer lugar se está asumiendo que las instancias de la estructura conceptual son estáticas, es decir, no se contemplan cambios en la estructura conceptual y en segundo lugar se deja de lado el problema del desequilibrio en la especificidad de nodos (esto, se da cuando la cantidad de instancias o nodos pertenecientes a una clase, supera en gran cantidad a los nodos o instancias pertenecientes a otra clase del mismo nivel en la estructura conceptual).

En [15], también se propone un algoritmo alternativo para abordar los dos problemas comentados anteriormente, en dicho algoritmo se suponen dos conceptos  $c_0$  y  $c_1$ , donde  $c_1$  es subclase de  $c_0$ , lo cual no implica que  $c_1$  sea hijo directo de  $c_0$  en la ontología. A continuación se presenta la expresión utilizada para calcular la similitud entre  $c_0$  y  $c_1$  (más adelante se discutirá la expresión general para calcular la similitud entre dos conceptos  $c_1$  y  $c_2$ , que no son subclases uno del otro):

$$ssim(c_0, c_1) = \frac{k + N_0}{k + N_0 + \log\left(\frac{E_0}{E_1}\right)} \quad (1)$$

Donde  $N_0$ , es el número de nodos que hay entre el nodo raíz y  $c_0$ , esto permite tener la noción de que descender un nivel en la ontología supone una mayor diferencia semántica en la parte alta de la ontología (donde están los conceptos más generales) que en la parte baja de esta (donde están los conceptos más específicos).

El cociente  $E_0/E_1$ , representa la relación entre la información que aporta  $c_0$  y la información que aporta  $c_1$ . Si se supone que el nodo raíz contiene el 100% de la información presente en la ontología, cada uno de los nodos inferiores contendrá una porción de esa información. La forma en que se modela el reparto de dicha información está relacionada con la estructura de la ontología.

Si una clase  $c$  tiene una cantidad de información  $E$ , suponga que cada una de sus subclases  $n$  tendrá una cantidad de información  $E/n$ , es decir, que la

información se va repartiendo equitativamente entre las clases hijas (claro está, decir que es un supuesto demasiado fuerte y difícil de generalizar). Con lo anterior se expresa que, cuanto menor es el número de clases en que se divide una clase, mayor es el parecido que conservan estas subclases y su clase padre.

El objetivo del diseño de la estructura conceptual es organizar la información de forma manejable. Para esto se suele crear clases intermedias evitando que una clase tenga un gran número de subclases. La agrupación de subclases intermedias se hace a partir de características cualitativas comunes, generalmente estas características se definen según el propósito de la ontología.

La variable  $k$  permite ajustar el peso que se le asigna a las variables: profundidad de  $c_0$  y tasa de información. Desgraciadamente el valor que toma  $k$  debe ser obtenida empíricamente.

A continuación se explica el caso general de la expresión (1), la cual produce como resultado un valor entre 0.0 (indicando carencia de similitud) y 1.0 (indicando la máxima similitud) de esta forma la distancia entre dos conceptos  $c_1$  y  $c_2$  se puede definir a partir de su similitud semántica  $ssim$  (1) como se muestra en la siguiente expresión:

$$dist(c_1, c_2) = \frac{1}{ssim(c_1, c_2)} - 1$$

De la expresión anterior se puede decir que:

- Si  $c_1 = c_2$  (i.e. el mismo concepto) entonces  $ssim(c_1, c_2) = 1$  y  $dist(c_1, c_2) = 0$ .
- Si  $ssim(c_1, c_2) = 0$  (i.e. conceptos totalmente distintos), entonces la distancia conceptual entre ambos será infinita.

Si  $C$  es el conjunto de conceptos que son superclase de  $c_1$  y  $c_2$ , otra forma de expresar la distancia ( $dist$ ) anterior es mediante la siguiente expresión:

$$dist(c_1, c_2) = \min_{c_0 \in C} (dist(c_0, c_1) + dist(c_0, c_2))$$

Un punto muy rescatable de este algoritmo, es que no hace ninguna suposición sobre la estructura conceptual donde se realizan los cálculos y no precisa de datos externos, ya que los cálculos dependen única y exclusivamente de la

estructura conceptual. Como consecuencia de lo anterior, esta medida de similitud puede ser utilizada en cualquier dominio.

### 2.2.7. Similitud semántica en jerarquías

Existen varios trabajos sobre como medir la similitud entre conceptos de una jerarquía, uno de los primeros trabajos y de los más simples es el propuesto en [16] y consiste simplemente en el conteo de conexiones entre un nodo y otro. Posteriormente en [17] se propone un refinamiento considerando el uso de pesos para ponderar cada uno de los enlaces de la jerarquía.

En el 2003, Genesan y Molina [18] realizaron un estudio exhaustivo sobre los modelos para medir la similitud basados en la intersección de objetos, utilizando una jerarquía que describa las relaciones entre los elementos que la componen. Para Ganesan y Molina es muy importante considerar una medida de similitud que sea intuitiva para el ser humano.

En el 2008, Sarabia [19], menciona que el *conocimiento semántico* en la jerarquía ayuda a identificar los objetos que tienen características en común y que esto beneficia al mejoramiento de las medidas de similitud.

En la Figura 2.2 muestra una jerarquía que representa diferentes géneros de música.

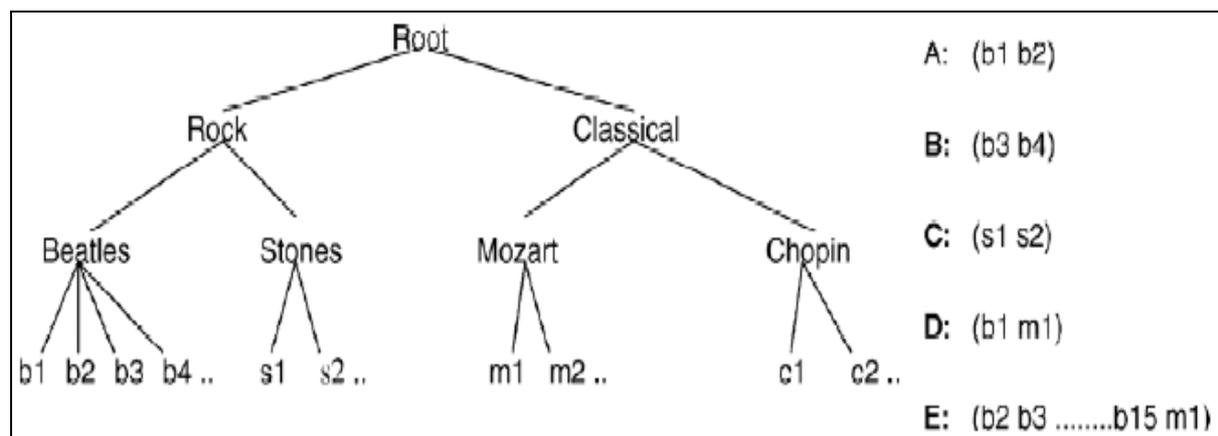


Fig. 2.2 Jerarquía de CD's utilizada en [18].

En la jerarquía anterior se suponen distintos usuarios (A, B, C, D, E) que compran un par de discos compactos respectivamente y se pretende determinar qué usuario es más similar a otro dado el género de música que estos compran.

La jerarquía de la *Figura 2.2* muestra que:

- El cliente A compra dos CD's de los Beatles (b1, b2).
- Que el cliente B compra (b3 y b4).
- El cliente C compra (s1 y s2)
- Finalmente el cliente D lleva (b1 y m1).

Como se puede observar el cliente A y B son bastante similares ya que ellos adquirieron dos discos de los Beatles, mientras que el cliente A y C son menos similares, ya que a pesar de escuchar música rock estos prefieren diferentes agrupaciones musicales. En este sentido, Ganesan y Molina se enfocan al desarrollo de medidas de similitud que se aproximen a la forma en que se supone, opera la intuición humana.

Los siguientes puntos, resumen las contribuciones del trabajo de Ganesan y Molina [18]:

- Se introducen medidas de similitud que aprovechan la estructura jerárquica de un dominio, proporcionando resultados que son intuitivamente más similares que los generados por las medidas tradicionales de similitud.
- Se consideran múltiples ocurrencias al medir la similitud entre elementos de la jerarquía.
- Se realizan comparaciones con las medidas de similitud que no explotan una jerarquía de dominio.

Las jerarquías a menudo son empleadas para estructurar la información y se han utilizado en diversas tareas como son: la clasificación de textos, la recuperación interactiva de información, entre otras tareas; donde la similitud juega un papel importante, ejemplo de esto son los trabajos: [20], [21].

Particularmente para esta investigación las jerarquías nos permiten estructurar y almacenar de manera práctica la información de un dominio específico, facilitando la recuperación de información que es similar a la que un usuario está buscando.

La *Figura 2.3* muestra la evolución que han tenido las medidas de similitud, según el criterio utilizado para medir dicha similitud.

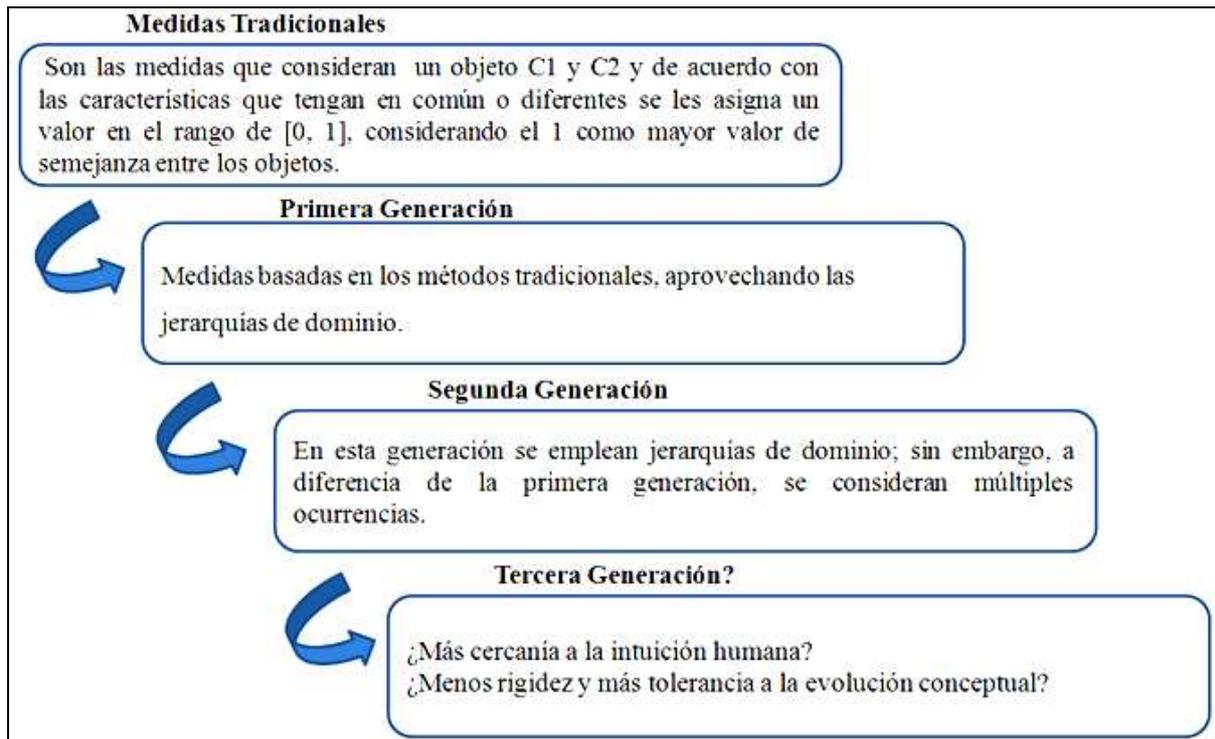


Fig. 2.3 Evolución de las medidas de similitud según Genesan y Molina.

## 2.3. Búsquedas semánticas

Actualmente en la Web el acceso a la información puede ser una tarea difícil y frustrante debido a diversos problemas habituales en la búsqueda de información. Idealmente la Web Semántica es una Web extendida basada en el significado, donde cualquier usuario en Internet podrá encontrar respuestas a sus consultas de forma más rápida y sencilla.

En [22] Berners Lee, menciona cuatro componentes o características básicas necesarias para la evolución de la Web Semántica. Estos componentes son los siguientes:

- **Expresar significado.** La Web Semántica debe brindar una estructura y añadir semántica al contenido de las páginas web, creando un entorno donde agentes de software puedan viajar de una página a otra, llevando a cabo sofisticadas tareas para los usuarios.
- **Acceso a representaciones del conocimiento.** La Web Semántica debe encargarse de resolver las limitaciones de los sistemas de representación de conocimiento tradicionales, creando lenguajes de reglas

suficientemente expresivos como para permitir a la Web razonar tan ampliamente como se desee.

- **Ontologías.** Para conseguir que los computadores sean mucho más útiles, la Web Semántica extiende la Web actual con conocimiento formalizado y datos que son procesados por máquinas. Para ser capaz de buscar y procesar información relativa a alguna materia de interés, los programas necesitan información que haya sido modelada de una forma coherente.
- **Agentes.** La potencia real de la Web Semántica se conoce cuando agentes capaces de manejar contenido semántico son usados para recoger y procesar información Web e intercambiar los resultados con otros agentes.

En la Web Semántica la información cuenta con un significado bien definido, se mejora la forma en la que las máquinas y las personas trabajan en conjunto [22]. En [23] se puede encontrar otra definición de la Web Semántica.

La Web Semántica busca superar las limitaciones existentes para la Web en la actualidad, mediante la introducción de descripciones explícitas de significado [22]. Se espera construir una Web más inteligente, en la cual se pueda conseguir una comunicación efectiva entre computadoras. Los esfuerzos de la Web Semántica están enfocados principalmente a la búsqueda de descripciones que enriquezcan semánticamente los datos que se encuentran en la Web, con esto se prometen mejores métodos de recuperación.

### 2.3.1. Web semántica geoespacial

El desarrollo de la Web Semántica requiere de atención especial en el dominio geográfico. La Web Semántica Geoespacial será un importante avance en el uso de la información geoespacial. Con la incorporación de la semántica geoespacial en la Web, la recuperación de información se hará precisa al nivel en que los resultados serán inmediatamente útiles.

La idea primordial es poder elaborar ambientes donde sea posible realizar búsquedas para ubicar lugares dejando de usar sólo palabras clave. En el dominio geoespacial existen varios desafíos y problemas por atacar, entre ellos se encuentran:

- Mejores formas para representar la información geográfica.
- Integración automática de información.
- Consultas con operadores espaciales.

Algunas de las aplicaciones posibles en el campo de la Web Semántica Geoespacial 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 Semántica Geoespacial.

El principal interés en la semántica geoespacial está centrado en la mejora de los resultados en las búsquedas e interoperabilidad de los datos espaciales.

El Consorcio Universitario para las Ciencias de Información Geográfica (UCGIS – por sus siglas en inglés) ha tomado el desafío de profundizar la investigación de la Web Semántica Geoespacial y existen varios autores que han tomado parte en esta área, tal es el caso de Egenhofer. En [24], se describe la necesidad de la semántica en los datos al realizar búsquedas en la web.

## **2.4. Aplicaciones relacionadas**

### **2.4.1. Búsqueda de productos o servicios con herramientas GIS**

En esta sección se presentan algunas aplicaciones relacionadas con la búsqueda y recuperación de información que vinculan el uso de herramientas GIS para desplegar sus resultados.

### 2.4.1.1. Google maps

Recientemente la compañía Google a través de Google maps (<http://www.googlemaps.com>) permite la búsqueda y localización de empresas, direcciones y lugares de interés, como se puede ver en la *Figura 2.4*.

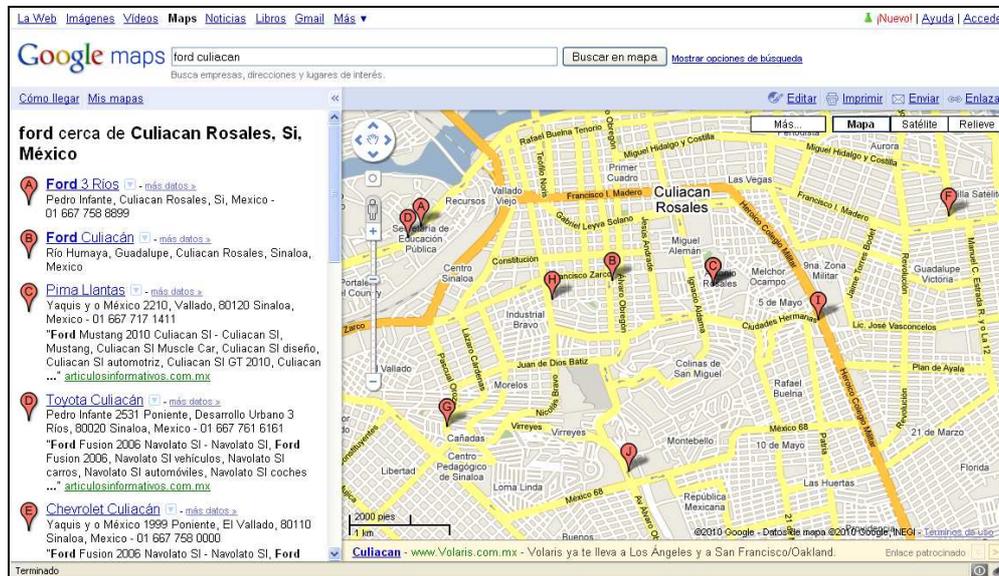


Fig. 2.4: Google maps.

### 2.4.1.2. Nestoria

La Nestoria ([www.nestoria.com](http://www.nestoria.com)) es un sitio Web para la compra y/o renta de inmuebles en algunos países de Europa, que permite la búsqueda y localización de inmuebles vía Web, ubicando los resultados sobre mapas de *Google maps*.



Fig. 2.5: Nestoria.

## 2.4.2. Aplicaciones GIR

En esta sección se mencionan algunos trabajos de investigación relacionados con la búsqueda y recuperación de información geográfica.

### 2.4.2.1. iRank

En 2009, Mata propone abordar la recuperación y ponderación de información geográfica desde repositorios no estructurados utilizando ontologías. En [25] se asume que los procesos de recuperación pueden mejorarse si se integran las relaciones geográficas y topológicas asociadas a los objetos, extraídas desde fuentes de información heterogéneas, para luego ponderar los resultados de carácter topológico, geográfico y semántico mediante un método denominado *iRank*, que a su vez retoma ideas del uso de medidas de similitud en estructuras conceptuales.

La propuesta para evaluar la relevancia de una entidad, en este caso un documento DG, dada una consulta QG es calculado mediante la siguiente expresión:

$$RelInt = Re \frac{ICon(Cq, Cd) + RelGeo(Gq, Gd) + RelTopology(Tq, Td)}{\text{Número de fuentes geográficas}}$$

### 2.4.2.2. Análisis espacial en espacios conceptuales

En [26] se aborda la problemática de donde es el mejor lugar para ubicar un negocio, considerando una serie de características cualitativas de los inmuebles, así como las relaciones topológicas, para realizar un análisis y funcionar como un sistema de soporte para la toma de decisiones (DSS – por sus siglas en ingles).

En este trabajo se hace uso de jerarquías y de medidas de similitud para determinar cuál es el lugar que más se parece a los requerimientos del usuario, según un perfil de usuario.

### 2.4.2.3. GEOSESYMU

En 2009, Rentería [27] propone abordar la búsqueda recuperación de información del dominio del turismo geo-cultural. En este trabajo se hace uso de jerarquías para recuperar museos ubicados en el centro histórico de la ciudad de México, generando un ranking de resultados y desplegando dichos resultados sobre un mapa.

### 2.4.3. Sistemas de compra-venta de vehículos vía Web

Dos de los principales sitios Web dedicados a la comercialización de vehículos por internet en México son Mercado libre ([www.mercadolibre.com.mx](http://www.mercadolibre.com.mx)) y Yahoo! Autos ([www.autocosmos.com](http://www.autocosmos.com)). Desgraciadamente estos sistemas son incapaces de ofrecer resultados similares a las consultas que un usuario realiza, llegando a producir errores como el que se muestra en la *Figura 2.6*, donde se muestra una pantalla de error, al no encontrar un vehículo buscado por un usuario.

En México, existen pocas aplicaciones dedicadas a la compra venta de vehículos por internet que confluyan con las herramientas GIS disponibles en la actualidad.



Fig. 2.6: Error: “no match found” en Yahoo! Autos

## 2.5. Comentarios finales

En este capítulo se describen trabajos relacionados con la búsqueda y recuperación de información, se analizan diversos trabajos sobre como evaluar la similitud desde distintos enfoques y como ha sido la evolución de estas medidas. Especialmente se revisan algunos trabajos sobre el cálculo de la similitud semántica en estructuras conceptuales como son ontologías, taxonomías y jerarquías. Estas últimas son de gran interés en la presente investigación, pues además de permitir almacenar y estructurar la información de un dominio específico, facilitan la búsqueda y recuperación de información haciendo uso de medidas de similitud.

Las medidas de similitud semántica toman cada vez más importancia en los sistemas de recuperación de información, pero es necesario adaptar estas medidas para que operen de manera adecuada en dominios que se vinculen al dominio geográfico. Además el uso de medidas de similitud semántica permite ir más allá de la comparación sintáctica de conceptos y utilizar otras características que permitan recuperar resultados que por sus atributos o propiedades similares sean relevantes.

## CAPÍTULO 3.

# MARCO TEÓRICO

En este capítulo se mencionan los principios que sustentan la presente investigación, así como, los conceptos utilizados en esta. Se profundiza en los conceptos básicos de la recuperación de información y específicamente de información geográfica. Se justifica el uso de bases de datos espaciales, así como el análisis espacial y el procesamiento de datos. También se aborda el uso de las ontologías, como una alternativa para estructurar y procesar la información. Se hace énfasis especial en el uso de jerarquías y finalmente se menciona en que consiste la teoría de confusión; utilizada como medida de similitud semántica sobre jerarquías.

### **3.1. Recuperación de información geográfica**

La Recuperación de Información Geográfica puede ser vista como una rama especializada de la recuperación de información tradicional. Incluye todas las áreas de investigación que tradicionalmente han constituido el núcleo de la investigación sobre recuperación de información, pero además esta hace un énfasis especial en la indexación espacial y recuperación geográfica [28].

La Recuperación de Información Geográfica no se limita la recuperación de objetos con una manifestación física en el espacio geográfico ó alguna extensión geográfica como pueden ser los ríos, ciudades, lagos o los países. Sino que abarca cualquier tipo de información que tenga algún tipo de relación

con uno o varios lugares en la superficie espacial. A la información que conserva este tipo de relaciones con el espacio geográfico, se le llama a menudo información georeferenciada y por lo tanto un término de uso frecuente para el campo es de Recuperación de Información Georeferenciada.

La Información Georeferenciada se encuentra en todo tipo de medios de comunicación, no sólo datos estructurados como mapas, planos topográficos, observaciones aéreas e imágenes satelitales y tabulados. El enfoque de este trabajo de tesis está orientado a la indexación y recuperación de productos o servicios teniendo en cuenta su ubicación geográfica.

### **3.2. Información georeferenciada**

Mucha de la información disponible en bibliotecas digitales y en Internet es georeferenciada, aunque la mayoría no se representa en términos de coordenadas geográficas. A menudo, el vínculo entre la información de un documento de texto y la ubicación que se refiere en la información se manifiesta como nombres de lugares o frases que sugieren una localización geográfica. A la ubicación geográfica y la extensión del nombre de un lugar a menudo se le llama una presencia geográfica [29] y está dada por las coordenadas (latitud, longitud). La Recuperación de Información Geográfica requiere que los nombres de lugares y frases que incluyen referencias directas o indirectas a las toponimias se resuelvan y se traduzcan en información que pueda ser almacenada e indexada. Las bases de datos espaciales son una buena alternativa para almacenar e indexar información con contenido geográfico, existen diversos manejadores de bases de datos (DBMS – Por sus siglas en inglés) que cuentan con la capacidad de almacenar datos espaciales, algunos de forma nativa (MySQL, MS SQL Server y Oracle, por mencionar algunos). Y otros mediante el uso de algún Plug-in como lo es (Postgres SQL).

### **3.3. Bases de datos espaciales**

Cuando se manejan grandes volúmenes de datos espaciales es recomendable almacenarlos en una base de datos espacial. Una base de datos espacial es una colección de datos georeferenciados organizados de tal forma que sirvan para una o varias aplicaciones GIS. Esta base de datos comprende la asociación entre

sus dos principales componentes: datos espaciales y atributos o datos no espaciales [30].

Al igual que las bases de datos no espaciales, las bases de datos espaciales conservan las siguientes características [31]:

**Velocidad:** Las Bases de datos están optimizadas para servir a grandes volúmenes de datos repetitivos, y los datos espaciales se ajustan perfectamente a esta descripción.

**Soporte a múltiples usuarios:** Los datos espaciales tienden a ser datos de referencia, y los datos de referencia generalmente están destinados a ser compartidos por múltiples usuarios. Almacenar los datos espaciales en una base de datos espacial, otorga beneficios tales como el acceso de muchos usuarios a la vez y que estos puedan acceder a los datos de forma remota, a través de una interfaz estándar (JDBC, ODBC, Perl/DBI y otros). Las bases de datos también le permiten aumentar la seguridad de los datos.

**Consultar:** Este es con mucho el mayor beneficio. Así como las bases de datos tradicionales le permiten realizar consultas tradicionales, las bases de datos espaciales permiten realizar consultas espaciales tales como: "muéstrame todos los puntos que caen dentro de un radio de 20 millas de este punto"). Este tipo de consultas implican un análisis espacial y un procesamiento de los datos.

### 3.4. Análisis espacial y procesamiento de datos

Con frecuencia se hace referencia al análisis espacial, como modelado. Refiriéndose al análisis de fenómenos que poseen dimensiones físicas (la ubicación de, cercano a, o la orientación de un objeto respecto a otro; relación con un área de un mapa referenciado o relacionado a una ubicación específica en la superficie del planeta). El análisis espacial a veces también es visto como un modelo de interpretación de resultados útiles para evaluación de disponibilidad, capacidad, estimación o predicción.

En los GIS, se puede decir que existen cuatro tipos de análisis espacial: análisis por sobre posición y contigüidad, análisis de superficies, análisis lineal, y análisis raster. Este incluye funciones GIS como sobre posición topológica, generación de buffers, y modelado de redes [25].

### **3.5. Ontologías como una alternativa para estructurar y procesar información**

Existen diversas definiciones de ontología con diferentes enfoques, pero el término ontología tiene sus orígenes en la filosofía, siendo en esta área del conocimiento donde se utilizó el concepto de ontología por primera vez.

El uso del concepto ontología se originó a partir de la noción de Aristóteles, que por un lado estudiaba la esencia del ser y las características básicas de la realidad como un todo. En la actualidad, es considerada una rama de la Filosofía que tiene por objetivo la explicación de la existencia de una manera sistemática, trata sobre los tipos y estructuras de objetos, propiedades, eventos, procesos y relaciones relativas a cada porción de la realidad [32].

A partir de la década de los noventa, las ontologías comienzan a tener una alta relevancia en el campo de la computación, específicamente en el área de Inteligencia Artificial (IA), con especial énfasis en la Ingeniería del Conocimiento. En [33] se menciona que la Inteligencia Artificial se ocupa del razonamiento acerca de los modelos del mundo. Lo que conlleva a que, investigadores del área adopten el término “ontología” para describir lo que puede ser representado y utilizado de manera computacional por un programa.

Con lo anterior, se puede decir que las ontologías representan una alternativa para estructurar información que será procesada y utilizada posteriormente por sistemas de información.

#### **3.5.1. Las ontologías como una conceptualización**

Como se mencionó anteriormente (punto 3.2), existen diferentes definiciones de ontología, algunas de las principales definiciones son las siguientes:

En [34] se define una ontología como los términos básicos y las relaciones que comprende el vocabulario de un área temática, así como las reglas para la combinación de términos y relaciones para definir la extensión del vocabulario. En esta definición, los autores conciben las ontologías como un recurso dinámico, en la medida que las ontologías se constituyen de términos creados a partir de reglas, además de los términos que son explícitamente definidos. Los

autores también proponen los siguientes elementos para la elaboración de ontologías:

1. Identificar la terminología básica y las relaciones que se dan entre los términos.
2. Identificar las reglas que permitan combinarlos.
3. Proveer las definiciones correspondientes a los términos y sus relaciones.

Por su parte, Thomas Gruber [35] dice que una ontología es una especificación explícita de una conceptualización. Y considera que una conceptualización está compuesta por objetos, conceptos y otras entidades que existen en una determinada área, y las relaciones que se dan entre ellos. Cuando el conocimiento de un dominio es representado en un formalismo declarativo, el conjunto de objetos que pueden ser representados es denominado universo de discurso. Este conjunto de objetos, y las relaciones entre ellos, se reflejan en el vocabulario con el que se representa el conocimiento.

En una ontología los nombres de las entidades en el universo de discurso se asocian mediante definiciones, con un texto legible por el hombre que las describe y con axiomas que restringen su interpretación y otorgan la característica de bien formados a dichos términos. Formalmente, una ontología es la declaración de una teoría lógica.

En 1995, Guarino y Giaretta [36], realizaron un estudio en el que recopilan siete definiciones donde el concepto de ontología tiene distintas interpretaciones, dichas definiciones son las siguientes:

1. La ontología como una disciplina filosófica.
2. La ontología como un sistema conceptual informal.
3. Ontología como una cuenta formal de la semántica.
4. Ontología como una especificación de una conceptualización
5. Ontología como una representación de un sistema conceptual de la teoría por medio de la lógica:
  - a. Caracterizado por propiedades formales específicas.
  - b. Caracterizada por sus fines específicos.
6. Ontología como el vocabulario utilizado por una teoría lógica.
7. Ontología como una meta de la especificación de una teoría lógica.

Las definiciones recopiladas por Guarino y Giaretta, exceptuando la primera definición, pueden ser catalogadas en dos grupos:

1. Las que conciben las ontologías como un marco conceptual a nivel semántico (definiciones 2 y 3).
2. Las que la conciben las ontologías como un objeto concreto a nivel sintáctico, con un uso guiado por un propósito específico (definiciones 4, 5, 6 y 7).

Para Guarino una ontología es un artefacto de ingeniería, construido por un vocabulario específico utilizado para describir una cierta realidad, más un conjunto de hipótesis explícitas en relación con el significado de las palabras del vocabulario [37].

En 1996, Bernaras propone que una ontología proporciona los medios para describir explícitamente lo que hay detrás de la conceptualización del conocimiento [38]. Bernaras hace referencia a la ontología como recurso destinado a representar el conocimiento en una base de conocimiento.

En [39] se define una ontología como un conjunto estructurado de términos que describen algún dominio en particular. La idea es que una ontología proporciona la estructura para construir una base de conocimientos y que dicha base de conocimientos puede ser utilizada para compartir información entre distintos sistemas.

En [40] la ontología es el término utilizado para referirse a la comprensión común de algún dominio de interés y puede ser usado como un marco unificador para resolver problemas. Una ontología implica necesariamente una especie de visión del mundo con respecto a un determinado dominio. La visión del mundo se suele concebir como un conjunto de conceptos, sus definiciones e interrelaciones; esto es lo que se denomina conceptualización. Los autores insisten sobre la naturaleza explícita de la representación que la ontología aporta sobre una conceptualización.

En 1997, Borst reformula la definición de Gruber precisando que las ontologías son definidas como la especificación formal de la conceptualización compartida. El aporte de dicha precisión es incluir la calificación de compartida a la noción de conceptualización, cuando en el carácter de compartido subyace una de las razones de la existencia de las ontologías como recursos [41].

Studer considera que, de las múltiples definiciones dadas sobre el término ontología, la de [35] y la de [41] son las que logran captar lo fundamental acerca del concepto ontología, y propone la siguiente definición: “*Una ontología es*

*una especificación explícita formal de una conceptualización compartida*” [33]. Donde el término *Conceptualización* se refiere a un modelo abstracto de algún fenómeno en el mundo proveniente de haber identificado los conceptos relevantes de dicho fenómeno. Mientras que *Explícita* hace referencia al tipo de concepto usado y las restricciones para su uso son explícitamente definidos. *Formal* se refiere al hecho de que la ontología debería ser legible por una máquina. Y *Compartida* refleja la noción de que una ontología captura conocimiento consensual, que no es objeto de un solo individuo, sino aceptado por un grupo.

Como se muestra en este apartado, existen diversas interpretaciones sobre el concepto ontología pero de manera general, se puede decir que una ontología es una descripción formal y consensada del conocimiento de un dominio específico. El conocimiento es representado a partir de conceptos, haciendo uso de un lenguaje. Dicho lenguaje también es utilizado para definir las entidades, clases, propiedades, relaciones y restricciones que forman parte del dominio.

De esta forma al ser las ontologías una descripción formal del conocimiento de un dominio específico, éstas permiten solucionar problemas de interoperabilidad entre sistemas de información y facilitan el procesamiento basado en la semántica (significado) de la información que éstas almacenan.

### 3.6. Jerarquías

El concepto de jerarquía no es nuevo. Las jerarquías han sido usadas en diferentes áreas, tales como: data warehousing y minería de datos, entre otras. Un ejemplo práctico de las jerarquías en el procesamiento simbólico es Clasitex [42], que considera los temas de un artículo escrito en español o inglés. Se utiliza el concepto de árbol, y una palabra que no está en el árbol, sugiere el tema de uno o más conceptos en el árbol.

BiblioDigital © [43], es un desarrollo reciente, que utiliza una gran taxonomía (aunque no una jerarquía) para clasificar los documentos de texto, un rastreador distribuido que recupera documentos externos que residen en otros lugares de la Web. Si se trata un documento sobre la guerra de Irak y el presidente Bush, su URL será almacenada en los tres nodos del árbol concepto.

Las jerarquías son más simples que las ontologías pero muy útiles [44]. Para comprender de manera más precisa y formal el concepto de jerarquía, analizaremos las siguientes definiciones de [45]:

**Conjunto de elementos:** Un conjunto  $E$  cuyos elementos se definen explícitamente. Ejemplo: {rojo, azul, verde, naranja, amarillo}.

**Conjunto ordenado:** Un conjunto de elementos cuyos valores están ordenados por la relación “menor que” ( $<$ ). Ejemplo: {muy frío, frío, templado, tibio, caliente, muy caliente}.

**Cubrimiento:**  $K$  es un cubrimiento del conjunto  $E$  si  $K$  es un conjunto de subconjuntos  $e_i \subset E$  tal que  $\cup_i e_i = E$ .

Cada elemento de  $E$  está en algún subconjunto  $e_i \in K$ . Si  $K$  no es un cubrimiento de  $E$ , podemos hacerlo añadiendo un nuevo  $e_j$ , denominado “otros”, que contiene todos los elementos restantes de  $E$  que no pertenecen a ningún  $e_i$  previo.

**Conjunto exclusivo:**  $K$  es un conjunto exclusivo si  $e_i \cap e_j = \emptyset$ , para cada  $e_i, e_j \in K$ .

Sus elementos son mutuamente exclusivos. Si  $K$  no es un conjunto exclusivo, podemos hacerlo reemplazando cada par que se traslape  $e_i, e_j \in K$  con tres:  $e_i - e_j$ ,  $e_j - e_i$  y  $e_i \cap e_j$ .

**Partición:**  $P$  es una partición del conjunto  $E$ , si  $P$  es cubrimiento y conjunto exclusivo de  $E$ .

**Variable cualitativa:** Una variable uní-valuada que toma valores simbólicos.

Su valor no puede ser un conjunto de variables, atributos y propiedades intercambiables. Por simbólico entendemos que es cualitativo, opuesto a numérico, o variables cuantitativas.

**Valor simbólico:** Un valor simbólico  $v$  representa al conjunto  $E$ , descrito por  $v \propto E$ , si  $v$  puede ser considerado un nombre o una representación de  $E$ .

**Jerarquía:** Para un elemento del conjunto  $E$ , una jerarquía  $H$  de  $E$  es otro conjunto de elementos, donde cada  $e_i$  es un valor simbólico que representa a

cualquier elemento de  $E$  o una partición, y  $\cup_i \{r_i | e_i \in r_i\} = E$  (La unión de todos los conjuntos representados por los  $e_i$  de  $E$ ).

Suponga una **Jerarquía H1**, Para un conjunto  $E = \{\text{Canada, USA, México, Cuba, Puerto_Rico, Jamaica, Guatemala, Honduras, Costa_Rica}\}$ , la jerarquía H1 es  $\{\text{Norte_America, Islas_Caribe, America_Central}\}$ , donde  $\text{Norte_America} \in \{\text{Canada, USA, México}\}$ ;  $\text{Islas_Caribe} \in \{\text{Islas_Hispano-parlantes, English_Speaking_Island}\}$ ;  $\text{Islas_Hispano-parlantes} \in \{\text{Cuba, Puerto_Rico}\}$ ;  $\text{English_Speaking_Island} \in \{\text{Jamaica}\}$ ;  $\text{America_Central} \in \{\text{Guatemala, Honduras, Costa_Rica}\}$ .

Las jerarquías facilitan la comparación de los valores cualitativos que pertenecen a ella.

**Variable Jerárquica.** Una variable jerárquica es una variable cualitativa cuyos valores pertenecen a la jerarquía (El tipo de dato de una variable jerárquica es jerarquía). Ejemplo: `lugar_de_origen` toma valores de H1.

**Definición:** Se utilizará la siguiente notación: a) **padre\_de** ( $v$ ). En un árbol que representa una jerarquía, el nodo **padre\_de** un nodo es aquel del cual desciende; b) los **hijos\_de** ( $v$ ) son los valores que cuelgan de  $v$ . Los nodos con el mismo padre son hermanos; c) **abuelo\_de**, **hermano\_de**, **tío\_de**, ascendentes, descendientes... son definidos, cuando ellos existen; d) La raíz es el nodo que no tiene padre.

### 3.6.1. Tipos de jerarquías

Una jerarquía describe la estructura de valores cualitativos en un conjunto  $E$ . A continuación se definen los tipos de jerarquías:

#### 3.6.1.1. Jerarquía simple

Una jerarquía simple (normal) 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:* ser viviente {animal {mamífero, pez, reptil, otro animal}, planta {árbol, otra planta}}.

### 3.6.1.2. Jerarquía ordenada

En una jerarquía ordenada, los nodos de algunas particiones obedecen una relación de orden.

*Ejemplo:* objeto {diminuto, pequeño, mediando, grande}.

### 3.6.1.3. Jerarquía porcentual

En una jerarquía porcentual, el tamaño de cada conjunto es conocido.

*Ejemplo:* Continente Americano (740M) {Norteamérica (430M) {USA (300M), Canadá (30M), México (100M)} América Central (10M), Sudamérica (300M)}.

### 3.6.1.4. Jerarquía mixta

Una jerarquía mixta combina los tres casos anteriores (3.3.1.1 - 3.3.1.3).

## 3.7. Teoría de confusión

La *Teoría de Confusión* consiste en modelar la disimilitud existente entre una o varias jerarquías empleando una medida asimétrica denominada *confusión* y se define como el grado de error que existe entre un concepto u otro.

El término *confusión* fue introducido para diferenciarlo de otros enfoques que utilizan diferentes tipos de distancias (i.e. medidas simétricas e independientes del contexto).

La propiedad asimétrica de *confusión* se da por su definición y su dependencia al contexto dentro de una estructura jerárquica.

El concepto de *confusión* permite definir qué tan cercanos son los objetos que cumplen con un predicado, así como la derivación de otras operaciones y propiedades sobre los valores jerárquicos.

### 3.7.1. Cálculo de confusión en jerarquías simples

**Definición:** Si  $r, s \in H$ , entonces la *confusión* de usar  $r$  en vez de  $s$ , que se denota como  $conf(r, s)$ , es:

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

Para medir  $conf(r, s)$  se cuentan los enlaces *descendentes* de  $r$  a  $s$ , al valor remplazado.

### 3.7.2. Cálculo de confusión en jerarquías ordenadas

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

- $conf'(r, r) = conf(r, cualquier\_ascendente\_de\_r) = 0$ ;
- Si  $r$  y  $s$  son hermanos distintos,  $conf'(r, s) = 1$  si el padre no está en un conjunto ordenado; entonces  $conf'(r, s)$  es igual a la distancia relativa de  $r$  a  $s$  y a la vez es igual al número de pasos requeridos para llegar de  $r$  a  $s$  en el orden definido, dividido entre la cardinalidad-1 del padre; (regla 3)
- $conf'(r, s) = 1 + conf'(r, padre\_de(s))$

Esto es como  $conf$  para las *jerarquías constituidas por conjuntos*, excepto que allí el error entre dos hermanos es 1, y aquí es un número  $\leq 1$ .

*Ejemplo:* Temperatura = {congelado, frío, normal, tibio, caliente, ardiendo};  
 en esta Lista  $conf'(congelado, frío) = \frac{1}{5}$ , mientras que la  
 $conf'(congelado, ardiendo) = \frac{5}{5} = 1$ .

### 3.7.3. Cálculo de confusión 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  $s$  ( $conf''(r, s)$ ), es:

- $conf''(r, r) = conf''(r, s) = 0$ , cuando  $s$  es cualquier ascendente de  $r$ ;
- $conf''(r, s) = 1 -$  proporción relativa de  $s$  en  $r$ .

### 3.7.4. Cálculo de confusión en jerarquías mixtas

**Definición.** Para calcular la  $conf'''(r, s)$  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.

## 3.8. Comentarios finales

En este capítulo se abordaron los principios de la recuperación de información geográfica, se explicó el modelo de datos geográficos y como estos pueden ser georeferenciados y almacenados utilizando bases de datos espaciales. Se mencionan cuales son los principales tipos de análisis espacial en el área de

GIS. Se abordan distintas definiciones del concepto ontología, se perfila a las ontologías como una base de conocimientos, se detallan los principales componentes de estas y cuáles son los principales pasos para definir y crear una nueva ontología. Posteriormente se menciona la definición de una jerarquía según Levashkin y Guzmán, partiendo de la teoría propuesta por estos, denominada “*Teoría de Confusión*” [45], se define a *confusión* como una medida de disimilitud asimétrica y dependiente del contexto de la estructura jerárquica donde es almacenada la información. Finalmente, se presentan los diferentes cálculos para obtener la *confusión* en jerarquías simples, ordenadas, porcentuales y mixtas.

## CAPÍTULO 4.

# METODOLOGÍA

En este capítulo se aborda la metodología desarrollada en el presente trabajo de investigación, la cual describe un modelo que permite la búsqueda y recuperación de productos o servicios similares a los que un usuario está buscando, considerando las posiciones geográficas de dichos productos. Dicha metodología consta de los siguientes bloques principales, cada uno de los cuales está constituido por diversos módulos, como se muestra a continuación:

1. Conceptualización del dominio.
  - a. Selección de características.
  - b. Construcción de la base de datos espacial.
  - c. Construcción de la base de datos conceptual.
2. Modelo de búsqueda y recuperación de información.
  - a. Evaluación de criterios.
  - b. Ordenación de resultados.

La metodología propuesta considera diferentes características o cualidades de los productos o servicios, las preferencias de los usuarios y la ubicación de los productos con respecto a los usuarios, para realizar una recuperación de información personalizada en función de las particularidades de cada usuario.

La presente investigación se centra en la búsqueda y recuperación de información geográfica. Y como se mencionó en el capítulo 1, la metodología propuesta es ejemplificada mediante un caso de estudio particular en un

dominio específico, como son los sistemas de compra-venta de vehículos en línea (vía Web). Cabe mencionar, que dicha metodología no se limita a este caso de estudio, sino que es expandible y aplicable a otros dominios (es decir, puede ser aplicada para recuperar información sobre otros tipos de productos o servicios).

Adicionalmente a la metodología, en este capítulo también se describe la arquitectura propuesta para implementar el modelo de recuperación de información haciendo uso de herramientas de software libre.

## 4.1. Conceptualización del dominio

Para que una computadora pueda procesar información, es necesario estructurarla previamente, por eso este bloque pretende estructurar la información de forma tal que sea procesable por una computadora. Las ontologías ofrecen un marco de trabajo que permite estructurar y describir de manera formal la información un dominio específico.

Con la conceptualización del dominio se busca estructurar la información, formalizando su representación; para hacer esto es necesario realizar una abstracción de la problemática del mundo real que se está abordando, modelando y captando las características importantes del problema, así como su entorno. Como se mencionó anteriormente la problemática abordada en esta investigación, es la búsqueda y recuperación de productos o servicios considerando sus coordenadas geográficas.

La *Figura 4.1* muestra el proceso propuesto para la conceptualización del dominio, dicho proceso está dividido en los siguientes módulos:

1. *Selección de características* que serán utilizadas en el proceso de búsqueda y recuperación.
2. *Construcción de la Base de Datos Conceptual (BDC)*.
3. *Construcción de la Base de Datos Espacial (BDE)*.

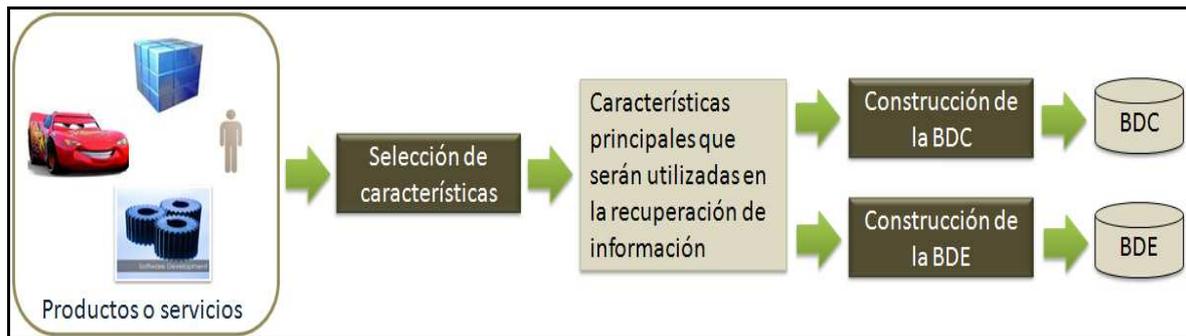


Fig. 4.1. Conceptualización del dominio.

### 4.1.1. Selección de características

El proceso de selección de características, consiste en identificar que características serán consideradas en el modelo de recuperación de información. Este proceso puede ser manual o automático, utilizando técnicas de inteligencia artificial y minería de datos, tales como extracción o selección de características (Feature Selection en inglés) [47].

Para el caso de estudio de los sistemas de compra-venta de automóviles vía Web, se han identificado las siguientes características de interés para el proceso de búsqueda y recuperación de información:

1. Clasificación del vehículo.
2. Tracción del vehículo.
3. Cilindrada del motor.
4. Tipo de combustible utilizado.
5. Tipo de dirección del vehículo.
6. Transmisión.
7. Configuración del motor.
8. Aspiración del motor.
9. Interiores del vehículo.
10. Nivel de equipamiento de los vehículos.

Las características anteriores, son los principales atributos técnicos que permiten diferenciar a un automóvil de otro. Adicionalmente a dichas características, el modelo de recuperación contempla la ubicación geográfica de cada uno de los automóviles para el proceso de búsqueda y recuperación de información.

Las primeras 10 características (cualitativas), serán almacenadas utilizando la base de datos conceptual, mientras que las coordenadas de la ubicación geográfica de los automóviles serán almacenadas en la base de datos espacial y ambas bases de datos estarán vinculadas en la base de datos espacial por las tablas que hacen la función de metadatos, tal como se muestra en el apartado 4.1.3.1 de este documento.

## 4.1.2. Construcción de la base de datos conceptual

En este trabajo se ha determinado utilizar jerarquías para modelar las principales características de los automóviles. Las jerarquías son un caso particular de las ontologías, donde solo se permite una relación, la cual es única en toda la estructura ontológica.

La *Figura 4.2* muestra la jerarquía de los tipos de transmisión:

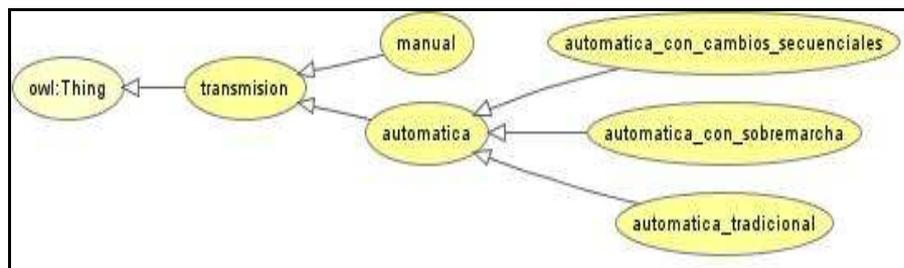


Fig. 4.2. Tipos de transmisión.

La *Figura 4.3* muestra la jerarquía con los distintos tipos de dirección:

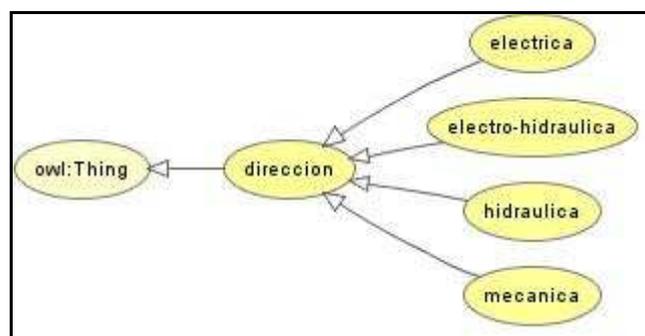


Fig. 4.3. Tipos de dirección.

La *Figura 4.4* muestra la jerarquía con la cual se clasificarán los automóviles en el sistema:

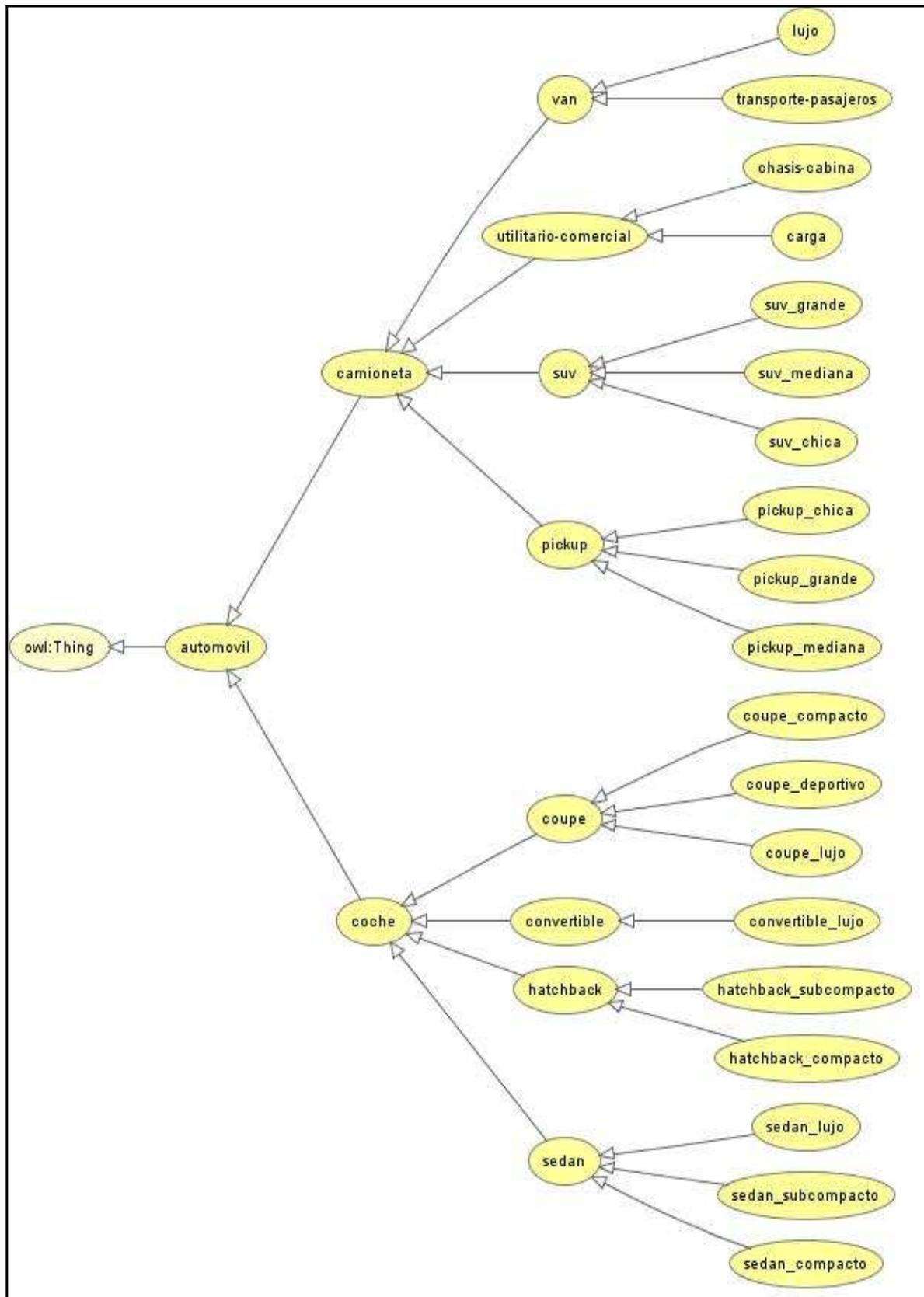


Fig. 4.4. Clasificación de automóviles.

Otro criterio importante en la selección de un automóvil es su tracción. La *Figura 4.5* muestra la jerarquía de los tipos de tracción:

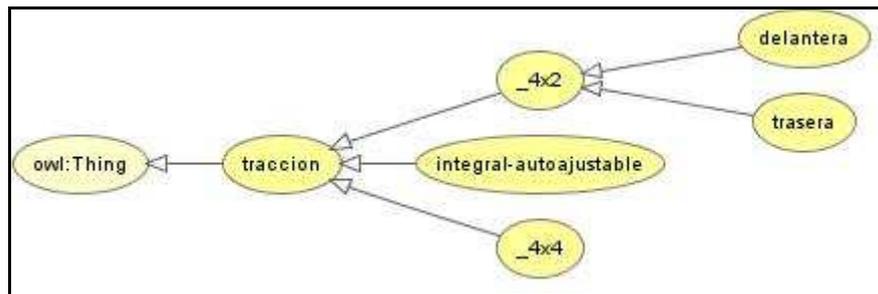


Fig. 4.5. Tipos de tracción.

Uno de los factores que impacta directamente en la potencia del motor de un automóvil es la cilindrada del motor. La *Figura 4.6* muestra la cilindrada de los motores que comúnmente son utilizados en los vehículos Mexicanos:

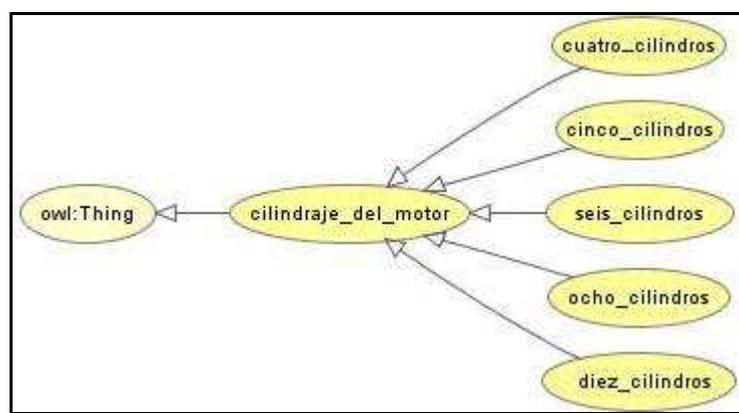


Fig. 4.6. Cilindrada del motor.

El tipo de combustible utilizado por un automóvil es un factor importante hoy en día, por ejemplo: en años recientes la fabricación de automóviles híbridos ha sido un tema de gran interés para las grandes compañías del automovilismo. En la *Figura 4.7* se muestran los tipos de combustible utilizados por los vehículos que actualmente se comercializan en México:

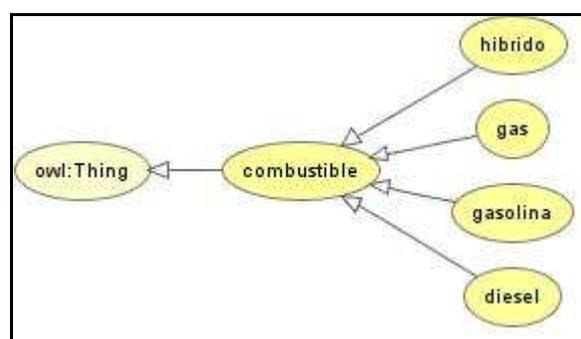


Fig. 4.7. Tipos de combustible.

El tipo de aspiración de los motores también impacta directamente en el rendimiento que estos tengan y es una característica importante de un automóvil. La *Figura 4.8* muestra la jerarquía con los distintos tipos de aspiración que utilizan los motores de un automóvil:

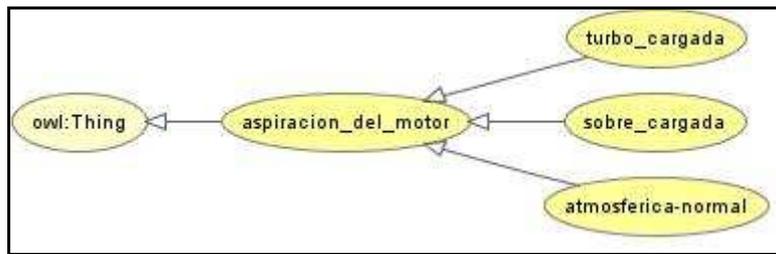


Fig. 4.8, Aspiración del motor.

Una característica importante de un motor es la configuración de éste; la *Figura 4.9* muestra las configuraciones de los motores:

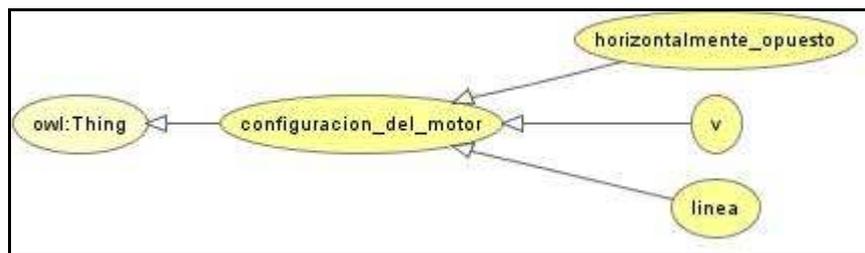


Fig. 4.9. Configuración del motor.

Los tipos de interiores de un automóvil, son una característica importante al momento de adquirir uno. La *Figura 4.10* muestra la jerarquía con los tipos de interiores:

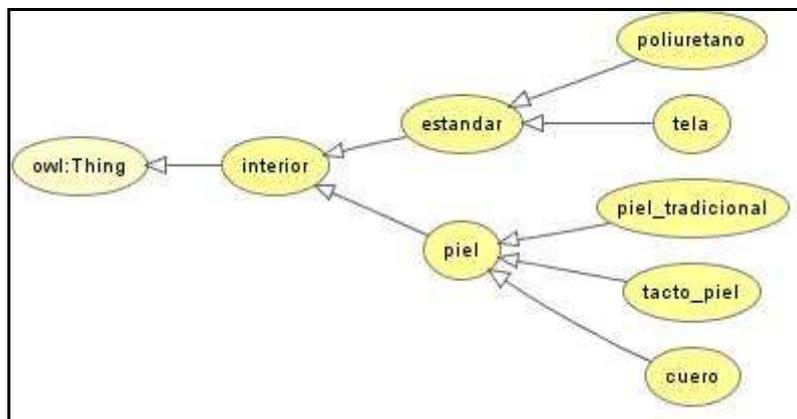


Fig. 4.10. Tipos de interiores.

Los automóviles también pueden ser clasificados de acuerdo al nivel de equipamiento con el que cuentan. La *Figura 4.11* muestra los niveles de equipamiento en los que pueden ser clasificados:

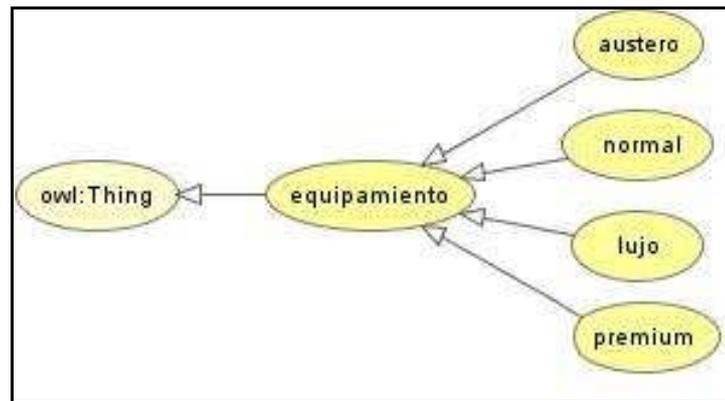


Fig. 4.II. Nivel de equipamiento.

La jerarquía anterior, con los niveles de equipamiento de los automóviles, es una jerarquía ordenada, donde existe una relación de orden “*menor que* ó <”, como se puede observar en la *Figura 4.11* (i.e. *austero* < *normal* < *lujo* < *premium*, donde *austero* es el nivel de equipamiento más bajo y *premium* es el nivel de equipamiento más alto).

Las jerarquías anteriores, fueron diseñadas utilizando el editor de ontologías *Protégé 3.4* [48] y son almacenadas en un repositorio de datos al cual se le denomina *Base de Datos Conceptual (BDC)* y será utilizado posteriormente para realizar un análisis semántico. La información utilizada para la construcción de las jerarquías se basa en la revisión de las fichas técnicas de los automóviles comercializados en México, en la clasificación de vehículos propuesta por la Asociación Mexicana de la Industria Automotriz [49] en México, así como la clasificación propuesta por eco-vehículos [50], el Portal de Indicadores de Eficiencia Energética y Emisiones Vehiculares del gobierno mexicano.

### 4.1.3. Construcción de la base de datos espacial

Para poder realizar una recuperación más flexible de la información y contemplando un posible crecimiento en el volumen de información manejado en el modelo de recuperación (escalabilidad), el proceso de conceptualización del dominio también considera la construcción una base de datos espacial y su diseño se divide en dos etapas que son:

1. El *diseño lógico* de la base de datos espacial.
2. El *diseño físico* de la base de datos espacial.

Tanto el *diseño lógico*, como el *diseño físico* son abordados en los puntos 4.1.3.1 y 4.1.3.2 respectivamente.

### 4.1.3.1. Diseño lógico de la base de datos espacial

La *Figura 4.12* muestra el diseño lógico de la base de datos que almacenara los detalles de cada uno de los productos (automóviles), así como la información referente a los vendedores de dichos productos. Además de incluir detalles de cada uno de los productos y de los vendedores, la base de datos también almacenara las coordenadas geográficas de la ubicación de los vendedores para poder realizar un análisis espacial de la información y obtener la información sobre un mapa que ayude a los usuarios a localizar donde es que se ubican exactamente los productos de su interés.

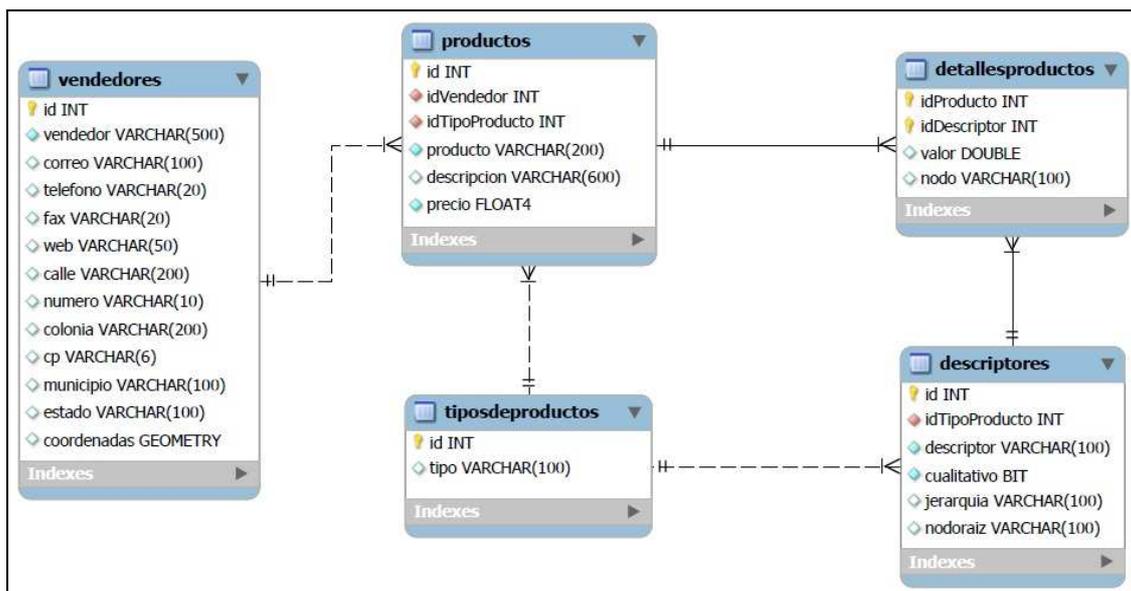


Fig. 4.12. Diseño Lógico de la Base de Datos Espacial.

Como se muestra en la figura anterior, la tabla *tiposdeproductos* funciona como una especie de metadatos de los *productos*, con la finalidad de poder incorporar diferentes productos al modelo de recuperación. Mientras que la tabla *descriptores* permite incorporar las diferentes características (*features* – en inglés) de cada uno de los tipos de productos. La columna “*cualitativo*” de la tabla *descriptores* indica si la característica o descriptor, se representara como una variable cualitativa (1) o como una variable numérica (0). Para el caso de

los descriptores cualitativos la tabla descriptores almacena también una referencia al nombre del archivo con la jerarquía y al nodo raíz de esta.

Como su nombre lo dice la tabla *detallesproductos* almacenara los detalles de cada uno de los productos, manteniendo para cada uno de los descriptores del producto el valor numérico o el nodo de la jerarquía según corresponda para dicho atributo.

### **4.1.3.2. Diseño físico de la base de datos espacial**

El diseño físico de la base de datos consiste en la implementación del diseño lógico de la base de datos, a nivel del Leguaje de Definición de Datos (DDL por sus siglas en ingles) del sistema manejador de bases de datos, en este caso PostgreSQL. El *Anexo B* muestra un script con las sentencias SQL (PL/SQL) necesarias para generar la base de datos del apartado 4.1.3.1.

## **4.2. Búsqueda y recuperación de información**

En este apartado, se aborda tal vez la parte más importante de esta investigación, pues aquí se describe el modelo de búsqueda y recuperación de información, el cual se muestra en la en la *Figura 4.13*.

El modelo de búsqueda y recuperación de información consiste básicamente de un buscador que recibe como entradas las consultas de los usuarios y realiza la evaluación de criterios sobre la *Base de Datos Conceptual* y la *Base de Datos Espacial*. Posteriormente, el buscador realiza un ordenamiento de resultados mediante un proceso de Ranking de resultados; dicho proceso de Ranking utiliza como entrada los resultados de la evaluación de criterios y las preferencias de los usuarios para ponderar los resultados y construir un listado de resultados único, que integra y ordena dichos resultados según su relevancia.

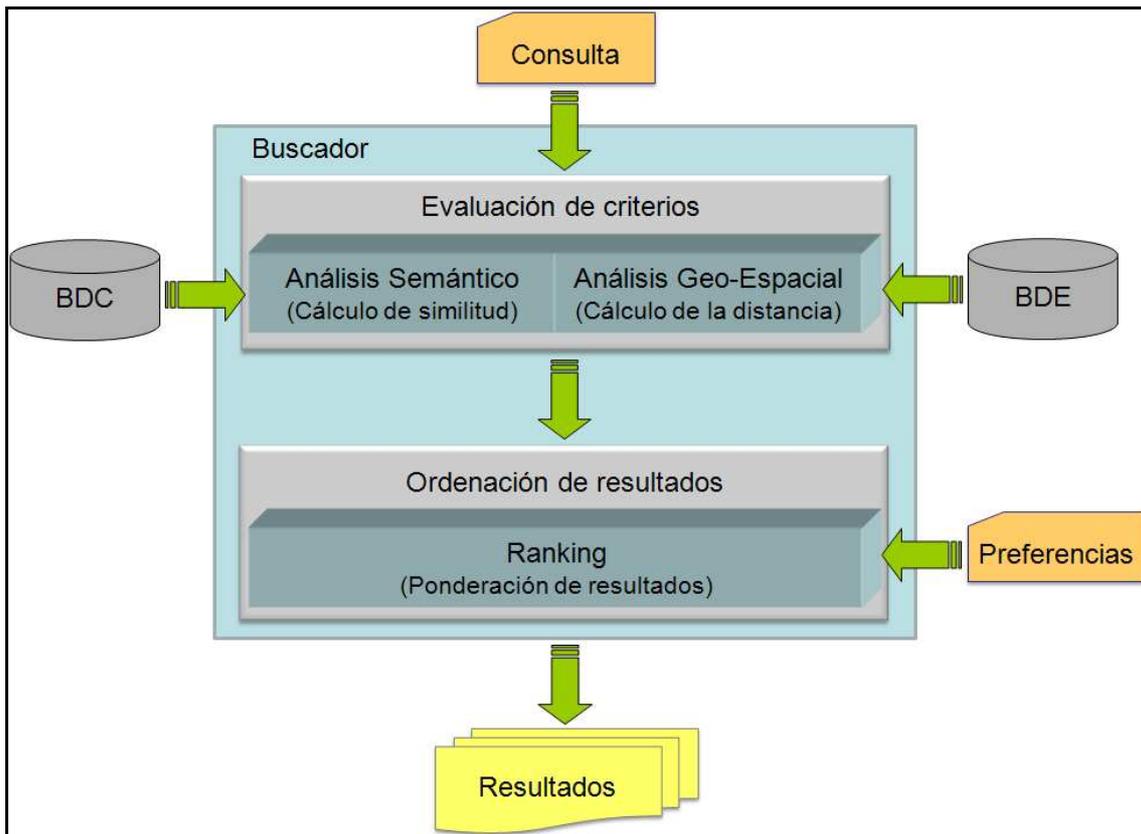


Fig. 4.13. Proceso de búsqueda y recuperación de información.

La *Figura 4.14* muestra el pseudocódigo de los pasos que debe realizar el buscador hasta obtener los resultados de una búsqueda.

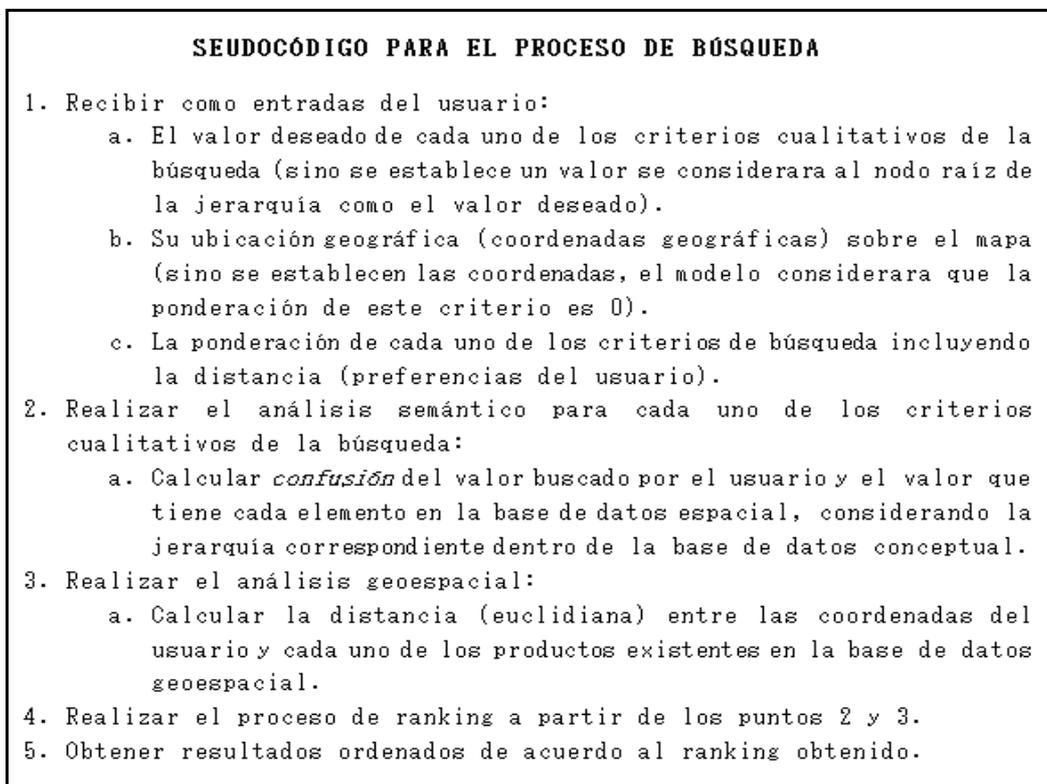


Fig. 4.14. Seudocódigo de los pasos que se deben realizar durante el proceso de búsqueda de información.

Ahora bien, para abordar a detalle los distintos procesos que realiza el buscador para la búsqueda y recuperación de información, suponga que existe un usuario ubicado en las coordenadas (-11953338.032, 2854240.745) que busca un automóvil con las siguientes características:

- Clasificación: Convertible
- Tracción: Delantera
- Transmisión: Automática
- Motor:
  - Cilindrada: 8 cilindros
- Interiores: Cuero
- Equipamiento: Lujo

Y considere el conjunto de datos de prueba de la *Tabla 4.1*:

Tabla 4.1. Conjunto de datos de prueba.

Modelo	Clase	Cilindros	Tracción	Transmisión	Interiores	Equipamiento	Latitud	Longitud
chevy	hatchback_subcompacto	cuatro_cilindros	delantera	manual	tela	austero	-11957270.36	2851401.827
chevy sedan	sedan_subcompacto	cuatro_cilindros	delantera	manual	tela	austero	-11957270.36	2851401.827
aveo	sedan_subcompacto	cuatro_cilindros	delantera	automatica	tela	normal	-11957270.36	2851401.827
optra	sedan_compacto	cuatro_cilindros	delantera	automatica	tela	normal	-11957270.36	2851401.827
cruze	sedan_compacto	cuatro_cilindros	delantera	automatica	tela	normal	-11957270.36	2851401.827
malibu	sedan_lujo	cuatro_cilindros	delantera	automatica	tela	lujo	-11957270.36	2851401.827
camaro	coupe_deportivo	seis_cilindros	delantera	manual	piel	lujo	-11957270.36	2851401.827
corvette	coupe_deportivo	ocho_cilindros	trasera	manual	piel	premium	-11957270.36	2851401.827
fiesta	sedan_subcompacto	cuatro_cilindros	delantera	manual	tela	austero	-11953338.03	2850852.137
focus	sedan_subcompacto	cuatro_cilindros	delantera	automatica	tela	normal	-11953338.03	2850852.137
focus 5 puertas	hatchback_compacto	cuatro_cilindros	delantera	automatica	tela	normal	-11953338.03	2850852.137
focus st	hatchback_compacto	cinco_cilindros	delantera	manual	piel	lujo	-11953338.03	2850852.137
fusion	sedan_lujo	seis_cilindros	delantera	automatica	piel	normal	-11953338.03	2850852.137
fusion sport	sedan_lujo	seis_cilindros	delantera	manual	piel	lujo	-11953338.03	2850852.137
mustang	coupe_deportivo	ocho_cilindros	delantera	manual	piel	lujo	-11953338.03	2850852.137
mustang cobra	coupe_deportivo	ocho_cilindros	delantera	manual	piel	premium	-11953338.03	2850852.137

A continuación, se describirá cada uno de los procesos y posteriormente se ejemplificarán dichos procesos considerando la búsqueda del ejemplo anterior.

### 4.2.1. Análisis semántico

La fase del análisis semántico permite determinar cuál es el grado de similitud semántica existente entre los productos que un usuario está buscando y los que se encuentran en las fuentes de información utilizadas por el buscador. Cuando es posible determinar el grado de similitud semántica, pueden presentarse diversos escenarios, al obtener los resultados de una consulta:

- Existen resultados que satisfacen completamente (100%) la consulta (resultados que pueden ser recuperados por los típicos sistemas manejadores de bases de datos).
- No existen resultados que satisfagan completamente la consulta. Este caso puede ser el más interesante y en donde muchos sistemas de información fallan, pues es necesario determinar: cual o cuales de los resultados satisfacen en mayor grado una consulta.

Ahora bien, ¿cómo es posible determinar en qué grado satisfacen los resultados a una consulta? Esto es posible mediante el cálculo de la similitud semántica, en el capítulo tres (marco teórico) de este trabajo, se ha propuesto utilizar la medida de similitud semántica *confusión* basada en la *Teoría de Confusión* propuesta por Levachkine y Guzmán en [45] con unas pequeñas variantes que se presentan a continuación:

#### 1. Refinación de resultados:

Se propone dividir los resultados del cálculo de *confusión* entre la profundidad de  $r$  (representada por  $p(r)$ ), esto con la finalidad de que la *confusión* entre conceptos superiores de la jerarquía, sea mayor que la *confusión* entre conceptos menos abstractos (nodos de mayor profundidad) de la jerarquía. La expresión propuesta es la siguiente:

$$\frac{conf(r, s)}{p(r)}$$

Adicionalmente, esta variación al cálculo de *confusión* permite que los resultados del cálculo de *confusión* estén normalizados en un intervalo cerrado de 0 a 1.

#### 2. Función de similitud:

Para obtener una función de similitud a partir del cálculo de *confusión* (i.e. *confusión* es una función de disimilitud), se propone utilizar la siguiente expresión:

$$Sim(r, s) = 1 - \frac{conf(r, s)}{p(r)}$$

Finalmente, la expresión  $Sim(r,s)$  representa mayor similitud cuando su valor tiende a 1 y menor similitud cuando su valor tiende a 0.

Aplicando el análisis semántico al ejemplo del punto 4.2 utilizando los datos de la *Tabla 4.1*, se calcula la *confusión* existente para cada criterio (columna) entre el valor solicitado por el usuario y el valor existente en el conjunto de datos cómo se muestra en la *Tabla 4.2*:

Tabla 4.2. Resultados del análisis semántico.

Modelo	Clase	Cilindros	Tracción	Transmisión	Interiores	Equipamiento
chevy	0.5	0.04	1.00	0.50	0.50	0.50
chevy sedan	0.5	0.04	1.00	0.50	0.50	0.50
aveo	0.5	0.04	1.00	1.00	0.50	0.25
optra	0.5	0.04	1.00	1.00	0.50	0.25
cruze	0.5	0.04	1.00	1.00	0.50	0.25
malibu	0.5	0.04	1.00	1.00	0.50	1.00
camaro	0.5	0.80	1.00	0.50	1.00	1.00
corvette	0.5	1.00	0.33	0.50	1.00	0.25
fiesta	0.5	1.00	1.00	0.50	0.50	0.50
focus	0.5	1.00	1.00	1.00	0.50	0.25
focus 5 puertas	0.5	1.00	1.00	1.00	0.50	0.25
focus st	0.5	0.60	1.00	0.50	1.00	1.00
fusion	0.5	0.80	1.00	1.00	1.00	0.25
fusion sport	0.5	0.80	1.00	0.50	1.00	1.00
mustang	0.5	1.00	1.00	0.50	1.00	1.00
mustang cobra	0.5	1.00	1.00	0.50	1.00	0.25

## 4.2.2. Análisis geoespacial

Considerar las posiciones geográficas de los productos o servicios, en el proceso de búsqueda y recuperación de información, enriquece a las medidas de similitud semántica, tales como la medida propuesta por Levachkine y Guzmán [46], permitiendo refinar los resultados que serán devueltos al usuario, después de sus consultas.

La presente investigación plantea utilizar la distancia entre la ubicación de los productos o servicios y la ubicación de los usuarios, permitiendo refinar los resultados en función de la posición de cada usuario en particular, es decir, obteniendo aquellos resultados que están más cercanos (menos distantes) a la posición de los usuarios. Para el cálculo de dicha distancia se propone utilizar la distancia Euclidiana entre dos puntos con coordenadas: latitud, longitud; sobre el globo terráqueo. La expresión para calcular dicha distancia ( $d$ ) entre dos puntos se muestra a continuación:

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_1 - y_2)^2}$$

Ahora bien, si utilizamos esta expresión para evaluar el ejemplo del punto 4.2 y calcular la distancia existente entre los automóviles del conjunto de datos de la *Tabla 4.1*, obtendremos los resultados de la *Tabla 4.3*.

Tabla 4.3. Resultados del cálculo de la distancia.

Modelo	Distancia
chevy	4850.02
chevy sedan	4850.02
aveo	4850.02
optra	4850.02
cruze	4850.02
malibu	4850.02
camaro	4850.02
corvette	4850.02
fiesta	3839.05
focus	3839.05
focus 5 puertas	3839.05
focus st	3839.05
fusion	3839.05
fusion sport	3839.05
mustang	3839.05
mustang cobra	3839.05

Para normalizar los valores del análisis geoespacial en un intervalo [0, 1], se propone dividir los resultados del cálculo de la distancia por el valor máximo de esta. Como se muestra en la *Tabla 4.4*:

Tabla 4.4. Resultados del cálculo de la distancia normalizados.

Modelo	Distancia
chevy	1
chevy sedan	1
aveo	1
optra	1
cruze	1
malibu	1
camaro	1
corvette	1
fiesta	0.79
focus	0.79
focus 5 puertas	0.79
focus st	0.79
fusion	0.79
fusion sport	0.79
mustang	0.79
mustang cobra	0.79

### 4.2.3. Ranking de resultados

Interpretar y trabajar con múltiples listas de resultados generadas a partir de la evaluación de cada uno de los criterios de búsqueda, es un trabajo realmente difícil y complejo para los seres humanos, más aun cuando el número de criterios de búsqueda tiende a crecer o ser grande.

El ranking de resultados, permite la integración de los múltiples criterios de búsqueda, generando una única lista de resultados que es más fácil de entender por los usuarios finales [4]. En este proceso, cada criterio es ponderado por los usuarios permitiendo que estos expresen explícitamente, cuales criterios serán de mayor o menor importancia para su búsqueda, logrando una excelente personalización de los resultados.

La *Figura 4.15* muestra como un pseudocódigo los pasos a seguir para obtener el valor de ranking y ordenar los resultados de acuerdo a este valor.

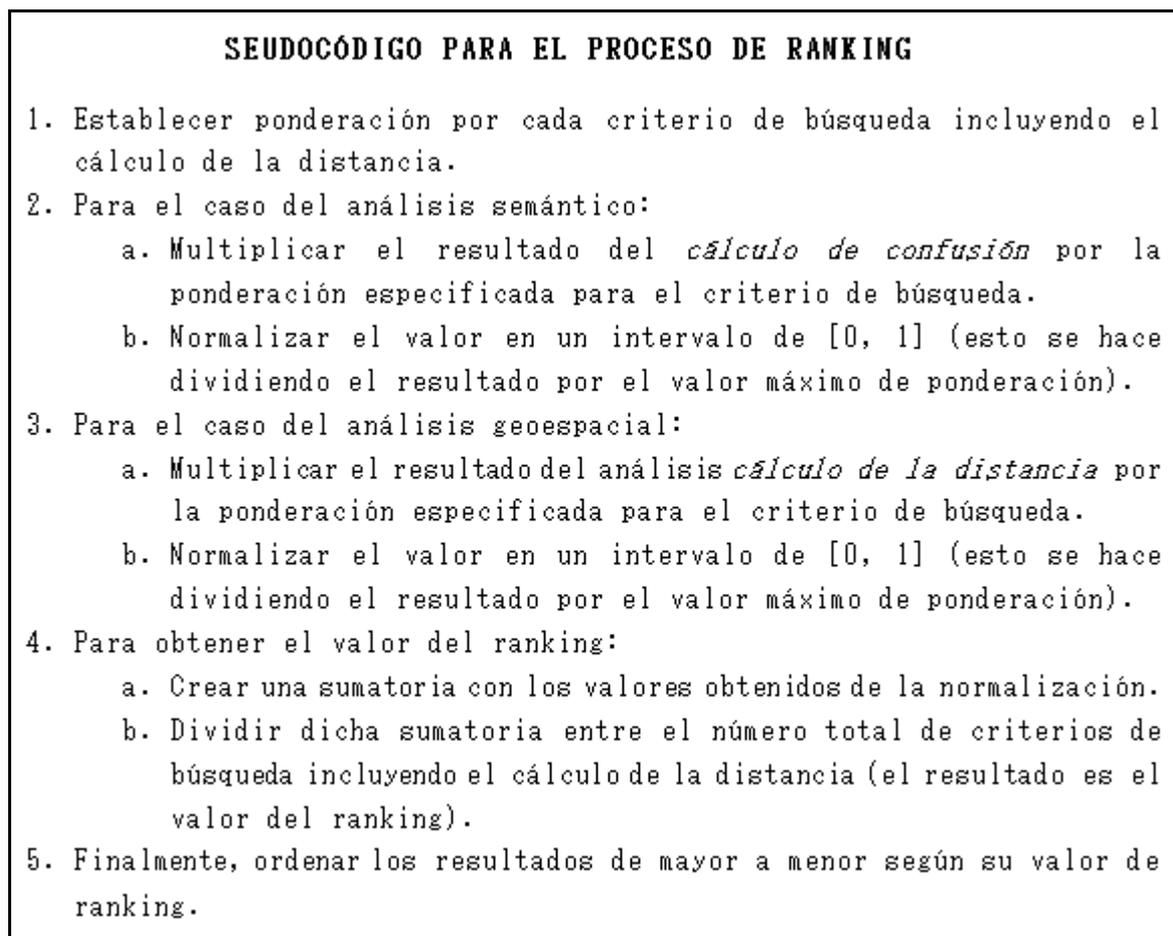


Fig. 4.15. Seudocódigo para el proceso de ranking.

Formalmente, la ponderación e integración de resultados está definida por la siguiente expresión de ranking:

$$\text{rank}(D, P) = \frac{\sum_{i=1}^n D_i P_i}{n * \max(P)}$$

Donde:

- $D$ : Es el conjunto con la evaluación de cada uno de los *descriptores* o criterios utilizados para recuperar los productos o servicios.
  - Para el caso de los criterios semánticos  $D_i = \text{Sim}(r_k, s_i)$  donde  $k$  pertenece al conjunto de valores existentes en la jerarquía a la cual pertenece  $s_i$ .
  - Para el caso del criterio geoespacial:

$$D_i = d(\text{posición}(r_k), \text{posición}(\text{usuario}))$$

- $P$ : Es el conjunto de valores (pesos) utilizados para ponderar la evaluación de cada uno de los *descriptores* o criterios utilizados para recuperar información.
- $n$ : Representa la cardinalidad de los conjuntos  $D$  y  $P$  ( $n = \#D = \#P$ ).
- $\max(P)$ : Es el máximo valor en el conjunto de pesos  $P$ .

Sustituyendo en la expresión anterior los criterios utilizados en el caso de estudio de este trabajo de tesis, la expresión puede ser expresada como:

$$\text{rank}(D, P) = \frac{\text{transmission}(P_1) + \text{direccion}(P_2) + \text{clasificacion}(P_3) + \text{traccion}(P_4) + \text{cilindrada}(P_5) + \text{combustible}(P_6) + \text{configuracion}(P_7) + \text{interiores}(P_8) + \text{equipamiento}(P_9) + \text{distancia}(P_{10})}{10 * \max(P)}$$

Dónde:

- $D = \left\{ \begin{array}{l} \text{transmission, direccion, clasificacion, traccion, cilindrada, combustible,} \\ \text{configuracion, interiores, equipamiento, distancia} \end{array} \right\}$
- $P = \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}\}$

Para ejemplificar el proceso de Ranking con los datos del punto 4.2, considere los datos de la *Tabla 4.5* como la ponderación asignada por el usuario según sus preferencias:

Tabla 4.5. Ponderación de criterios.

Clase	Cilindros	Tracción	Transmisión	Interiores	Equipamiento	Distancia
0.5	1	0	1	0.5	1	1

La *Tabla 4.6* muestra el resultado de la multiplicación de cada uno de los criterios (del análisis semántico y geoespacial) por la ponderación de cada criterio. La columna final de esta tabla (*Rank*), muestra el resultado de la sumatoria de cada uno de los criterios y su multiplicación.

Tabla 4.6. Criterios ponderados.

Modelo	Clase	Cilindros	Tracción	Transmisión	Interiores	Equipamiento	Distancia	Rank
chevy	0.25	0.04	0.00	0.50	0.25	0.50	1.00	2.54
chevy sedan	0.25	0.04	0.00	0.50	0.25	0.50	1.00	2.54
aveo	0.25	0.04	0.00	1.00	0.25	0.25	1.00	2.79
optra	0.25	0.04	0.00	1.00	0.25	0.25	1.00	2.79
cruze	0.25	0.04	0.00	1.00	0.25	0.25	1.00	2.79
malibu	0.25	0.04	0.00	1.00	0.25	1.00	1.00	3.54
camaro	0.25	0.80	0.00	0.50	0.50	1.00	1.00	4.05
corvette	0.25	1.00	0.00	0.50	0.50	0.25	1.00	3.50
fiesta	0.25	1.00	0.00	0.50	0.25	0.50	0.79	3.29
focus	0.25	1.00	0.00	1.00	0.25	0.25	0.79	3.54
focus 5 puertas	0.25	1.00	0.00	1.00	0.25	0.25	0.79	3.54
focus st	0.25	0.60	0.00	0.50	0.50	1.00	0.79	3.64
fusion	0.25	0.80	0.00	1.00	0.50	0.25	0.79	3.59
fusion sport	0.25	0.80	0.00	0.50	0.50	1.00	0.79	3.84
mustang	0.25	1.00	0.00	0.50	0.50	1.00	0.79	4.04
mustang cobra	0.25	1.00	0.00	0.50	0.50	0.25	0.79	3.29

Finalmente, los resultados son ordenados de acuerdo a su relevancia (*Rank*), como se muestra en la *Tabla 4.7*:

Tabla 4.7. Resultados ordenados de acuerdo a su relevancia.

Modelo	Clase	Cilindros	Tracción	Transmisión	Interiores	Equipamiento	Distancia	Rank
camaro	0.25	0.80	0.00	0.50	0.50	1.00	1.00	4.05
mustang	0.25	1.00	0.00	0.50	0.50	1.00	0.79	4.04
fusion sport	0.25	0.80	0.00	0.50	0.50	1.00	0.79	3.84
focus st	0.25	0.60	0.00	0.50	0.50	1.00	0.79	3.64
fusion	0.25	0.80	0.00	1.00	0.50	0.25	0.79	3.59
focus	0.25	1.00	0.00	1.00	0.25	0.25	0.79	3.54
focus 5 puertas	0.25	1.00	0.00	1.00	0.25	0.25	0.79	3.54
malibu	0.25	0.04	0.00	1.00	0.25	1.00	1.00	3.54
corvette	0.25	1.00	0.00	0.50	0.50	0.25	1.00	3.50
fiesta	0.25	1.00	0.00	0.50	0.25	0.50	0.79	3.29
mustang cobra	0.25	1.00	0.00	0.50	0.50	0.25	0.79	3.29
aveo	0.25	0.04	0.00	1.00	0.25	0.25	1.00	2.79
optra	0.25	0.04	0.00	1.00	0.25	0.25	1.00	2.79
cruze	0.25	0.04	0.00	1.00	0.25	0.25	1.00	2.79
chevy	0.25	0.04	0.00	0.50	0.25	0.50	1.00	2.54
chevy sedan	0.25	0.04	0.00	0.50	0.25	0.50	1.00	2.54

Como se puede observar en la Tabla 4.6 los resultados obtenidos obedecen a los automóviles más cercanos y más parecidos a lo que el usuario solicitó en su consulta. En este caso obedecen a automóviles “deportivos de calle”. Se podría pensar entonces que los resultados son incorrectos pues el *Mustang Cobra* y el *Corvette* no aparecen al principio de la lista, pero esto se debe a la ponderación asignada por el usuario a cada uno de los distintos criterios que intervienen en la búsqueda.

### 4.3. Arquitectura de la aplicación

Para materializar la propuesta de la metodología desarrollada en el presente trabajo de tesis, la *Figura 4.16* muestra la arquitectura de la aplicación. Dicha arquitectura describe el prototipo de un sistema de recuperación de información que implementa cada una de las fases de la metodología. A continuación se abordarán a detalle cada uno de los componentes de la arquitectura:



Fig. 4.16. Arquitectura de la aplicación.

Las consultas de los usuarios así como sus preferencias son implementadas mediante una interfaz Web desarrollada en Java Server Page (JSP). Una vez capturadas las consultas y las preferencias de los usuarios, estos datos sirven como entrada al buscador, dicho buscador realiza el proceso de recuperación

descrito por la metodología planteada en este trabajo de tesis y es implementado en el lenguaje de programación Java (J2SE). Como se describe en la metodología, el buscador realiza un proceso de búsqueda y recuperación de información en dos fuentes de datos: Una Base de Datos Espacial y una Base de Datos Conceptual. La Base de Datos Espacial implementada en PostgreSQL/PostGIS almacena la información concerniente a cada uno de los vendedores, así como los metadatos de los diversos productos que este puede vender, mientras que la Base de Datos Conceptual está conformada por un conjunto de archivos OWL, lo cuales representan las características o descriptores cualitativos de cada uno de los productos o servicios que se encuentran en venta.

Una vez terminado el proceso del buscador este regresa una tabla de resultados a la interfaz Web, con la ponderación de resultados de acuerdo a su relevancia basado en las preferencias de los usuarios. Adicionalmente, el buscador envía a un servidor de datos espaciales, un conjunto de resultados con las coordenadas geográficas de cada uno de los vendedores de los productos recuperados para cada una de las consultas. El servidor de datos espaciales GeoServer se encarga de la generación de mapas temáticos, mediante Servicios Web para Mapas (WMS - por sus siglas en ingles), dichos servicios Web entregan archivos KML (formato propuesto por *GoogleEarth* para la representación de datos espaciales) y son utilizados por la interfaz Web para el despliegue de resultados sobre mapas de *GoogleMaps* utilizando un API de OpenSource para el desarrollo de mapas sobre la Web denominada OpenLayers la cual está basada en JavaScript.

#### 4.4. Comentarios finales

La metodología desarrollada en el presente trabajo de tesis, ofrece una alternativa para flexibilizar la búsqueda de productos o servicios basándose en las características cualitativas y las ubicaciones geográficas de estos. El proceso de búsqueda y recuperación de información plantea la evaluación de los distintos criterios de búsqueda valiéndose de dos tipos de análisis: el análisis semántico y el análisis geoespacial. El análisis semántico busca flexibilizar el proceso de recuperación de resultados mediante el uso de medidas de similitud, el realizar distintas mediciones semánticas permite obtener resultados aún cuando estos no cumplen en un 100% con los criterios de búsqueda (lo anterior es un requisito forzoso en las bases de datos tradicionales para poder obtener

información en las consultas que se realizan en este tipo de sistemas). Por otro lado el análisis geoespacial busca incluir las posiciones geográficas de cada uno de los productos o servicios que existen registrados en el sistema y la posición de los usuarios, con la finalidad de ofrecer a los usuarios aquellos resultados que estén más próximos a este.

Adicionalmente la metodología propuesta puede ser materializada mediante una arquitectura o prototipo de la aplicación. Algunas de las características de dicha arquitectura son:

- La separación de la interfaz del software y del buscador, permite que la interfaz pueda ser fácilmente implementada en otro tipo de tecnologías sin afectar el resto de la aplicación.
- La integración de los mapas mediante el uso de WMS y archivos KML supone una ventaja, pues se basa en dos estándares para la generación, representación y consumo de mapas desde diversas aplicaciones.
- Una característica importante de esta arquitectura es que se basa completamente en código abierto, permitiendo reducir drásticamente los costos en el desarrollo de este tipo de aplicaciones.

Es importante resaltar que el trabajo puede expandirse a utilizar otros criterios del tipo numérico así como otras medidas de similitud no semánticas y esto está contemplado en los metadatos contenidos en la base de datos espacial, dichos metadatos también permiten agregar nuevos productos o servicios así como sus correspondientes descriptores, dotando de una gran flexibilidad al modelo propuesto.

Finalmente es importante aclarar que la integración del dominio geoespacial, con el dominio semántico permitirá que los sistemas de búsqueda y recuperación de información se complementen y que se acerquen a un nivel de inteligencia, tal que, estos puedan ofrecer las mejores alternativas que tiene un usuario bajo ciertas restricciones. El considerar cuales son las preferencias de cada usuario particular, permite que estos resultados le sean más útiles a este.

## **CAPÍTULO 5.**

# **PRUEBAS Y RESULTADOS**

En este capítulo, se presentan los resultados obtenidos con el modelo de recuperación de información propuesto. Adicionalmente a los resultados, se presenta la implementación de un sistema Web que funciona como un buscador de automóviles basado en la metodología propuesta en el presente trabajo de tesis. Dicho sistema, utiliza los atributos cualitativos de los automóviles y considera su ubicación con respecto a los usuarios para realizar la búsqueda y recuperación de información, además de considerar la ubicación de dichos automóviles con respecto a los usuarios del sistema.

Es importante mencionar, que para cada consulta, se recupera un conjunto ordenado de resultados en función del grado de relevancia que estos tengan (según un proceso de ranking). Los resultados consisten de un listado de automóviles vinculados a un mapa que indica la ubicación geográfica de dichos automóviles basándose en sus vendedores (agencias de autos). Cada resultado cuenta con también con una descripción detallada de sus características.

### **5.1. Evaluación de los componentes del modelo de recuperación de información**

En esta sección, se evalúan los principales componentes del modelo de recuperación propuesto en la metodología.

Los componentes a evaluar son los siguientes:

- Análisis semántico (Cálculo de confusión sobre las jerarquías).
- Análisis geoespacial (Cálculo de la distancia),
- Asignación y evaluación de los valores de ponderación de cada criterio.
- Evaluación de la expresión de ranking.

### 5.1.1. Análisis semántico

Para evaluar el análisis semántico que realiza el sistema es necesario, referirse al cálculo de la similitud semántica sobre las jerarquías utilizadas en el sistema de búsqueda y recuperación de información. Como se mencionó anteriormente, la medida de similitud semántica utilizada para el análisis semántico es la *Teoría de Confusión* [45].

Para realizar el análisis semántico en el sistema se ha implementado un paquete denominado “confusión” en el lenguaje de programación *Java* que permite calcular la medida de similitud semántica denominada *confusión*, utilizando jerarquías simples, ordenas y submixtas (simples y ordenadas); dichas jerarquías deben ser implementadas en OWL utilizando *Protege 3.4*.

La principal utilidad de este análisis, es evaluar la similitud existente entre los diferentes productos o servicios contenidos en la base de datos desde una perspectiva semántica. Adicionalmente este análisis permite controlar la precisión con la que son recuperados los diferentes productos o servicios después de alguna consulta realizada por un usuario.

Utilizando el análisis semántico se pueden obtener resultados, tales que, dada una consulta donde se pregunta por un determinado tipo de automóvil como: *sedan compacto*, los resultados sean de acuerdo a la *Figura 5.1*:

Clase	ε	Clase	ε	Clase	ε
sedan_compacto	1.00	convertible_deportivo	0.50	pickup_chica	0.25
sedan_lujo	0.92	convertible_lujo	0.50	pickup_grande	0.25
sedan_subcompacto	0.92	coupe_compacto	0.50	pickup_mediana	0.25
		coupe_deportivo	0.50	suv_chica	0.25
		coupe_lujo	0.50	suv_grande	0.25
		hatchback_compacto	0.50	suv_mediana	0.25
		hatchback_subcompacto	0.50	van_chica	0.25
				van_grande	0.25
				van_mediana	0.25

Fig. 5.1. Resultados del cálculo de confusión.

Cómo se muestra en la *Figura 5.1* los resultados obtenidos por el cálculo de confusión sólo se realizan considerando los nodos finales (hoja) de la jerarquía. Esto es por una optimización para poder calcular la medida de confusión en tiempo de ejecución del sistema de una forma rápida.

Es importante resaltar que los resultados obtenidos por el cálculo de confusión, reflejan claramente la semántica del dominio, pues se puede apreciar que el primer bloque de resultados (verde) corresponde a los automóviles de la misma clase (*sedan*). Mientras que el segundo bloque (naranja) corresponde a los automóviles de la clase *coche* y el tercer bloque (marrón) corresponde a todos los automóviles correspondientes a la clase *camioneta*.

También es importante resaltar que al utilizar la siguiente expresión:

$$Sim(r, s) = 1 - \frac{conf(r, s)}{p(r)}$$

Tal que:

$r$ : es el nodo existente.

$s$ : es el nodo ideal.

$p(r)$ : es la profundidad de  $r$ .

Los resultados, son más precisos y reflejan claramente la similitud existente entre los diferentes nodos de la jerarquía, tal como se refleja en los colores de la *Figura 5.1*, donde el verde indica mayor similitud (cuando  $\varepsilon \rightarrow 1$ ) y el marrón indica menor similitud (cuando  $\varepsilon \rightarrow 0$ ).

A continuación, en la *Figura 5.2* se muestra el diagrama de un segmento de jerarquía sobre los seres vivos.

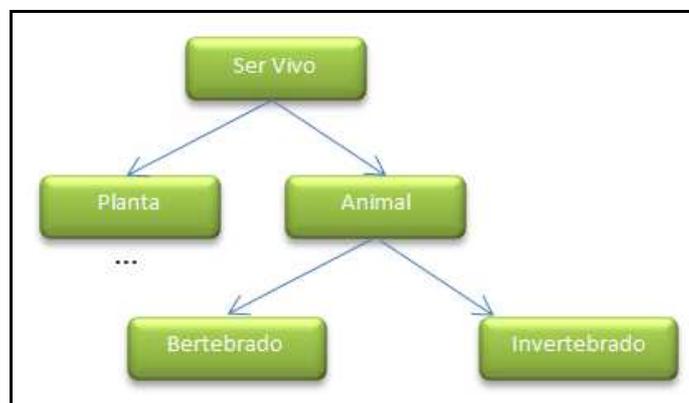


Fig. 5.2. Segmento de jerarquía sobre los seres vivos.

Si evaluamos la confusión entre planta y animal utilizando la expresión original, se muestra que:

- $conf(\text{animal}, \text{planta})=1$
- $conf(\text{betebrado}, \text{invertebrado})=1$

Mientras que al dividir estos resultados por la profundidad de  $r$ , los resultados serían:

- $conf(\text{animal}, \text{planta})/1=1/1 = 1$
- $conf(\text{betebrado}, \text{invertebrado})/2=1/2 = 0.5$

Como se puede observar entre más abstractos (superiores) son los conceptos, la diferencia la *confusión* es mayor. Y mientras más específicos (inferiores) son los conceptos la *confusión* es menor.

Esto obedece a que entre más abstracto es un concepto, la diferencia conceptual con otros conceptos es mayor.

## 5.1.2. Análisis geoespacial

Como se mencionó en el capítulo anterior (Capítulo 4: Metodología), el análisis geoespacial en este sistema corresponde únicamente al cálculo de la distancia entre los productos o servicios que existen en el sistema y la posición de los usuarios.

Para medir la distancia que existe entre los productos o servicios que un usuario está buscando y la ubicación de este, se ha determinado utilizar la distancia Euclidiana. Dicha distancia ha sido ampliamente utilizada para medir distancias entre puntos ubicados en un plano, debido a que cumple con las siguientes características:

1. Resultados en los reales:
  - a.  $d(x, y) = 0$  si  $x = y$
  - b.  $d(x, y) \geq 0$  si  $x \neq y$
2. Simetría:
  - a.  $d(x, y) = d(y, x)$
3. Desigualdad triangular:

- a. Si se agrega una coordenada más ( $z$ ) se comprueba que:  $d(x, y) \leq d(x, z) + d(z, y)$

Pero medir la distancia entre dos puntos sobre la superficie terrestre no es una tarea simple, pues existen muchos factores a considerar, por ejemplo desde el punto de vista geográfico, la distancia geodésica es más apropiada para medir la distancia existente entre dos puntos sobre la superficie terrestre. No obstante desde la perspectiva de los seres humanos dicha distancia tampoco es apropiada, pues no considera los bloques o cuadras de una ciudad, considerando esto se podría decir entonces que ¿la distancia Manhattan o City block es más apropiada, ya que su cálculo se basa en dichos bloques? La respuesta es no, pues es una distancia que cumple con el axioma de simetría, además de que no considera otros factores como: las reglas que cumple con el axioma de simetría, además de que no considera otros factores como: las reglas y señalamientos de tránsito, flujo de vialidades, horas pico, entre otros factores más.

Debido a lo anterior y para simplificar el modelo de recuperación de información, se ha optado por hacer uso de la distancia Euclidiana, para ejemplificar el modelo de recuperación propuesto. Es importante mencionar que la medida de distancia propuesta puede ser fácilmente reemplazada en el modelo de recuperación, por otra que pudiese resultar más adecuada con la finalidad de mejorar la precisión de este criterio.

Ciertamente el utilizar, tanto el análisis semántico como el análisis geoespacial, es con la finalidad de cubrir las necesidades de aquellos usuarios que además de requerir resultados conceptualmente similares, estén más cercanos (menos distancia) a su ubicación. Pero habrá usuarios, para los cuales el criterio geoespacial resulte algo sin importancia, menos importante o incluso más importante que otros criterios de la búsqueda. Considerando estas posibilidades el sistema propuesto considera un mecanismo para ponderar la evaluación de los diferentes criterios de búsqueda y es abordado en el siguiente punto (5.1.3) de este documento.

### **5.1.3. Asignación y evaluación de los valores de ponderación de cada criterio**

Considerando que no todas las personas tienen las mismas preferencias y necesidades, el sistema de búsqueda y recuperación de información propuesto considera un mecanismo de ponderación de los diferentes criterios que

intervienen en una búsqueda, permitiendo que los usuarios determinen que criterios son más o menos importantes en sus búsquedas.

Pero ahora bien que esquemas existen actualmente para captar la importancia de los diferentes criterios de búsqueda:

1. **Checklist (booleano):** Esquema donde un criterio simplemente importa o no importa.
2. **Basado en expertos:** La importancia de cada criterio es especificada previamente por uno o más expertos, quienes consideran que es realmente importante y que no.
3. **Valores numéricos:** Los usuarios requieren introducir valores numéricos indicando que criterios son más o menos importantes para ellos.

En [27] se propone un ecualizador de preferencias, que permite a los usuarios interactuar con barras, que permiten a los usuarios indicar los valores de sus preferencias sin la necesidad de interactuar con valores numéricos. Haciendo una analogía en como una persona elige y configura sus preferencias musicales, de acuerdo a lo que quiere escuchar, se asume que quien más sabe lo que quiere es el mismo usuario. En este trabajo, se propone utilizar una barra para indicar que tan importante es cada criterio y un recuadro que indica visualmente mediante un semáforo de colores cual es la importancia de cada criterio como se muestra en la *Figura 5.3*.

Es importante mencionar que el semáforo de colores únicamente capta la importancia asignada a cada criterio por los usuarios y que dicha asignación afecta directamente los resultados de cada consulta. De esta forma los resultados de una misma consulta pueden variar drásticamente, en función de la importancia asignada a cada criterio. Por ejemplo si un usuario decide no dar importancia al criterio de distancia, los resultados obedecerán únicamente al análisis semántico, pero si otro usuario o incluso el mismo usuario, decide no dar importancia a los criterios semánticos y realizar su consulta únicamente con el criterio de distancia, los resultados arrojados por el sistema estarán ordenados de acuerdo a la cercanía con el usuario, entre menos distancia exista, será mejor.

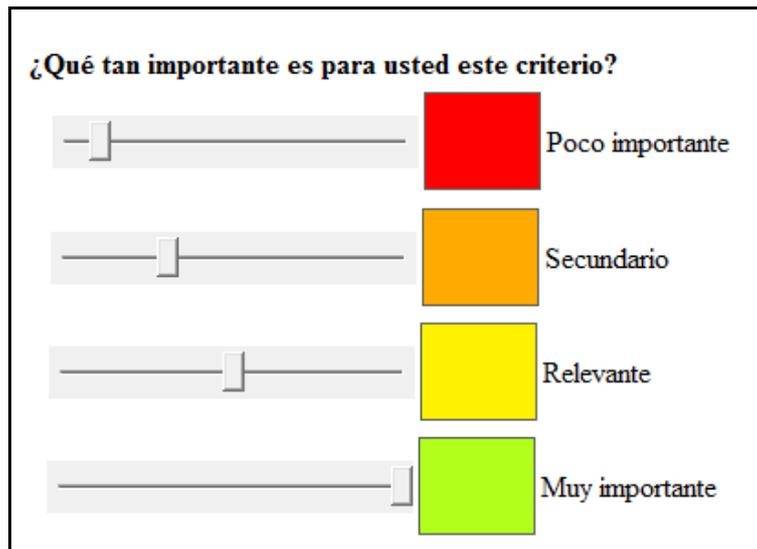


Fig. 5.3. Barra para indicar la importancia de cada criterio y semáforo de colores.

Es importante resaltar que si la ubicación del usuario no es establecida, el sistema considerara que el criterio de distancia no es importante, como si se hubiese atenuado al valor mínimo dicho criterio.

## 5.2. Evaluación del sistema de búsqueda y recuperación de información

La evaluación del sistema, contempla la interfaz del sistema y un análisis sobre los resultados obtenidos mediante una serie de consultas.

### 5.2.1. Interfaz del sistema

En esta sección, se presenta el resultado de la implementación del sistema de búsqueda y recuperación de información según la propuesta metodológica. La *Figura 5.4* muestra la interfaz gráfica de dicho sistema.



Fig. 5.4. Interfaz del sistema.

Como se muestra en la *Figura 5.4* del lado izquierdo de la interfaz existen los criterios de búsqueda relacionados con el análisis semántico, mientras que del lado derecho de la interfaz, se encuentra un mapa para que el usuario indique mediante un clic sobre el mapa cual es su posición.

### 5.2.1.1. Ubicación del usuario sobre el mapa

Para indicar cuál es la ubicación del usuario, la interfaz gráfica del sistema cuenta con un mapa digital del lado derecho, como se muestra en la *Figura 5.1*. Dicho mapa cuenta con dos modalidades: una para navegar y otra para indicar la posición. Las modalidades del mapa pueden ser seleccionadas en la parte superior del mapa, la *Figura 5.5* ilustra esto:



Fig. 5.5. Modalidades de navegación.

La modalidad de navegación (*Navegar*) permite moverse sobre el mapa, bien sea utilizando el ratón de la computadora (haciendo clic sobre el mapa y arrastrándolo) o mediante los controles de navegación disponibles para las dos modalidades. La *Figura 5.6* muestra y explica cada uno de estos controles.

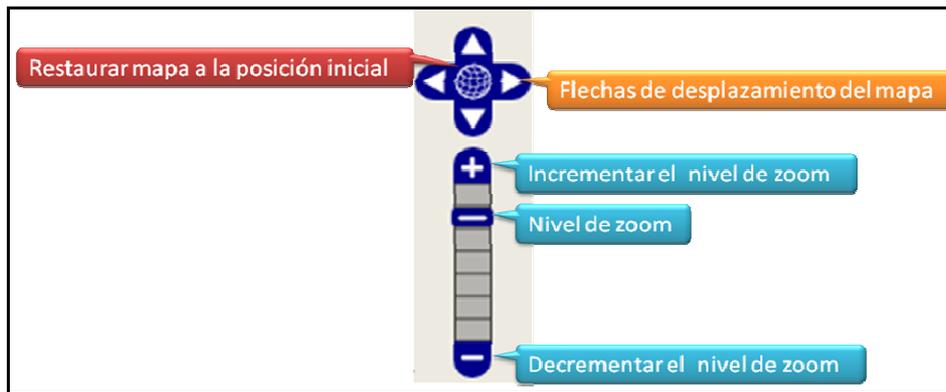


Fig. 5.6. Controles de navegación.

La modalidad de posicionamiento (*Indicar posición*) permite mediante un clic sobre el mapa, colocar un marcador (*insert feature*) indicando la posición del usuario. Si la posición del usuario ya ha sido capturada previamente, permite seleccionar dicho marcador (*select feature*) y arrastrarlo hacia otro lugar del mapa (*drag feature*), indicando así una nueva posición. La *Figura 5.7* ilustra lo anterior.



Fig. 5.7. Reubicación del usuario.

### 5.2.1.2. Criterios de búsqueda

Como se muestra en la *Figura 5.8* del lado izquierdo de la interfaz existen los criterios de búsqueda relacionados con el análisis semántico. Inicialmente solo se muestra un criterio de búsqueda, que corresponde a la clasificación del automóvil.

Como se muestra en la *Figura 5.9* la forma de ingresar la clasificación del vehículo, es mediante una serie de combos (*selects*) dependientes, que permiten

dos cosas: una fácil navegación en la estructura de información; y diferentes niveles de especificación, es decir, el usuario puede ser tan preciso como él lo decida, puede simplemente pedir un automóvil, un coche o camioneta, seleccionar el segmento al que pertenece un coche, la clase o incluso un modelo específico de vehículo.

Fig. 5.8. Criterio de clasificación del vehículo.

Para seleccionar un modelo simplemente es necesario seleccionar un tipo de vehículo. El combo de *Segmento* depende del combo de *tipo de vehículo*, mientras que, el combo de la *Clase* depende del combo de *Segmento*. Pero estos combos son solamente filtros del combo *Modelo*, es decir, si el usuario selecciona el segmento Sedán, el combo de la *Clase* se limitara a tener los valores de todas las clases que pertenecen a Sedán en la jerarquía de la clasificación de vehículos. Mientras que el combo *Modelo* tendrá todos los modelos del segmento Sedán. Pero si el usuario selecciona la clase Compacto, el combo *Modelo* se limitara a tener todos los modelos de Sedán Compactos.

Para facilitar el uso del sistema a usuarios con poco conocimiento de automóviles, todos los vehículos pueden ser listados de acuerdo a su clasificación como se mostro en la *Figura 5.8*, o de acuerdo a su marca como se muestra en la *Figura 5.9*.

Nuevamente, para seleccionar un modelo de vehículo es necesario seleccionar al menos el tipo de vehículo. De esta forma el combo *Modelo* contendrá todos los modelos existentes del tipo de vehículo seleccionado.

Si solamente se desea visualizar los automóviles de una determinada marca es necesario establecer el valor de la marca deseada en el combo *Marca*. Como se muestra en la *Figura 5.9*.

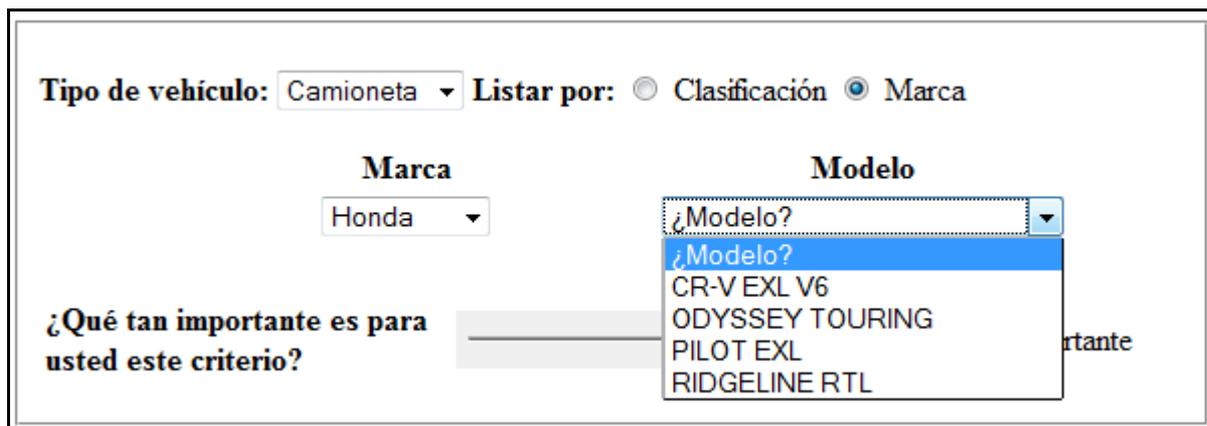


Fig. 5.9. Listado de automóviles por marca.

Es importante mencionar que aunque los vehículos sean listados de acuerdo a la marca, el criterio utilizado en la búsqueda siempre será la clasificación del vehículo. Por lo que si no se selecciona algún modelo en específico, el sistema ignorara el valor establecido en la *Marca* y utilizara solamente el *Tipo de vehículo*, para realizar la búsqueda.

El sistema propuesto, contempla dos tipos de búsqueda: la sencilla que únicamente considera el criterio de la clasificación de los automóviles y la distancia existente entre los automóviles y los usuarios; y la búsqueda avanzada que considera el resto de los criterios planteados en la metodología para este caso de estudio.

Por facilidad para los usuarios, el sistema únicamente considera la búsqueda sencilla, pero estos pueden seleccionar utilizar la búsqueda avanzada y visualizar todos los posibles criterios de búsqueda contemplados por el sistema. La *Figura 5.10* muestra las opciones de los tipos de búsqueda, que los usuarios pueden seleccionar, dicha opción se encuentra justo debajo del criterio de clasificación de los automóviles como se muestra en la *Figura 5.1*.



Fig. 5.10. Tipo de búsqueda.

Debido a que no es posible visualizar todos los criterios de búsqueda al mismo tamaño del mapa, el sistema en realidad genera el listado de criterios justo debajo del criterio de la clasificación y agrega una barra de desplazamiento del lado derecho, como se muestra en la *Figura 5.11*.

Como se muestra en la *Figura 5.12* si se selecciona la opción de búsqueda *Avanzada* el sistema muestra inmediatamente después el listado completo de criterios de búsqueda.

**Seleccione los criterios para su búsqueda:**

**Tipo de vehículo:**  **Listar por:**  Clasificación  Marca

**Marca**  **Modelo**

**¿Qué tan importante es para usted este criterio?**   Muy importante

**Tipo de búsqueda:**  Sencilla  Avanzada

**Equipamiento:**

**¿Qué tan importante es para usted este criterio?**   Muy importante

Fig. 5.II. Selección de los criterios de búsqueda.

Tipo de búsqueda:  Sencilla  Avanzada

---

**Equipamiento:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Interiores:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Transmisión:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Dirección:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Tracción:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Combustible:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Aspiración:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Configuración:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

---

**Cilindrada del motor:**  ▼

¿Qué tan importante es para usted este criterio?   Muy importante

Fig. 5.12. Listado completo de criterios para la búsqueda avanzada.

### 5.2.1.3. Semáforo de colores

Para cada criterio de búsqueda existe una ponderación, dicha ponderación se hace a través de una barra que funciona como un *semáforo de colores* indicando la importancia de cada criterio. La *Figura 5.13* ilustra lo anterior.

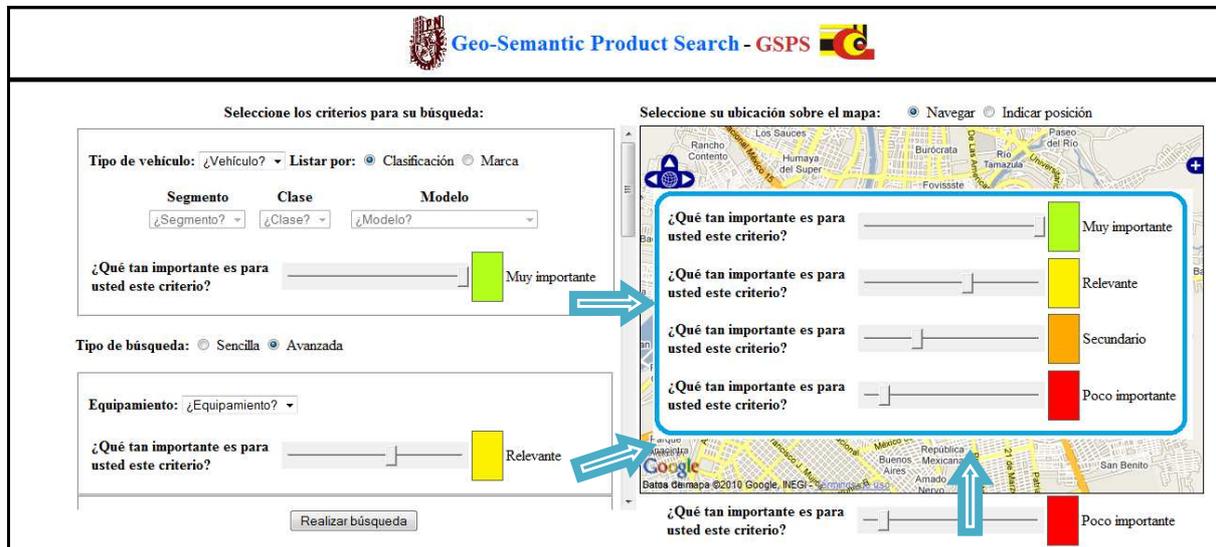


Fig. 5.13. Semáforo de colores.

### 5.2.1.4. Realizar búsqueda

Para proceder a realizar la búsqueda de información una vez establecidos todos los criterios de búsqueda de los usuarios, es necesario presionar (dar clic) el botón *Realizar búsqueda* (*Figura 5.14*) ubicado en la parte inferior izquierda de la interfaz (ver *Figura 5.4*).



Fig. 5.14. Botón para realizar la búsqueda.

### 5.2.1.5. Resultados

En este apartado se muestra como es la interfaz gráfica donde se muestran los resultados del sistema, omitiendo los detalles para llegar a estos (si desea ver lo anterior, vea el punto 5.2.2 que corresponde al análisis de resultados).

La *Figura 5.15* muestra una pantalla con los resultados de una consulta, como se puede apreciar nuevamente la pantalla se divide en dos secciones, la parte

izquierda muestra una tabla de resultados, con las columnas: *Productos*, *Detalles* y *Ubicación*. Mientras que la parte derecha exhibe nuevamente un mapa pero con una capa temática con los vendedores de automóviles, en este caso agencias automotrices.



Fig. 5.15. Listado de resultados – Interfaz gráfica.

### 5.2.1.6. Información del vendedor

Para obtener información sobre el vendedor de un automóvil, basta con dar clic sobre el link *Ir* de la columna ubicación en la tabla de resultados. Esto nos posicionara exactamente en la ubicación del vendedor, como se puede apreciar en la *Figura 5.16*.

Para obtener información detallada sobre un vendedor específico, es necesario dar clic sobre el punto en el mapa que representa a dicho vendedor, y posteriormente aparecerá una ventana tipo *popup* con la información detallada del vendedor. La *Figura 5.17* ilustra lo anterior.



Fig. 5.16. Ubicación de un vendedor.



Fig. 5.17. Información sobre un vendedor.

### 5.2.1.7. Detalles del producto

Para obtener información detallada sobre cada uno de los resultados obtenidos, es necesario dar clic en el link *Ver* de la columna *Detalles* (ver *Figura 5.15*). Una vez que se ha dado clic sobre el enlace, el sistema abrirá una nueva ventana, donde se muestra la descripción del automóvil, tal como se puede apreciar en la *Figura 5.18*.



The screenshot displays the 'Geo-Semantic Product Search - GSPS' interface. At the top left is a logo with the letters 'GSPS' and a gear icon. To its right is the text 'Geo-Semantic Product Search - GSPS' and a stylized 'G' logo. Below this is a section titled 'Descripción del Producto'. It contains two boxes: 'General' and 'Atributos'. The 'General' box lists: Producto: Camry, Descripción: Camry XLE V6, Marca: Toyota, and Precio: \$355700.00. The 'Atributos' box lists: Clasificación: sedan lujo, Equipamiento: lujo, Interiores: piel tradicional, Transmisión: automatica con cambios secuenciales, Dirección: hidraulica, Tracción: delantera, Cilindrada del motor: 6 cilindros, Combustible: gasolina, Aspiración: atmosfera-normal, and Configuración: v.

Fig. 5.18. Descripción del producto.

### 5.2.2. Análisis de resultados

En este apartado se analizarán los resultados arrojados por el sistema después de una serie de consultas. Dichas consultas se muestran a continuación:

La *Figura 5.19* muestra una consulta simple donde se busca un vehículo NISSAN 350Z Touring.



Fig. 5.19. Búsqueda de un Nissan 350Z.

La *Tabla 5.1* muestra los resultados de la consulta anterior, evidenciando una clara relación en los datos, pues al buscar un automóvil: “NISSAN 350Z TOURING” el sistema además de obtener estos resultados con una gran precisión recomienda además otro tipo de automóviles deportivos como lo son:

- Mustang Shelby
- Corvette
- Camaro
- Y posteriormente otros automóviles que podrían acercarse a ser deportivos aunque no lo son.

Tabla 5.1. Resultados de la consulta simple de un Nissan 350Z.

370Z TOURING TRANSMISIÓN AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
370Z TOURING TRANSMISIÓN AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang Shelby GT 500	<a href="#">Ver</a>	<a href="#">Ir</a>
Corvette 2 ptas. Coupe Paq. A Automático	<a href="#">Ver</a>	<a href="#">Ir</a>
Corvette 2 ptas. Coupe Paq. A Automático	<a href="#">Ver</a>	<a href="#">Ir</a>
Camaro Paquete C	<a href="#">Ver</a>	<a href="#">Ir</a>
Camaro Paquete C	<a href="#">Ver</a>	<a href="#">Ir</a>
370Z TOURING TRANSMISIÓN AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Camaro Paquete C	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang Shelby GT 500	<a href="#">Ver</a>	<a href="#">Ir</a>
370Z TOURING TRANSMISIÓN AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Corvette 2 ptas. Coupe Paq. A Automático	<a href="#">Ver</a>	<a href="#">Ir</a>
Accord Coupé EX V6	<a href="#">Ver</a>	<a href="#">Ir</a>
ALTIMA COUPE SR 3.5 LTS./TRANSMISIÓN CVT	<a href="#">Ver</a>	<a href="#">Ir</a>
ALTIMA COUPE SR 3.5 LTS./TRANSMISIÓN CVT	<a href="#">Ver</a>	<a href="#">Ir</a>
Civic Coupé EXL AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Civic Si COUPÉ Manual	<a href="#">Ver</a>	<a href="#">Ir</a>

Pero ahora bien, ¿qué pasa si la consulta se realiza únicamente considerando el criterio de distancia y no la clasificación del automóvil? La *Figura 5.20* ilustra una consulta con este caso:



Fig. 5.20. Consulta con el criterio de la distancia.

Los resultados a la consulta de la *Figura 5.20* son mostrados en la *Tabla 5.2*, dichos resultados muestran que al no haber seleccionado otro criterio más que el de la distancia el sistema arroja los resultados más cercanos a este según su posición. En este caso al ser una agencia de Honda la más cercana a la ubicación del usuario, los primeros resultados hacen alusión a dicha agencia Y conforme la distancia se va haciendo más grande empiezan a aparecer vehículos de otras compañías como se muestra en la parte inferior de la tabla, donde aparecen vehículos de la marca Ford.

La *Figura 5.21* ilustra cual es la posición de la agencia Honda, dicha agencia es la que geográficamente se encuentra más cercana a la posición del usuario considerada para la consulta en la *Figura 5.20*.

Tabla 5.2. Resultados utilizando únicamente la distancia.

Accord Coupé EX V6	<a href="#">Ver</a>	<a href="#">Ir</a>
Cívic Sedán EXL AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Fit EX AUTOMATICA	<a href="#">Ver</a>	<a href="#">Ir</a>
City EX AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
PILOT EXL	<a href="#">Ver</a>	<a href="#">Ir</a>
RIDGELINE RTL	<a href="#">Ver</a>	<a href="#">Ir</a>
Accord Coupé EX V6	<a href="#">Ver</a>	<a href="#">Ir</a>
Accord Crosstour EXL V6	<a href="#">Ver</a>	<a href="#">Ir</a>
CR-V EXL V6	<a href="#">Ver</a>	<a href="#">Ir</a>
Cívic Si COUPÉ Manual	<a href="#">Ver</a>	<a href="#">Ir</a>
Cívic Hybrid CVT	<a href="#">Ver</a>	<a href="#">Ir</a>
Cívic Coupé EXL AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
ODYSSEY TOURING	<a href="#">Ver</a>	<a href="#">Ir</a>
Accord Sedán EX V6	<a href="#">Ver</a>	<a href="#">Ir</a>
Fusion Sport V6	<a href="#">Ver</a>	<a href="#">Ir</a>
Ranger XLT Cab. doble 4 puertas AA	<a href="#">Ver</a>	<a href="#">Ir</a>
Edge Limited DVD	<a href="#">Ver</a>	<a href="#">Ir</a>



Fig. 5.21. Ubicación de la agencia Honda Visión

Los resultados obtenidos anteriormente muestran únicamente resultados utilizando un solo criterio, pero realmente que sucede cuando estos criterios se combinan, a continuación se muestra una consulta donde se combinan los criterios de búsqueda anteriores (distancia y clasificación del automóvil), donde la clasificación tiene una mayor ponderación por parte del usuario (ver *Figura 5.22*).



Fig. 5.22. Combinación de los criterios de distancia y clasificación de los automóviles.

Tabla 5.3. Resultados utilizando únicamente de los criterios de distancia y clasificación.

Cruze Paquete F	<a href="#">Ver</a>	<a href="#">Ir</a>
Cruze Paquete F	<a href="#">Ver</a>	<a href="#">Ir</a>
Cruze Paquete F	<a href="#">Ver</a>	<a href="#">Ir</a>
SENTRA SE-R SPEC V	<a href="#">Ver</a>	<a href="#">Ir</a>
Avalanche Paq. C Z71	<a href="#">Ver</a>	<a href="#">Ir</a>
Suburban SUV Paq. G 2500 (3/4 ton)	<a href="#">Ver</a>	<a href="#">Ir</a>
ARMADA SE 4x4 PIEL/QUEMACOCOS/TRANSMISIÓN AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang Shelby GT 500	<a href="#">Ver</a>	<a href="#">Ir</a>
Corvette 2 ptas. Coupe Paq. A Automático	<a href="#">Ver</a>	<a href="#">Ir</a>
MURANO LE AWD/TRANSMISIÓN CVT	<a href="#">Ver</a>	<a href="#">Ir</a>
PLATINA CUSTOM TRANSMISIÓN MANUAL/AIRE ACONDICIONADO	<a href="#">Ver</a>	<a href="#">Ir</a>
Fit EX AUTOMATICA	<a href="#">Ver</a>	<a href="#">Ir</a>
Cheyenne 4x4 Crew Cab LTZ Paq. C	<a href="#">Ver</a>	<a href="#">Ir</a>
TITAN CREW CAB LE 4X4 PIEL/TRANSMISIÓN AUTOMÁTICA/QUEMACOCOS	<a href="#">Ver</a>	<a href="#">Ir</a>
Malibu Paquete G	<a href="#">Ver</a>	<a href="#">Ir</a>
Tahoe SUV Paq. E EZ71	<a href="#">Ver</a>	<a href="#">Ir</a>
City EX AUTOMÁTICA	<a href="#">Ver</a>	<a href="#">Ir</a>

Como se puede apreciar en los resultados de la *Tabla 5.3* al combinar los criterios de búsqueda, el sistema primero obtiene aquellos resultados cuya ponderación es mayor, pues dicha ponderación refleja un mayor interés por parte de los usuarios y es considerada al momento de realizar la recuperación de los resultados.

Finalmente, el sistema está diseñado para trabajar con diferentes tipos de usuarios, pueden existir usuarios neófitos o no expertos y usuarios con un mayor grado de conocimiento sobre el dominio de los automóviles. Para el caso de los usuarios comunes (la mayoría) se recomienda utilizar únicamente los criterios de clasificación del automóvil para que el sistema le arroje resultados similares a los que él está buscando y el criterio de la distancia, para obtener primero aquellos que estén más cercanos. La combinación de estos criterios puede ofrecer resultados bastante alentadores y útiles para los usuarios pero es importante recordar que los resultados de una misma consulta pueden verse afectados drásticamente según la ponderación (preferencias) que el usuario asigne a cada criterio.

A continuación se muestra otra consulta utilizando criterios más avanzados donde se especifican características muy técnicas para un automóvil.

Ahora bien, veamos que sucede cuando la búsqueda es más compleja y el usuario busca lo siguiente:

- Una Pickup chica
- Equipamiento de lujo
- Transmisión automática
- Dirección Hidráulica
- Tracción 4x4
- De 8 cilindros
- Motor en V

Para la búsqueda anterior considerando que todos los criterios son “Muy importantes” el sistema retorna la tabla de resultados mostrados en la *Tabla 5.4*, en este conjunto de resultados es importante resaltar que la expresión de relevancia utilizada en el modelo de recuperación, traerá al usuario un conjunto de resultados ordenados de acuerdo al número de criterios que satisfaga cada uno de los vehículos, siendo entonces, más importantes aquellos que cumplen con más criterios pues su ponderación es la misma.

Tabla 5.4. Resultados de una consulta compleja.

RIDGELINE RTL	<a href="#">Ver</a>	<a href="#">Ir</a>
F-250 SUPER DUTY Crew Cab Diesel 4x4 TA	<a href="#">Ver</a>	<a href="#">Ir</a>
F-150 XL REG CAB 4x4	<a href="#">Ver</a>	<a href="#">Ir</a>
Cheyenne 4x4 Crew Cab LTZ Paq. C	<a href="#">Ver</a>	<a href="#">Ir</a>
Cheyenne 4x4 Crew Cab LTZ Paq. C	<a href="#">Ver</a>	<a href="#">Ir</a>
Explorer Sport Track V8 4x2	<a href="#">Ver</a>	<a href="#">Ir</a>
F-250 SUPER DUTY Crew Cab Diesel 4x4 TA	<a href="#">Ver</a>	<a href="#">Ir</a>
F-150 XL REG CAB 4x4	<a href="#">Ver</a>	<a href="#">Ir</a>
Lobo KR cab. doble 4x4	<a href="#">Ver</a>	<a href="#">Ir</a>
Lobo Crew Cab 4x4 Lariat 5.4L	<a href="#">Ver</a>	<a href="#">Ir</a>
ODYSSEY TOURING	<a href="#">Ver</a>	<a href="#">Ir</a>
Silverado 1500 Paquete G Cabina Regular	<a href="#">Ver</a>	<a href="#">Ir</a>
Cheyenne 4x4 Crew Cab LTZ Paq. C	<a href="#">Ver</a>	<a href="#">Ir</a>
Silverado 1500 Paquete G Cabina Regular	<a href="#">Ver</a>	<a href="#">Ir</a>
Explorer Sport Track V8 4x2	<a href="#">Ver</a>	<a href="#">Ir</a>
FRONTIER SE V6 4X4	<a href="#">Ver</a>	<a href="#">Ir</a>
Econoline E350 XL V8 15 pasajeros	<a href="#">Ver</a>	<a href="#">Ir</a>
Tahoe SUV Paq. E EZ71	<a href="#">Ver</a>	<a href="#">Ir</a>

Y que sucede cuando no existen automóviles que no cumplen con determinados criterios, por ejemplo si se busca un automóvil de 10 cilindros y el sistema no tiene almacenado ninguno, como se muestra en la *Figura 5.23*:

Cilindrada del motor: 10 cilindros

¿Qué tan importante es para usted este criterio? Muy importante

Fig. 5.23. Búsqueda de un automóvil de 10 cilindros.

La *Tabla 5.5* muestra los resultados de buscar un automóvil de 10 cilindros y dando una ponderación mínima al resto de los criterios (es decir, solo se utiliza la cilindrada del motor). La peculiaridad de esta búsqueda es que no existen registros de automóviles en la base de datos de pruebas con dicha característica. Presentándose entonces una situación difícil de manejar para los sistemas clásicos de búsqueda y recuperación de información.

Tabla 5.5. Resultados obtenidos a partir de una consulta donde no existe correspondencia en la BD.

Fiesta Trend Conford TA	<a href="#">Ver</a>	<a href="#">Ir</a>
Fiesta Trend Conford TA	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus Sport TA	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus Sport TA	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus RS	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus RS	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus 5 Puertas	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus 5 Puertas	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus ST	<a href="#">Ver</a>	<a href="#">Ir</a>
Focus ST	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang GT Premium Convertible TA/TS	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang GT Premium Convertible TA/TS	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang Shelby GT 500	<a href="#">Ver</a>	<a href="#">Ir</a>
Mustang Shelby GT 500	<a href="#">Ver</a>	<a href="#">Ir</a>
Fusion SEL V6 Ford Interactive System	<a href="#">Ver</a>	<a href="#">Ir</a>
Fusion SEL V6 Ford Interactive System	<a href="#">Ver</a>	<a href="#">Ir</a>
Fusion Sport V6	<a href="#">Ver</a>	<a href="#">Ir</a>
Fusion Sport V6	<a href="#">Ver</a>	<a href="#">Ir</a>

Debido a las características propias de *confusión* el sistema es capaz de obtener resultados aun cuando no existen automóviles que no cumplan en 0% con lo que se está solicitando. Esto es muy importante pues cumple con la tendencia de los sistemas modernos de recuperación de información (evitando resultados vacíos) y mejora los sistemas comerciales, evitando que las compañías pierdan a posibles clientes al presentarse esta situación que resulta común, prácticamente para cualquier sistema de búsqueda.

Además entre más criterios se utilizan la probabilidad de hacer match con alguno de estos incrementa, siendo esta otra de las razones para incorporar al modelo de recuperación mayor numero de criterios de búsqueda basados en la semántica del dominio.

### 5.3. Comentarios finales

Los resultados obtenidos muestran que el sistema tiene la capacidad de ajustarse a las preferencias de los usuarios, mediante la ponderación de cada criterio de

búsqueda; logrando que los resultados obtenidos sean más satisfactorios para los usuarios, pues estos son ajustados de acuerdo a lo que el usuario desea.

Haciendo uso de la medida de similitud semántica *confusión* es posible recuperar resultados que no satisfacen en un 100% las consultas de los usuarios, pero que satisfacen con un determinado grado de error (*confusión*) dichas consultas.

Al aprovechar la monotonía de *confusión* es posible evitar resultados vacíos, logrando siempre ofrecer resultados a los usuarios aún cuando no existan resultados que satisfagan sus consultas.

Finalmente, se listan algunos puntos que detallan las bondades expuestas por el sistema de búsqueda y recuperación de información:

- El uso del análisis semántico hace posible, que el sistema pueda recomendar u ofrecer resultados que: aún cuando el usuario explícitamente no los haya solicitado, pueden resultar útiles para este.
- El utilizar múltiples atributos o características cualitativas para describir un mismo objeto de información, en definitiva, enriquece y potencializa las capacidades del sistema para poder obtener dichos resultados.
- La integración de múltiples criterios de búsqueda, permite evitar resultados vacíos y presentar alternativas al usuario, tanto desde el punto de vista geoespacial, como de la semántica (significado) inherente a la información.
- La integración de resultados de acuerdo a su nivel de relevancia, facilita la interpretación y disminuye los tiempos de análisis de resultados para los usuarios finales del sistema.

## **CAPÍTULO 6.**

# **CONCLUSIONES Y TRABAJO A FUTURO**

En este capítulo se describen las conclusiones derivadas del desarrollo de este trabajo de tesis, dentro del área de recuperación de información geográfica (GIR) basado en variables cualitativas.

### **6.1. Conclusiones**

Es importante resaltar que el conjunto de resultados obtenidos por el sistema es ordenado en función de su relevancia y que estos resultados permiten una mejor interpretación a los usuarios.

Los resultados demuestran que a pesar de no existir una coincidencia exacta entre la información buscada por los usuarios puede existir información similar que satisfaga con un cierto grado de error las búsquedas de los usuarios. Lo anterior confirma que la metodología propuesta permite obtener resultados similares a los que un usuario está buscando, en adición a los resultados que satisfacen en un 100% las consultas de los usuarios.

El poder recuperar información similar a la que un usuario está buscando, permite ayudar a evitar resultados vacíos.

Un objeto de información puede ser representado a partir de sus principales características (*features*). El modelo propuesto, considera el uso de múltiples descriptores como características cualitativas de un objeto.

La metodología propuesta puede ser aplicada a la recuperación de productos o servicios de otros dominios, es decir, aunque el caso de estudio utilizado en este trabajo sea la comercialización de vehículos por internet, puede ser expandido a otros dominios (por ejemplo: bienes raíces), pudiendo agregar criterios de búsqueda dependientes a los nuevos dominios.

El vincular la semántica del dominio en el modelo de recuperación, incrementa enormemente las posibilidades de reducir los resultados vacíos en los sistemas de recuperación. Esta característica dota al modelo de recuperación con la capacidad de ofrecer resultados que sin haber sido solicitados explícitamente, pueden resultar útiles a los usuarios.

## 6.2. Aportaciones

En este apartado se mencionan algunas de las principales contribuciones de este trabajo:

- Se propuso un modelo de búsqueda y recuperación de productos o servicios considerando sus ubicaciones geográficas, dotando al sistema con la capacidad para realizar un análisis espacial y poder proporcionar a resultados que se encuentren más cercanos (menos distantes) a los usuarios.
- El modelo de recuperación propuesto, considera el uso de múltiples características cualitativas que describen a un mismo producto, permitiendo mediante el uso de medidas de similitud semántica realizar un análisis semántico y recuperar productos similares a los que un usuario está buscando aún cuando no exista una coincidencia exacta (resultados que satisfagan en un 100% las consultas de los usuarios).
- Los resultados arrojados por el sistema de recuperación de información pueden ser personalizados por los usuarios mediante la ponderación de cada criterio de búsqueda. Permitiendo que una misma búsqueda arroje diferentes resultados según las preferencias de cada usuario en particular.

- Se propuso un mecanismo de metadatos para almacenar información acerca de los descriptores de los productos dentro de la base de datos espacial, que dota de flexibilidad al modelo para incorporar nuevos productos o servicios pertenecientes a otros dominios (i.e. diferentes al de la compra venta de vehículos online).
- Se implementó un mecanismo para el cálculo de *confusión* en Java sobre archivos OWL que almacenan jerarquías, considerando el cálculo en: *jerarquías simples*, *jerarquías ordenadas*, así como, una combinación de estas (*jerarquías simples y ordenadas*) denominadas *jerarquías sub-mixtas* para diferenciarlas de las jerarquías mixtas propuestas en [46].
- Se propuso una variante a la *Teoría de Confusión* para obtener resultados más precisos en el cálculo de *confusión*, dicha modificación permite considerar el nivel de abstracción de los conceptos dentro de las jerarquías, haciendo que la diferencia (*confusión*) entre conceptos más generales (superiores) sea mayor que la diferencia entre conceptos más específicos (conceptos inferiores de la jerarquía).
- Se propuso utilizar una expresión de *ranking* para evaluar la relevancia de los resultados obtenidos por el modelo de recuperación de información, generando un listado único de resultados.

### 6.3. Trabajo a futuro

- Desarrollar una interfaz dinámica que se auto configure, utilizando los metadatos del sistema para obtener un buscador genérico de todos los productos almacenados en la base de datos.
- Experimentar con la metodología propuesta, introduciendo al sistema productos de otros dominios.
- Incluir un mecanismo para el cálculo de la distancia que sea más preciso y considere otros factores en un entorno socio-urbano, con la finalidad de que no sólo consista en el cálculo de la distancia geométrica.
- Experimentar con nuevas medidas de similitud semántica que permitan realizar un procesamiento más “inteligente” de la información, que permitan trabajar con ontologías y relaciones más complejas en las

estructuras conceptuales, con la finalidad realizar inferencias que se apeguen más a la “intuición o razonamiento humano”.

- Implementar mecanismos de retroalimentación (*feed-back*) que le permitan al modelo “aprender” de las búsquedas pasadas de los usuarios, buscando identificar una especie de “inteligencia colectiva” con la finalidad de realizar recomendaciones automáticas en la ponderación de los criterios de búsqueda, para que esta responsabilidad no recaiga completamente en los usuarios, sino que tengan la opción de que el sistema automáticamente les proponga la ponderación de todos los criterios que intervengan en sus búsquedas, una alternativa para realizar esto es utilizar el método de *boosting*.
- Incluir una ontología del dominio geográfico que permita fortalecer el análisis de las relaciones espaciales, considerando otros factores como las relaciones topológicas para el análisis de información.
- Dotar al sistema con la capacidad de trabajar con más criterios numéricos, e incluir otro tipo de medidas de similitud no semánticas para el procesamiento de estos criterios de búsqueda.
- Buscar y proponer una nueva expresión de ranking, que sea más efectiva y que permita evaluar de mejor forma la relevancia de los resultados.

# REFERENCIAS

- [1] Capgemini, Cars Online 08/09 10th Annual Global Automotive Study: Tracking Consumer Buying Behavior in Both Mature and Emerging Markets [en línea] <[http://api.ning.com/files/9GUK4tWWN3vpsXOdeIcMqTRQq5eJjssJQHlnKyzcDGECp\\*vWYlf8i7kO0gXL5Cqa0KoOUxIuG\\*f7VKnoAeErQ7hntLEijtso/tl\\_Cars\\_Online\\_08\\_09.pdf](http://api.ning.com/files/9GUK4tWWN3vpsXOdeIcMqTRQq5eJjssJQHlnKyzcDGECp*vWYlf8i7kO0gXL5Cqa0KoOUxIuG*f7VKnoAeErQ7hntLEijtso/tl_Cars_Online_08_09.pdf)> [consulta: 03 de agosto del 2010].
- [2] Torres M. and Levachkine S. Towards a methodology to conceptualize the geographic domain, En: A. Gelbukh & E.F. Morales (Eds.). Lecture Notes in Artificial Intelligence, Vol. 5317, Berlin Heidelberg, Springer-Verlag, 2008. pp. 111-122.
- [3] R. Baeza-Yates, B. Ribeiro-Nieto, Modern Information Retrieval. ACM press, 1999.
- [4] Frakes, W., Baeza-Yates, R. Information Retrieval: Data Structures & Algorithms. Prentice-Hall, 1992.
- [5] Vestavik, Ø. Geographic Information Retrieval: An Overview [en línea] <<http://idi.ntu.no/~oyvindve/article.pdf>> [consulta: 10 de enero del 2010].
- [6] Teknomo, K. Similarity Measurement [en línea] <<http://people.revoledu.com/kardi/tutorial/Similarity/index.html>> [consulta: 31 de mayo del 2009].
- [7] Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. Psychometrika 27(2): 125-140, 1962.
- [8] Tversky, A. Features of similarity. Psychological Review 84(4): 327-352, 1977.
- [9] Rubenstein, H., Goodenough, J. B. Contextual Correlate of Synonymy. Communication of the ACM 8(10):627-663, 1965.
- [10] Miller, G. A., Charles, W. G. Contextual Correlate of Semantic Similarity. Language and Cognitive Processes 6(1):1-28, 1991.

- [11] Resnik, P. Selection and Information: A Class-Based Approach to Lexical Relationships, Tesis (Ph. D. Dissertation, in Computer and Information Science), Pennsylvania, USA. University of Pennsylvania, 1993.
- [12] Lin, D. An information-theoretic definition of similarity. Proceedings of the Fifteenth International Conference on Machine Learning, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc, 1998. pp. 296-304.
- [13] Leacock, C., Chodorow, M. Combining local context and WordNet similarity for word sense identification. Computational Linguistics: Special issue on word sense disambiguation. 24(1):147-165, March 1998.
- [14] Hirts, G., St-Onge, D. Lexical chains as representation of context for the detection and correction of malapropisms. En: Fellbaum, C., WordNet: An Electronic Lexical Database. Cambridge, MA, MIT Press, 1998. pp. 265-283.
- [15] Blázquez del Toro, J. M., Arias, j. Luque, V. Sánchez L. Ordenación de resultados en consultas semánticas sobre bases de datos. Actas de Talleres de Ingeniería del Software y Bases de Datos Madrid. 1(3):22-30, 2003.
- [16] Rada, R., Mili, H., Bicknell, E., Blettner, M. Development and Application of a Metric on Semantic Nets. IEEE Transaction on Systems, Man and Cybernetics. 9(1):17-30, 1989.
- [17] Kim, Y. W., Kim J. H. A Model of Knowledge Based Information Retrieval with Hierarchical Concept Graph. Journal of Documentation 46(2):113-136, 1990.
- [18] Ganesan, P., Garcia-Molina, H., Widom, J. Exploiting hierarchical domain structure to compute similarity. ACM Trans. Information Systems. 21(1):64-93, 2003.
- [19] Sarabia-López, G. Búsqueda y Ponderación de Información Contenida en Bases de Datos Espaciales, Utilizando Jerarquías. Tesis (Maestría en Ciencias de la Computación). México D.F., Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), 2008.
- [20] Feldman, R. and Dagan, I. Knowledge Discovery in Textual Databases (KDT). En: Proceedings of the 1<sup>st</sup> International Conference on Knowledge Discovery (Montreal, 1995). pp. 112-117.

- [21] Scott, S. and Matwin, S. Text classification using WordNet hypernyms. Proceedings of the Use of WordNet in Natural Language Processing Systems. Association for Computational Linguistics. 1998.
- [22] Berners-Lee, T. and Hendler, J. The Semantic Web: A new form of Web. Content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American, 2001.
- [23] W3C, Semantic Web Activity [en línea] <<http://www.w3.org/2001/sw/>> [consulta: 10 de junio del 2010].
- [24] Egenhofer, M. J. Toward the semantic geospatial web. En: Proceedings of the 10th ACM international symposium on Advances in geographic information systems. 2002, pp. 1-4.
- [25] Mata-Rivera, M. F. Recuperación y Ponderación de Información Geográfica desde Repositorios No Estructurados Conducidas por Ontologías. Tesis (Doctorado en Ciencias de la Computación). México D.F., Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), 2009.
- [26] Ibarra-Vargas, J. J. Análisis Espacial en Espacios Conceptuales. Tesis (Maestría en Ciencias de la Computación). México D.F., Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), 2009.
- [27] Renteria-Agualimpia, W. Recuperación Controlada de Información Cualitativa desde Repositorios de Datos. Tesis (Maestría en Ciencias de la Computación). México D.F., Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), 2009.
- [28] Larson, R. Geographic Information Retrieval and Spatial Browsing. En: Smith, L. C., Gluck M. (Eds). Geographic Information Systems Patrons Maps and Spatial Information. 1995, pp. 81-123.
- [29] Hill, L., Core Elements of Digital Gazetteers: Placenames, Categories and Footprints. En: Borbinha, J. and Baker, T. (Eds.). Research and Advanced Technology for Digital Libraries, proceedings 2000.
- [30] Arctur, D., Zeiler, M. Designing Geodatabases: Cases Studies in GIS Data Modeling, ESRI Press, 2004.

- [31] Davis S., GIS For Web Developers: Adding Where to your Web Applications. The Pragmatic Programmers- LLC, 2007.
- [32] Aristotle's. Methaphysics, IV-1, The Subject Matter of Aristotle's Methaphysics. Stanford Encyclopedia of Philosophy. 2003.
- [33] Studer, S., Benjamins, R. and Fensel, D. Knowledge Engineering: Principles and Methods. Data and Knowledge Engineering. 25(1), 1998. pp. 167- 197.
- [34] Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, R. W. Enabling Technology for Knowledge Sharing. AI Magazine. 12(3), 1991.
- [35] Gruber, T. A translation approach to portable ontology specifications. Knowledge Acquisition. 5(2), 1993. pp. 199-220.
- [36] Guarino, N. and Giaretta, P. 1995. Ontologies and knowledge bases: towards a terminological clarification. En: Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (1995). pp. 25-32.
- [37] Guarino, N. 1998. Formal ontology and information systems. En: Proceedings of FOIS'98: 6-8 June 1998. Trento, Italy. Amsterdam, IOS Press. pp. 3-15.
- [38] Bernaras, A., Laresgoiti, I. and Corera, J. 1996. Building and Reusing Ontologies for Electrical Network Applications. En: Proceedings of the European Conference on Artificial Intelligence (1996). pp. 298-302.
- [39] Swartout, B., Patil, R. and Knight, K. 1997. Toward distributed use of largescale ontologies. In AAAI-97 Spring Symposium Series on Ontological Engineering. 1997.
- [40] Uschold, M. and Grüninge, M. Ontologies: Principles, methods and applications. Knowledge Engineering Review. 11(2), 1996.
- [41] Borst, W. N. Construction of engineering ontologies for knowledge sharing and reuse. Tesis (Ph. D. in Information Knowledge Systems). Benalmádena, Centre for Telematics and Information Technology, Universiteit Twente, 1997.
- [42] Beltrán-Martínez, B., Guzmán-Arenas, A., Martínez-Trinidad, F., Ruiz-Shulcloper, J. 1998. Clasitex++: una herramienta para el análisis de textos.

En: Memorias del Tercer Taller Iberoamericano de Reconocimiento de Patrones (1998). pp. 369-379.

[43] Adolfo Guzmán-Arenas and Victor-Polo de Gyves. A distributed digital text accessing and acquisition system: BiblioDigital SoftwarePro International. Lecture Notes in Computer Science Vol. 3061, Springer Verlag, 2004. pp. 274-283.

[44] Guzman, A., and Olivares, J. (2004) Finding the Most Similar Concepts in two Different Ontologies. Lecture Notes in Artificial Intelligence Vol. 2972, Springer Verlag. pp. 129-138.

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

[46] Levachkine, S. and Guzmán –Arenas, A. Hierarchies measuring qualitative variables. Lecture Notes in Computer Science. Vol. 2945. 2004, pp. 262-274.

[47] Liu, H., Motoda, H. Feature transformation and subset selection. IEEE Intelligent Systems. 13(2), 1988. pp. 26-28.

[48] Protégé. The Protégé Ontology Editor and Knowledge Acquisition System [en línea] <<http://protege.stanford.edu/>> [consulta: 20 de agosto del 2009].

[49] AMIA, Asociación Mexicana de la Industria Automotriz [en línea] <<http://www.amia.com.mx/>> [consulta: 20 de septiembre del 2009]

[50] Eco-Vehículos, Portal de Indicadores de Eficiencia Energética y Emisiones Vehiculares [en línea] < <http://www.ecovehiculos.gob.mx/>> [consulta: 10 de diciembre del 2009].

[51] MathWorld, Spherical Distance [en línea] <<http://mathworld.wolfram.com/SphericalDistance.html>> [consulta: 10 de diciembre del 2010].

## ANEXO A.

# Jerarquías

En este anexo se muestran los archivos OWL de las jerarquías utilizadas para estructurar la información sobre los descriptores de los automóviles utilizados por el sistema de búsqueda y recuperación de información. Dichas jerarquías fueron implementadas en *Protege 3.4*, facilitando su interoperabilidad y difusión.

A continuación se muestra la jerarquía para la clasificación de vehículos, de acuerdo al tipo de vehículo, su segmento, tamaño o clasificación (basado en la Asociación Mexicana de la Industria Automotriz):

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns="http://www.owl-ontologies.com/clasificacion.owl#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/clasificacion.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="van_grande">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="van"/>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
```

```

    <owl:DatatypeProperty rdf:ID="posicion"/>
  </owl:onProperty>
  <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
    >3</owl:hasValue>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="pickup">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="camioneta"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="convertible_deportivo">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="convertible"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="suv">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#camioneta"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="coche">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="automovil"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="pickup_chica">
  <rdfs:subClassOf rdf:resource="#pickup"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#convertible">
  <rdfs:subClassOf rdf:resource="#coche"/>
</owl:Class>
<owl:Class rdf:ID="convertible_lujo">
  <rdfs:subClassOf rdf:resource="#convertible"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </owl:Restriction>

```

```

    <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
    >2</owl:hasValue>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="sedan_compacto">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="sedan"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="van_mediana">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#van"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="hatchback_compacto">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="hatchback"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="coupe_deportivo">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="coupe"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >3</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```

    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="sedan_lujo">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#sedan"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >3</owl:hasValue>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="suv_grande">
  <rdfs:subClassOf rdf:resource="#suv"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >3</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="suv_chica">
  <rdfs:subClassOf rdf:resource="#suv"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#hatchback">
  <rdfs:subClassOf rdf:resource="#coche"/>
</owl:Class>
<owl:Class rdf:ID="coupe_lujo">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#coupe"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```

<owl:Class rdf:ID="pickup_mediana">
  <rdfs:subClassOf rdf:resource="#pickup"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#van">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#camioneta"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#coupe">
  <rdfs:subClassOf rdf:resource="#coche"/>
</owl:Class>
<owl:Class rdf:ID="sedan_subcompacto">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#sedan"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="pickup_grande">
  <rdfs:subClassOf rdf:resource="#pickup"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >3</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#sedan">
  <rdfs:subClassOf rdf:resource="#coche"/>
</owl:Class>
<owl:Class rdf:ID="van_chica">
  <rdfs:subClassOf rdf:resource="#van"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>

```

```

    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="coupe_compacto">
  <rdfs:subClassOf rdf:resource="#coupe"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="hatchback_subcompacto">
  <rdfs:subClassOf rdf:resource="#hatchback"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="suv_mediana">
  <rdfs:subClassOf rdf:resource="#suv"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >2</owl:hasValue>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#camioneta">
  <rdfs:subClassOf rdf:resource="#automovil"/>
</owl:Class>
<owl:DatatypeProperty rdf:about="#posicion">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
  <rdfs:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
</owl:DatatypeProperty>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

A continuación se muestra la jerarquía sobre los tipos de combustible utilizados en los automóviles comercializados en México:

```

<?xml version="1.0"?>
<rdf:RDF

```

```

xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
xmlns:swrl="http://www.w3.org/2003/11/swrl#"
xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://www.owl-ontologies.com/Combustible.owl#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xml:base="http://www.owl-ontologies.com/Combustible.owl">
<owl:Ontology rdf:about=""/>
<owl:Class rdf:ID="combustible"/>
<owl:Class rdf:ID="gas">
  <rdfs:subClassOf rdf:resource="#combustible"/>
</owl:Class>
<owl:Class rdf:ID="hibrido">
  <rdfs:subClassOf rdf:resource="#combustible"/>
</owl:Class>
<owl:Class rdf:ID="diesel">
  <rdfs:subClassOf rdf:resource="#combustible"/>
</owl:Class>
<owl:Class rdf:ID="gasolina">
  <rdfs:subClassOf rdf:resource="#combustible"/>
</owl:Class>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

## Jerarquía sobre los tipos de transmisiones:

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/Transmision.owl#"
  xml:base="http://www.owl-ontologies.com/Transmision.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="continuamente_variable-cvt">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="automatica"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="transmision"/>
  <owl:Class rdf:ID="automatica_tradicional">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#automatica"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#automatica">
    <rdfs:subClassOf rdf:resource="#transmision"/>
  </owl:Class>

```

```

</owl:Class>
<owl:Class rdf:ID="automatica_con_sobremarcha">
  <rdfs:subClassOf rdf:resource="#automatica"/>
</owl:Class>
<owl:Class rdf:ID="automatica_con_cambios_secuenciales">
  <rdfs:subClassOf rdf:resource="#automatica"/>
</owl:Class>
<owl:Class rdf:ID="manual">
  <rdfs:subClassOf rdf:resource="#transmision"/>
</owl:Class>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

## Jerarquía sobre los tipos de tracción utilizados en los distintos tipos de automóviles comercializados en México:

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns="http://www.owl-ontologies.com/Traccion.owl#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/Traccion.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="traccion"/>
  <owl:Class rdf:ID="delantera">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="_4x2"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="trasera">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#_4x2"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#_4x2">
    <rdfs:subClassOf rdf:resource="#traccion"/>
  </owl:Class>
  <owl:Class rdf:ID="_4x4">
    <rdfs:subClassOf rdf:resource="#traccion"/>
  </owl:Class>
  <owl:Class rdf:ID="integral-autoajustable">
    <rdfs:subClassOf rdf:resource="#traccion"/>
  </owl:Class>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

## Jerarquía sobre los tipos de Interiores:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/Interior.owl#"
  xml:base="http://www.owl-ontologies.com/Interior.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="piel_tradicional">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="piel"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="estandar">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="interior"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="tela">
    <rdfs:subClassOf rdf:resource="#estandar"/>
  </owl:Class>
  <owl:Class rdf:ID="tacto_piel">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#piel"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#piel">
    <rdfs:subClassOf rdf:resource="#interior"/>
  </owl:Class>
  <owl:Class rdf:ID="poliuretano">
    <rdfs:subClassOf rdf:resource="#estandar"/>
  </owl:Class>
  <owl:Class rdf:ID="cuero">
    <rdfs:subClassOf rdf:resource="#piel"/>
  </owl:Class>
</rdf:RDF>
```

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

## Jerarquía sobre los Niveles de Equipamiento en los que se pueden clasificar los automóviles:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/Automoviles.owl#"
  xml:base="http://www.owl-ontologies.com/Automoviles.owl">
```

```

xmlns="http://www.owl-ontologies.com/Equipamiento.owl#"
xmlns:swrl="http://www.w3.org/2003/11/swrl#"
xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xml:base="http://www.owl-ontologies.com/Equipamiento.owl">
<owl:Ontology rdf:about=""/>
<owl:Class rdf:ID="equipamiento"/>
<owl:Class rdf:ID="austero">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:FunctionalProperty rdf:ID="posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >1</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#equipamiento"/>
</owl:Class>
<owl:Class rdf:ID="premium">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:FunctionalProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >4</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#equipamiento"/>
</owl:Class>
<owl:Class rdf:ID="lujo">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >3</owl:hasValue>
      <owl:onProperty>
        <owl:FunctionalProperty rdf:about="#posicion"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#equipamiento"/>
</owl:Class>
<owl:Class rdf:ID="normal">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:FunctionalProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#equipamiento"/>
</owl:Class>

```

```

<owl:FunctionalProperty rdf:about="#posicion">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
</owl:FunctionalProperty>
</rdf:RDF>

```

## Jerarquía sobre los tipos de dirección utilizadas por los automóviles que se comercializan en México:

```

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns="http://www.owl-ontologies.com/Direccion.owl#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/Direccion.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="mecanica">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="direccion"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="electrica">
    <rdfs:subClassOf rdf:resource="#direccion"/>
  </owl:Class>
  <owl:Class rdf:ID="hidraulica">
    <rdfs:subClassOf rdf:resource="#direccion"/>
  </owl:Class>
  <owl:Class rdf:ID="electro-hidraulica">
    <rdfs:subClassOf rdf:resource="#direccion"/>
  </owl:Class>
</rdf:RDF>

```

```

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

## Jerarquía sobre las diferentes configuraciones de los motores incluidos en los vehículos de los automóviles:

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"

```

```

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns="http://www.owl-ontologies.com/ConfiguracionDelMotor.owl#"
xml:base="http://www.owl-ontologies.com/ConfiguracionDelMotor.owl">
<owl:Ontology rdf:about=""/>
<owl:Class rdf:ID="configuracion_del_motor"/>
<owl:Class rdf:ID="otro">
  <rdfs:subClassOf rdf:resource="#configuracion_del_motor"/>
</owl:Class>
<owl:Class rdf:ID="horizontalmente_opuesto-Boxer">
  <rdfs:subClassOf rdf:resource="#configuracion_del_motor"/>
</owl:Class>
<owl:Class rdf:ID="linea">
  <rdfs:subClassOf rdf:resource="#configuracion_del_motor"/>
</owl:Class>
<owl:Class rdf:ID="v">
  <rdfs:subClassOf rdf:resource="#configuracion_del_motor"/>
</owl:Class>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

## Jerarquía sobre Tipos de Aspiración de los motores:

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns="http://www.owl-ontologies.com/AspiracionDelMotor.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/AspiracionDelMotor.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="sobre_cargada">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="aspiracion_del_motor"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="turbo_cargada">
    <rdfs:subClassOf rdf:resource="#aspiracion_del_motor"/>
  </owl:Class>
  <owl:Class rdf:ID="atmosferica-normal">
    <rdfs:subClassOf rdf:resource="#aspiracion_del_motor"/>
  </owl:Class>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) http://protege.stanford.edu -->

```

## Jerarquía sobre la Cilindrada de los Motores:

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/CilindradaDelMotor.owl#"
  xml:base="http://www.owl-ontologies.com/CilindradaDelMotor.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="_6_cilindros">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:FunctionalProperty rdf:ID="posicion"/>
        </owl:onProperty>
        <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int">3</owl:hasValue>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="cilindrada_del_motor"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="_4_cilindros">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</owl:hasValue>
        <owl:onProperty>
          <owl:FunctionalProperty rdf:about="#posicion"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#cilindrada_del_motor"/>
  </owl:Class>
  <owl:Class rdf:ID="_8_cilindros">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int">4</owl:hasValue>
        <owl:onProperty>
          <owl:FunctionalProperty rdf:about="#posicion"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf rdf:resource="#cilindrada_del_motor"/>
  </owl:Class>
  <owl:Class rdf:ID="_10_cilindros">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>

```

```

    <owl:FunctionalProperty rdf:about="#posicion"/>
  </owl:onProperty>
  <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
    >5</owl:hasValue>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf rdf:resource="#cilindrada_del_motor"/>
</owl:Class>
<owl:Class rdf:ID="_5_cilindros">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:FunctionalProperty rdf:about="#posicion"/>
      </owl:onProperty>
      <owl:hasValue rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >2</owl:hasValue>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#cilindrada_del_motor"/>
</owl:Class>
<owl:FunctionalProperty rdf:about="#posicion">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:FunctionalProperty>
</rdf:RDF>

```

<!-- Created with Protege (with OWL Plugin 3.4.4, Build 579) <http://protege.stanford.edu> -->

## ANEXO B.

# Construcción de la base de datos

En este anexo se muestran los archivos OWL de las jerarquías utilizadas para estructurar la información sobre los descriptores de los automóviles utilizados por el sistema de búsqueda y recuperación de información. Dichas jerarquías fueron implementadas en *Protege 3.4*, facilitando su interoperabilidad y difusión.

```
// Crear la base de datos:
```

```
CREATE DATABASE tesis WITH OWNER = postgres ENCODING = 'UTF8' LC_COLLATE = 'Spanish_Mexico.1252' LC_CTYPE = 'Spanish_Mexico.1252' CONNECTION LIMIT = -1;
```

```
CREATE TABLE vendedores (  
  id serial NOT NULL,  
  vendedor character varying(500),  
  correo character varying(100),  
  telefono character varying(20),  
  fax character varying(20),  
  web character varying(50),  
  calle character varying(200),  
  numero character varying(10),  
  colonia character varying(200),  
  cp character varying(6),  
  municipio character varying(100),  
  estado character varying(100),  
  coordenadas geometry,  
  CONSTRAINT vendedores_pkey PRIMARY KEY (id)  
)  
WITH (OIDS=FALSE );  
ALTER TABLE vendedores OWNER TO postgres;
```

```
CREATE TABLE tiposdeproductos (  
  id serial NOT NULL,  
  tipo character varying(100) NOT NULL,
```

```

    CONSTRAINT pk_tiposdeproductos PRIMARY KEY (id)
)
WITH (OIDS=FALSE);
ALTER TABLE tiposdeproductos OWNER TO postgres;

CREATE TABLE productos (
    id serial NOT NULL,
    "idVendedor" integer NOT NULL,
    "idTipoProducto" integer NOT NULL,
    producto character varying(200) NOT NULL,
    descripcion character varying(600),
    precio double precision NOT NULL DEFAULT 0,
    marca character varying(200),
    CONSTRAINT pk_productos PRIMARY KEY (id),
    CONSTRAINT fk_productos_tiposdeproductos FOREIGN KEY ("idTipoProducto")
        REFERENCES tiposdeproductos (id) MATCH SIMPLE
        ON UPDATE NO ACTION ON DELETE NO ACTION,
    CONSTRAINT fk_productos_vendedores FOREIGN KEY ("idVendedor")
        REFERENCES vendedores (id) MATCH SIMPLE
        ON UPDATE NO ACTION ON DELETE NO ACTION
)
WITH (OIDS=FALSE);
ALTER TABLE productos OWNER TO postgres;

CREATE TABLE descriptores (
    id serial NOT NULL,
    "idTipoProducto" integer NOT NULL,
    descriptor character varying(100) NOT NULL,
    cualitativo boolean NOT NULL,
    jerarquia character varying(100),
    nodoraiz character varying(100),
    CONSTRAINT pk_descriptores PRIMARY KEY (id),
    CONSTRAINT fk_descriptores_tiposdeproductos FOREIGN KEY ("idTipoProducto")
        REFERENCES tiposdeproductos (id) MATCH SIMPLE
        ON UPDATE NO ACTION ON DELETE NO ACTION
)
WITH (OIDS=FALSE);
ALTER TABLE descriptores OWNER TO postgres;

CREATE TABLE detallesproductos (
    "idProducto" integer NOT NULL,
    "idDescriptor" integer NOT NULL,
    valor double precision,
    nodo character varying(100) DEFAULT NULL::character varying,
    CONSTRAINT pk_detalleproductos PRIMARY KEY ("idProducto", "idDescriptor"),
    CONSTRAINT fk_detalleproductos_descriptores FOREIGN KEY ("idDescriptor")
        REFERENCES descriptores (id) MATCH SIMPLE
        ON UPDATE NO ACTION ON DELETE NO ACTION,
    CONSTRAINT fk_detalleproductos_productos FOREIGN KEY ("idProducto")
        REFERENCES productos (id) MATCH SIMPLE
        ON UPDATE NO ACTION ON DELETE NO ACTION
)
WITH (OIDS=FALSE);
ALTER TABLE detallesproductos OWNER TO postgres;

```

## ANEXO C.

# Poblando la base de datos

Para poblar la base de datos y probar el modelo de recuperación, se realizó una recolección de aproximadamente 120 automóviles de las principales compañías de automóviles en México y de las agencias de autos ubicadas en la ciudad de Culiacán, Sinaloa. Dichos datos pueden ser agregados a la base de datos ejecutando el siguiente script de SQL:

```

/*****
* Vendedores *
*****/

```

```

INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (1,'Ford Zapata', null, '(667)713-31-13', null, 'www.ford.com.mx', 'Blvd. Emiliano Zapata', '50 Pte.', 'Col. Guadalupe',
'80220',
      'Culiacán','Sinaloa', GeomFromText('POINT(-11955142.364743 2850852.13713402)'));

```

```

INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (2,'Ford 3 Ríos', null, '(667)758-88-99', null, 'www.ford.com.mx', 'Blvd. Pedro Infante', '2598 Pte.', 'Fracc. Desarrollo
Urbano Tres Ríos', '80020',
      'Culiacán','Sinaloa', GeomFromText('POINT(-11957783.6223685 2851111.60501516)'));

```

```

INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (3,'HONDA VISION CULIACAN', null, '(667)713-15-55', null, 'www.honda.com.mx', 'Blvd. Enrique Sanchez Alonso',
'1863 Nte.', 'Fracc. Desarrollo Urbano Tres Ríos', '80220',
      'Culiacán','Sinaloa', GeomFromText('POINT(-11955814.1745618 2853720.91435145)'));

```

```

INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (4,'PREMIER CHEVROLET S.A. DE C.V.', null, '(667)758-58-58', null, 'www.chevrolet.com.mx', 'Blvd. Pedro Infante',
'471', 'Fracc. Desarrollo Urbano Tres Ríos', '80020',
      'Culiacán','Sinaloa', GeomFromText('POINT(-11957270.359667 2851401.82685689)'));

```

```

INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)

```

```
VALUES (5,'AUTOMOTRIZ CULIACAN, S.A. DE C.V.', null, '(667)761-35-30', null, 'www.chevrolet.com.mx', 'Blvd. Emiliano Zapata', '4630 Pte.', 'Col. San Rafael', '80150',
'Culiacán','Sinaloa', GeomFromText('POINT(-11958890.7649499 2849805.30814522)'));
```

```
INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (6,'CULIACAN MOTORS,S.A. DE C.V.', null, '(667)758-00-00', null, 'www.chevrolet.com.mx', 'Blvd. Emiliano Zapata', '1999', 'Col. Vallado', '80110',
'Culiacán','Sinaloa', GeomFromText('POINT(-11957523.8559052 2850314.09211468)'));
```

```
INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (7,'PLASSA NISSAN', null, '(667)716-46-26', null, 'www.nissan.com.mx', 'Blvd. Emiliano Zapata y Esq. R. Paliza', 'S/N', 'Col. Guadalupe', '80220',
'Culiacán','Sinaloa', GeomFromText('POINT(-11954885.5841012 2850878.71094463)'));
```

```
INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (8,'NISSAN AEROPUERTO', null, '(667)717-17-17', null, 'www.nissan.com.mx', 'Calz. Aeropuerto', '5400-A', 'Col. San Rafael', '80150',
'Culiacán','Sinaloa', GeomFromText('POINT(-11960543.7156863 2848450.93955049)'));
```

```
INSERT INTO vendedores(id, vendedor, correo, telefono, fax, web, calle, numero, colonia, cp, municipio, estado, coordenadas)
VALUES (9,'TOYOTA CULIACÁN', null, '(667)717-20-20', null, 'www.toyota.com.mx', 'Blvd. Pedro Infante', '2531 Pte.', 'Fracc. Desarrollo Urbano Tres Ríos', '80000',
'Culiacán','Sinaloa', GeomFromText('POINT(-11957391.5840165 2851261.49325029)'));
```

```
/*****
```

```
* Tipos de productos *
```

```
*****/
```

```
INSERT INTO tiposdeproductos VALUES (1,'VEHICULOS AUTOMOTRICES');
```

```
/*****
```

```
* Descriptores *
```

```
*****/
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(1,1,'Clasificación',true,'OntologyCar.OWL','automovil');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(2,1,'Equipamiento',true,'equipamiento.owl','equipamiento');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(3,1,'Interiores',true,'interior.owl','interior');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(4,1,'Transmisión',true,'transmision.owl','transmision');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(5,1,'Dirección',true,'direccion.owl','direccion');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(6,1,'Tracción',true,'traccion.owl','traccion');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(7,1,'Cilindraje',true,'cilindradaelmotor.owl','cilindraje_del_motor');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(8,1,'Combustible',true,'combustible.owl','combustible');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(9,1,'Aspiración',true,'aspiraciondelmotor.owl','aspiracion_del_motor');
```

```
INSERT INTO descriptores (id,"idTipoProducto",descriptor,cualitativo,jerarquia,nodoraiz)
VALUES(10,1,'Configuración',true,'configuracion_del_motor.owl','configuracion_del_motor');
```

```
/******
```

```
* PRODUCTOS *
```

```
*****/
```

```
/******
```

```
* Vehiculos Ford *
```

```
*****/
```

```
/* Fiesta */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Fiesta','Fiesta Trend Conford TA',194633.00);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Fiesta','Fiesta Trend Conford TA',194633.00);
```

```
/* Focus */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Focus','Focus Sport TA', 262443);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Focus','Focus Sport TA', 262443);
```

```
/* Focus RS */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Focus RS','Focus RS', 470001);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Focus RS','Focus RS', 470001);
```

```
/* Focus 5 puertas */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Focus 5 puertas','Focus 5 Puertas', 262443);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Focus 5 puertas','Focus 5 Puertas', 262443);
```

```
/* Focus ST */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Focus ST','Focus ST', 349199);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Focus ST','Focus ST', 349199);
```

```
/* Mustang */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Mustang','Mustang GT Premium Convertible TA/TS', 512599);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Mustang','Mustang GT Premium Convertible TA/TS', 512599);
```

```
/* Mustang Shelby */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Mustang Shelby','Mustang Shelby GT 500', 777000);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Mustang Shelby','Mustang Shelby GT 500', 777000);
```

```
/* Fusion */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Fusion','Fusion SEL V6 Ford Interactive System', 375928);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Fusion','Fusion SEL V6 Ford Interactive System', 375928);
```

```
/* Fusion Sport */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Fusion Sport','Fusion Sport V6', 375928);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Fusion Sport','Fusion Sport V6', 375928);
```

```
/* EcoSport */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'EcoSport','EcoSport 4x2 TA', 263953);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'EcoSport','EcoSport 4x2 TA', 263953);
```

```
/* Escape */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Escape','Escape Limited V6', 377427);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Escape','Escape Limited V6', 377427);
```

```
/* Explorer */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Explorer','Explorer Limited V8', 573833);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Explorer','Explorer Limited V8', 573833);
```

```
/* Explorer Sport Track */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Explorer Sport Track','Explorer Sport Track V8 4x2', 430680);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Explorer Sport Track','Explorer Sport Track V8 4x2', 430680);
```

```
/* Expedition */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Expedition','Expedition Limited 4x2', 605092);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Expedition','Expedition Limited 4x2', 605092);
```

```
/* Expedition Max */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Expedition Max','Expedition Max 4x2', 636701);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
```

```
VALUES (2, 1, 'Expedition Max','Expedition Max 4x2', 636701);
```

```
/* Expedition King Ranch */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Expedition King Ranch','Expedition King Ranch 4x2', 623991);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Expedition King Ranch','Expedition King Ranch 4x2', 623991);
```

```
/* Edge */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Edge','Edge Limited DVD', 477812);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Edge','Edge Limited DVD', 477812);
```

```
/* Edge Sport */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Edge Sport','Edge Sport', 518612);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Edge Sport','Edge Sport', 518612);
```

```
/* Courier */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Courier','Courier XL', 166900);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Courier','Courier XL', 166900);
```

```
/* Ranger */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Ranger','Ranger XLT Cab. doble 4 puertas AA', 256300);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Ranger','Ranger XLT Cab. doble 4 puertas AA', 256300);
```

```
/* F-150 */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'F-150','F-150 XL REG CAB 4x4', 317200);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'F-150','F-150 XL REG CAB 4x4', 317200);
```

```
/* F-250 SUPER DUTY */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'F-250 SUPER DUTY','F-250 SUPER DUTY Crew Cab Diesel 4x4 TA', 558800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'F-250 SUPER DUTY','F-250 SUPER DUTY Crew Cab Diesel 4x4 TA', 558800);
```

```
/* Lobo */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'F-250 SUPER DUTY','Lobo Crew Cab 4x4 Lariat 5.4L',591714);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'F-250 SUPER DUTY','Lobo Crew Cab 4x4 Lariat 5.4L',591714);
```

```

/* Lobo King Ranch */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Lobo King Ranch','Lobo KR cab. doble 4x4',606300);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Lobo King Ranch','Lobo KR cab. doble 4x4',606300);

/* E-150/350 */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'E-150/350','Econoline E350 XL V8 15 pasajeros',562999);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'E-150/350','Econoline E350 XL V8 15 pasajeros',562999);

/* Transit */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (1, 1, 'Transit','PASAJ LARGA DIESEL A/A',371499);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (2, 1, 'Transit','PASAJ LARGA DIESEL A/A',371499);

/*****
* Vehiculos Chevrolet *
*****/
/* Chevy 3 puertas */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Chevy 3 Puertas','Chevy 3 Puertas Paq. J',145900);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Chevy 3 Puertas','Chevy 3 Puertas Paq. J',145900);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Chevy 3 Puertas','Chevy 3 Puertas Paq. J',145900);

/* Chevy Sedán */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Chevy Sedán','Chevy Sedán Paquete J',157620);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Chevy Sedán','Chevy Sedán Paquete J',157620);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Chevy Sedán','Chevy Sedán Paquete J',157620);

/* Tornado */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Tornado','Tornado Pickup Paq. C',173200);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Tornado','Tornado Pickup Paq. C',173200);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Tornado','Tornado Pickup Paq. C',173200);

/* Aveo */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Aveo','Aveo Paq. F',200210);

```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Aveo','Aveo Paq. F',200210);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Aveo','Aveo Paq. F',200210);
```

```
/* Optra */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Optra','Optra Paquete D',199800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Optra','Optra Paquete D',199800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Optra','Optra Paquete D',199800);
```

```
/* Cruze */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Cruze','Cruze Paquete F',266250);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Cruze','Cruze Paquete F',266250);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Cruze','Cruze Paquete F',266250);
```

```
/* Silverado 1500 */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Silverado 1500','Silverado 1500 Paquete G Cabina Regular',256060);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Silverado 1500','Silverado 1500 Paquete G Cabina Regular',256060);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Silverado 1500','Silverado 1500 Paquete G Cabina Regular',256060);
```

```
/* Silverado 2500 */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Silverado 2500','Silverado 2500 Paquete A Cabina Extendida 4x2',316010);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Silverado 2500','Silverado 2500 Paquete A Cabina Extendida 4x2',316010);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Silverado 2500','Silverado 2500 Paquete A Cabina Extendida 4x2',316010);
```

```
/* Malibu */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Malibu','Malibu Paquete G',348980);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Malibu','Malibu Paquete G',348980);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Malibu','Malibu Paquete G',348980);
```

```
/* Colorado */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
```

```
VALUES (4, 1, 'Colorado','Colorado Paq. B',349560);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Colorado','Colorado Paq. B',349560);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Colorado','Colorado Paq. B',349560);

/* Captiva */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Captiva','Captiva Sport Paq. D',364450);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Captiva','Captiva Sport Paq. D',364450);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Captiva','Captiva Sport Paq. D',364450);

/* Cheyenne */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Cheyenne','Cheyenne 4x4 Crew Cab LTZ Paq. C',545360);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Cheyenne','Cheyenne 4x4 Crew Cab LTZ Paq. C',545360);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Cheyenne','Cheyenne 4x4 Crew Cab LTZ Paq. C',545360);

/* Express */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Express','Express Passenger Van Paq. C',517380);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Express','Express Passenger Van Paq. C',517380);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Express','Express Passenger Van Paq. C',517380);

/* Camaro */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Camaro','Camaro Paquete C',523390);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Camaro','Camaro Paquete C',523390);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Camaro','Camaro Paquete C',523390);

/* Traverse */
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Traverse','Traverse Paq. B',498300);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Traverse','Traverse Paq. B',498300);

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Traverse','Traverse Paq. B',498300);
```

/\* Avalanche \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Avalanche','Avalanche Paq. C Z71',584300);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Avalanche','Avalanche Paq. C Z71',584300);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Avalanche','Avalanche Paq. C Z71',584300);
```

/\* Tahoe \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Tahoe','Tahoe SUV Paq. E EZ71',623450);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Tahoe','Tahoe SUV Paq. E EZ71',623450);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Tahoe','Tahoe SUV Paq. E EZ71',623450);
```

/\* Suburban \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Suburban','Suburban SUV Paq. G 2500 (3/4 ton)',762390);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Suburban','Suburban SUV Paq. G 2500 (3/4 ton)',762390);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Suburban','Suburban SUV Paq. G 2500 (3/4 ton)',762390);
```

/\* Corvette \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (4, 1, 'Corvette','Corvette 2 ptas. Coupe Paq. A Automático',921700);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (5, 1, 'Corvette','Corvette 2 ptas. Coupe Paq. A Automático',921700);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (6, 1, 'Corvette','Corvette 2 ptas. Coupe Paq. A Automático',921700);
```

/\*\*\*\*\*\*

\* Vehiculos Nissan \*

\*\*\*\*\*/

/\* Tsuru \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Tsuru','TSURU GS II EQUIPADO/TRANSMISIÓN AUTOMÁTICA',144800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Tsuru','TSURU GS II EQUIPADO/TRANSMISIÓN AUTOMÁTICA',144800);
```

/\* Tiida Sedán \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Tiida Sedán','TIIDA SEDAN EMOTION TRANSMISIÓN AUTOMÁTICA/1.8 LTS',193500);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Tiida Sedán','TIIDA SEDAN EMOTION TRANSMISIÓN AUTOMÁTICA/1.8 LTS',193500);
```

/\* Tiida HatchBack \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Tiida HB','TIIDA HB PREMIUM TRANSMISIÓN AUTOMÁTICA/1.8 LTS',226300);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Tiida HB','TIIDA HB PREMIUM TRANSMISIÓN AUTOMÁTICA/1.8 LTS',226300);
```

```
/* Sentra SE-R */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Sentra SE-R','SENTRA SE-R SPEC V',255100);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Sentra SE-R','SENTRA SE-R SPEC V',255100);
```

```
/* Sentra */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Sentra','SENTRA LUXURY PIEL CVT',252100);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Sentra','SENTRA LUXURY PIEL CVT',252100);
```

```
/* Platina */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Platina','PLATINA CUSTOM TRANSMISIÓN MANUAL/AIRE ACONDICIONADO',129600);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Platina','PLATINA CUSTOM TRANSMISIÓN MANUAL/AIRE ACONDICIONADO',129600);
```

```
/* Maxima */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Maxima','MAXIMA SPORT TRANSMISIÓN CVT/3.5L',483600);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Maxima','MAXIMA SPORT TRANSMISIÓN CVT/3.5L',483600);
```

```
/* Aprio */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Aprío','APRIO BASE TRANSMISIÓN AUTOMÁTICA/AIRE ACONDICIONADO',144300);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Aprío','APRIO BASE TRANSMISIÓN AUTOMÁTICA/AIRE ACONDICIONADO',144300);
```

```
/* Altima */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Altima','ALTIMA SR 3.5 LTS./CVT CON MODO MANUAL',374200);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Altima','ALTIMA SR 3.5 LTS./CVT CON MODO MANUAL',374200);
```

```
/* Altima Coupé */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Altima Coupé','ALTIMA COUPE SR 3.5 LTS./TRANSMISIÓN CVT',395100);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Altima Coupé','ALTIMA COUPE SR 3.5 LTS./TRANSMISIÓN CVT',395100);
```

```
/* 370Z TOURING */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, '370Z TOURING','370Z TOURING TRANSMISIÓN AUTOMÁTICA',649600);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, '370Z TOURING','370Z TOURING TRANSMISIÓN AUTOMÁTICA',649600);
```

```
/* Murano */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'MURANO','MURANO LE AWD/TRANSMISIÓN CVT', 497000);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'MURANO','MURANO LE AWD/TRANSMISIÓN CVT', 497000);
```

```
/* Rogue */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Rogue','ROGUE SL AWD TRANSMISIÓN CVT/2.5 LTS', 338200);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Rogue','ROGUE SL AWD TRANSMISIÓN CVT/2.5 LTS', 338200);
```

```
/* X-Trail */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'X-Trail','X-TRAIL 4WD TRANSMISIÓN CVT/2.5 LTS', 355600);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'X-Trail','X-TRAIL 4WD TRANSMISIÓN CVT/2.5 LTS', 355600);
```

```
/* Titan */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'Titan','TITAN CREW CAB LE 4X4 PIEL/TRANSMISIÓN AUTOMÁTICA/QUEMACOCOS', 479800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'Titan','TITAN CREW CAB LE 4X4 PIEL/TRANSMISIÓN AUTOMÁTICA/QUEMACOCOS', 479800);
```

```
/* PATHFINDER */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'PATHFINDER','PATHFINDER LE 4X4 TRANSMISIÓN AUTOMÁTICA/LUXURY CON MODO MANUAL', 534800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'PATHFINDER','PATHFINDER LE 4X4 TRANSMISIÓN AUTOMÁTICA/LUXURY CON MODO MANUAL', 534800);
```

```
/* FRONTIER */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'FRONTIER','FRONTIER SE V6 4X4', 389800);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'FRONTIER','FRONTIER SE V6 4X4', 389800);
```

```
/* ARMADA */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (7, 1, 'ARMADA','ARMADA SE 4x4 PIEL/QUEMACOCOS/TRANSMISIÓN AUTOMÁTICA', 648100);
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (8, 1, 'ARMADA','ARMADA SE 4x4 PIEL/QUEMACOCOS/TRANSMISIÓN AUTOMÁTICA', 648100);
```

```
/******
```

```
* Vehiculos Honda *
```

```
*****
```

```
/* Fit */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
```

VALUES (3, 1, 'Fit','Fit EX AUTOMATICA',219000);

/\* City \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'City','City EX AUTOMÁTICA',240000);

/\* Civic Sedán \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Civic Sedán','Civic Sedán EXL AUTOMÁTICA',282000);

/\* Civic Coupé \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Civic Coupé','Civic Coupé EXL AUTOMÁTICA',281500);

/\* Civic Hibrido \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Civic Hybrid','Civic Hybrid CVT',329000);

/\* Civic Si \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Civic Si','Civic Si COUPÉ Manual',343000);

/\* Accord Sedán \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Accord Sedán','Accord Sedán EX V6',377000);

/\* Accord Coupé \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Accord Coupé','Accord Coupé EX V6',377000);

/\* Accord Crosstour \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'Accord Crosstour','Accord Crosstour EXL V6',468000);

/\* CR-V \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'CR-V ','CR-V EXL V6',358000);

/\* ODYSSEY \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'ODYSSEY','ODYSSEY TOURING',536000);

/\* PILOT \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'PILOT','PILOT EXL',505000);

/\* RIDGELINE \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (3, 1, 'RIDGELINE','RIDGELINE RTL',458000);

/\*\*\*\*\*\*

\* Vehiculos Toyota \*

\*\*\*\*\*/

/\* Avanza \*/

INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)  
VALUES (9, 1, 'Avanza','Avanza Premium AT',194300);

/\* Yaris Sedán \*/

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Yaris Sedán','Yaris SD Premium AT', 209900);
```

```
/* Yaris HatchBack */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Yaris HatchBack','Yaris HB Premium AT', 193500);
```

```
/* Matrix */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Matrix','Matrix XR 5AT', 298400);
```

```
/* Corolla */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Corolla','Corolla XRS 5MT w/moonroof', 264700);
```

```
/* Camry */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Camry','Camry XLE V6', 355700);
```

```
/* Sequoia */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Sequoia','Sequoia Platinum', 718900);
```

```
/* FJ Cruiser */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'FJ Cruiser','FJ Cruiser 4x4 Premium', 440300);
```

```
/* Highlander */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Highlander','Highlander Limited', 520000);
```

```
/* Sienna */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Sienna','Sienna Limited', 642200);
```

```
/* Rav4 */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Rav4','Rav4x2 V6 SportPiel', 642200);
```

```
/* LandCruiser */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'LandCruiser','LandCruiser Wagon VX', 970200);
```

```
/* Tundra */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Tundra','Tundra 5.7L 4x4 LTD Crew Max', 549900);
```

```
/* Tacoma */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Tacoma','Tacoma TRD Sport 4x4', 402000);
```

```
/* Hilux */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Hilux','Hilux 4x2 Doble Cabina SR', 279500);
```

```
/* Hiace */
```

```
INSERT INTO productos("idVendedor", "idTipoProducto", producto, descripcion, precio)
VALUES (9, 1, 'Hiace','Hiace 15 Pasajeros AA', 367600);
```

## ANEXO D.

# Códigos Fuentes

En este anexo, se muestran los códigos fuentes del sistema de búsqueda y recuperación de información implementado, iniciando por los diferentes paquetes utilizados por el sistema y las clases correspondientes a dichos paquetes.

Clase “*Jerarquia.java*” del paquete *confusion*:

```
/******  
* Package *  
*****/  
package confusion;  
  
/******  
* Imports *  
*****/  
import com.hp.hpl.jena.ontology.*;  
import com.hp.hpl.jena.rdf.model.*;  
  
import java.io.*;  
import java.util.*;  
  
public class Jerarquia {  
  
    private OntModel m;  
    private OntClass root;  
  
    // Constructor de la jerarquia  
    public Jerarquia(String filename,String rootelement) {  
        m = ModelFactory.createOntologyModel( OntModelSpec.OWL_MEM, null );  
        m.getDocumentManager().addAltEntry( "http://www.owl-ontologies.com/Ontology.owl", filename);  
        m.read( filename );  
        root = getNode(rootelement);  
    }  
}
```

```

// Obtener el nodo raiz
public OntClass getRoot() {
    return root;
}

// Obtener el modelo de la ontología
public OntModel getModel() {
    return m;
}

// Obtener un nodo a partir de su nombre
public OntClass getNode(String node) {
    Iterator<OntClass> i = m.listClasses();
    OntClass res=null, temp = null ;
    while (i.hasNext() && res==null) {
        if ((temp = i.next()).getLocalName().compareToIgnoreCase(node)==0 )
            res=temp;
    }
    return res;
}

// Obtener el padre de un nodo
public OntClass getParentOf(OntClass node) {
    OntClass res = null;
    Iterator<OntClass> i = node.listSuperClasses();
    while (i.hasNext()) {
        res = i.next();
        if(res.getLocalName()==null)
            continue;
        else
            break;
    }
    return res;
}

// Obtener los padres de un nodo hasta llegar al nodo raiz
public ArrayList<OntClass> getParentsOf(OntClass node) {
    ArrayList<OntClass> parents = new ArrayList();
    OntClass temp = node;
    while (!temp.equals(this.root)) {
        temp = this.getParentOf(temp);
        parents.add(temp);
    }
    return parents;
}

// Función que determina si r es hijo de s
public boolean isChildrenOf_S(OntClass s, OntClass r) {
    boolean flag = false;
    Iterator<OntClass> i = s.listSubClasses();
    OntClass temp;
    while (i.hasNext() && !flag) {
        temp = i.next();
        if (r.equals(temp))
            flag = true;
        else
            flag = this.isChildrenOf_S(temp, r);
    }
}

```

```

    return flag;
}

// Función que determina la profundidad de un nodo, donde: la profundidad del nodo raíz es igual a 1
public int getDepth(OntClass nodo) {
    int res=-1;
    res = this.getParentsOf(nodo).size()+1;
    return res;
}

// Confusión como una expresión de similitud  $Sim(r,s) = 1 - conf(r,s)/p(r)$ 
public float confusion(OntClass r, OntClass s, Jerarquia jerarquia) {
    float conf;
    conf = this.conf(r, s, jerarquia);
    if(conf!=-1) {
        conf /= this.getDepth(r);
        conf = 1 - conf;
    }
    return conf;
}

// Cálculo de confusión
public float conf(OntClass r, OntClass s, Jerarquia jerarquia) {
    float conf = -1;
    if(r.equals(s)) { // r = s
        conf = 0;
    } else if(this.isChildrenOf_S(s, r)) { // r = HijoDe(s)
        conf = 0;
    } else if(this.isChildrenOf_S(r, s)) { // s = HijoDe(r)
        ArrayList<OntClass> parents = jerarquia.getParentsOf(s);
        conf=0;
        boolean flag = true;
        for(int i=0;i<parents.size()&&flag;i++) {
            conf++;
            if(r.equals(parents.get(i)))
                flag=false;
        }
    } else if(jerarquia.getParentOf(r).equals(jerarquia.getParentOf(s))) { // Son hermanos
        int oR = jerarquia.getOrden(r);
        int oS = jerarquia.getOrden(s);
        if(oR != -1 && oS != -1) { // Jerarquía ordenada
            conf = Math.abs(oR-oS);
            OntClass parent = jerarquia.getParentOf(r);
            ArrayList <OntClass> childrens = jerarquia.getChildrenOf(parent);
            conf /= childrens.size();
        } else { // Jerarquía simple
            conf = 1;
        }
    } else { // Recursion
        conf = 1 + this.conf(r, this.getParentOf(s), jerarquia);
    }
    return conf;
}

// Obtener los hijos inmediatos de un nodo
public ArrayList<OntClass> getChildrensOf(OntClass nodo) {

```

```

    ArrayList<OntClass> childrens = new ArrayList();
    Iterator<OntClass> ei = node.listSubClasses();
    while (ei.hasNext())
        childrens.add(ei.next());
    return childrens;
}

// Cálculo de confusión conforme la Teoría de Confusión
public ArrayList<OntClass> getAllChildrensOf(OntClass node) {
    ArrayList<OntClass> childrens = this.getChildrensOf(node);
    ArrayList<OntClass> resultados= new ArrayList();
    Iterator<OntClass> i = childrens.iterator();
    resultados.addAll(childrens);
    while(i.hasNext())
        resultados.addAll(this.getAllChildrensOf(i.next()));
    return resultados;
}

// Cálculo de confusión
public ArrayList<OntClass> getHojas(OntClass root) {
    ArrayList<OntClass> resultados = new ArrayList();
    ArrayList<OntClass> childrens = this.getAllChildrensOf(root);
    Iterator<OntClass> iterator = childrens.iterator();
    OntClass temp;
    while(iterator.hasNext()) {
        temp = iterator.next();
        if(this.getChildrensOf(temp).size()==0)
            resultados.add(temp);
    }
    return resultados;
}

// Obtener el valor de la restricción de orden para el cálculo automático de confusión en Jerarquías Ordenadas
public int getOrden(OntClass nodo) {
    int res = -1;
    StmtIterator iterador =nodo.listProperties();
    Statement stmt;
    while (iterador.hasNext()) {
        stmt= iterador.nextStatement();
        Resource r = stmt.getResource();
        Property p = stmt.getPredicate();
        if(r.getLocalName()==null && p.getLocalName().compareTo("subClassOf")==0) {
            StmtIterator i2 =r.listProperties();
            Statement s;
            while (i2.hasNext()) {
                s = i2.nextStatement();
                if(s.asTriple().getPredicate().getLocalName().compareTo("hasValue")!=0)
                    continue;
                res = Integer.parseInt(s.asTriple().getObject().getLiteral().getValue().toString());
            }
        } else
            continue;
    }
    return res;
}
}

```

## Clase “*PostgresSQLWrapper.java*” del paquete *postgresdb*, utilizada para la comunicación entre Java y PostgreSQL:

```

package postgresdb;
import java.sql.*;

/**
 * @author Nahun Enrique
 */
public class PostgresSQLWrapper {
    // Establecer una conexión con PostgreSQL
    public Connection connect() {
        Connection conn = null;
        try {
            Class.forName("org.postgresql.Driver");
            String url = "jdbc:postgresql://localhost:5432/tesis"; // <host>:<puerto>/<base de datos>
            String username = new String("postgres"); // usuario
            String password = new String("nosoyyo"); // contraseña

            conn = DriverManager.getConnection(url, username, password);
        } catch (Exception e) { }
        return conn;
    }

    // Cerrar la conexión con PostgreSQL
    public void closeConnection(Connection conn) {
        try { conn.close(); } catch (Exception e) {} finally { conn=null; }
    }

    // Obtener los productos existentes en la base de datos
    public ResultSet obtenerProductos() {
        ResultSet rs = null;
        Connection conn = this.connect();
        try {
            Statement st = conn.createStatement();
            rs = st.executeQuery("SELECT * FROM tiposdeproductos");
        } catch (Exception e) {} finally { this.closeConnection(conn); }
        return rs;
    }

    // Ejecutar una consulta
    public ResultSet getQuery(String sql) {
        ResultSet rs = null;
        Connection conn = this.connect();
        try {
            Statement st = conn.createStatement();
            rs = st.executeQuery(sql);
        } catch (Exception e) {} finally { this.closeConnection(conn); }
        return rs;
    }
}

```

Archivo “*app.jsp*” con la aplicación principal:

```

<%-- Document : app Author : Nahun Enrique --%>

<%@page contentType="text/html" pageEncoding="UTF-8"%>
<%@ page import = "postgressdb.*" %>
<%@ page import = "java.sql.*" %>
<%@ page import = "java.util.*" %>
<%@ page import = "confusion.*" %>
<%@ page import = "com.hp.hpl.jena.ontology.*" %>
<%@ page import = "com.hp.hpl.jena.rdf.model.*" %>
<% Jerarquia jerarquia=new Jerarquia("file:/CatalogoDeAutomoviles/clasificacion.owl","automovil"); %>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<title>GSPS</title>
<script type="text/javascript" src="script.js"></script>
<script type="text/javascript" src="js/range.js"></script>
<script type="text/javascript" src="js/timer.js"></script>
<script type="text/javascript" src="js/slider.js"></script>

<script
src='http://maps.google.com/maps?file=api&v=2&key=ABQIAAAeOetIDyuePZsonNLD0F8jxSvrCEJx5aS2Ic_Kf6e4nE
443Nx9BTUGoomeChi-M5X4d9AZWIJGI7xA'></script>
<script src="http://www.openlayers.org/api/OpenLayers.js" type="text/javascript"></script>
<script defer="defer" type="text/javascript">
var map;
var flag_feature=true;
OpenLayers.IMAGE_RELOAD_ATTEMPTS = 7;
OpenLayers.Util.onImageLoadErrorColor = "transparent";
var position=null;

OpenLayers.DOTS_PER_INCH = 25.4 / 0.28;

function init(){
document.getElementById("rbSencilla").click();
document.getElementById("rbClasificacion").click();
document.getElementById("vehiculo").selectedIndex=0;

format = 'image/png';

var bounds = new OpenLayers.Bounds(-11965987.46529, 2841164,-11943935.38264,2860999.74980);

var options = {
projection: "EPSG:900913",
displayProjection:new OpenLayers.Projection("EPSG:900913"),
units: "m",
maxResolution: 156543.0339,
maxExtent: bounds,
controls:[]
};

map = new OpenLayers.Map('map', options);

var gstreet = new OpenLayers.Layer.Google(
"Mapa base",
{'minZoomLevel':13, 'maxZoomLevel':20, 'numZoomLevels':7, 'sphericalMercator': true}
);

```

```

vectors = new OpenLayers.Layer.Vector("Posición del usuario");

map.addLayers([gstreet, vectors]);

controls = {
  point: new OpenLayers.Control.DrawFeature(vectors,
    OpenLayers.Handler.Point),
  drag: new OpenLayers.Control.DragFeature(vectors)
};

controls.drag.onComplete = function(feature) {
  document.getElementById("x_position").value=feature.geometry.x;
  document.getElementById("y_position").value=feature.geometry.y;
}

controls.point.featureAdded = function(feature) {
  controls.point.deactivate();
  flag_feature = false;
  controls.drag.activate();
  document.getElementById("x_position").value=feature.geometry.x;
  document.getElementById("y_position").value=feature.geometry.y;
}

for(var key in controls) {
  map.addControl(controls[key]);
}

map.addControl(new OpenLayers.Control.PanZoomBar({
  zoomWorldIcon: true,
  position: new OpenLayers.Pixel(2, 15)
}));
map.addControl(new OpenLayers.Control.LayerSwitcher());
map.addControl(new OpenLayers.Control.Navigation());
map.zoomToExtent(bounds);

}

function toggleControlNavegar(element) {
  for(key in controls) {
    var control = controls[key];
    if(element.value == key && element.checked) {
      control.activate();
    } else {
      control.deactivate();
    }
  }
}

function toggleControl(element) {
  for(key in controls) {
    var control = controls[key];
    control.deactivate();
  }
  if(flag_feature)
    controls.point.activate();
  else

```

```

        controls.drag.activate();
    }
</script>
<link type="text/css" rel="StyleSheet" href="css/winclassic.css" />
</head>
<body style="width:100%; height:100%;background-color:#FFFFFF;border-width:0px;margin: 0px;" onload="init();">
    <form action="resultados.jsp" method="POST">
        <input type="text" name="x_position" id="x_position" style="display:none"/>
        <input type="text" name="y_position" id="y_position" style="display:none"/>
        <div id="banner" align="center">
            <table>
                <tr>
                    <td></td>
                    <td><h2 style="color:#0066FF;" align="center">Geo-Semantic Product Search</h2></td>
                    <td><h2 align="center">-</h2></td>
                    <td><h2 style="color:#CC3333;" align="center">GPSP</h2></td>
                    <td></td>
                </tr>
            </table>
        </div>
        <div id="black_banner" style="background-color:black;height:5px;"></div>
        <div id="containt" style="background-color:#FFFFFF;overflow:auto;">
            <br>
            <table align="center" width="90%">
                <thead>
                    <th>Seleccione los criterios para su búsqueda:</th>
                    <th align="left">Seleccione su ubicación sobre el mapa:</th>
                </thead>
                <tbody>
                    <tr>
                        <td width="50%">
                            <div id="criterios" style="width: 100%; height: 450px;">
                                <div style="width: 100%; height: 92%;overflow:auto;">
                                    <table width="100%">
                                        <tbody>
                                            <tr>
                                                <td>
                                                    <fieldset>
                                                        <p>
                                                            <strong>Tipo de vehículo:</strong>
                                                            <select name="vehiculo" id="vehiculo" onChange="cambioVehiculo(this);">
                                                                <option value="0">¿Vehículo?</option>
                                                                <%
                                                                    Iterator<OntClass> iterador =
                                                                    jerarquia.getChildrenOf(jerarquia.getRoot()).listIterator();
                                                                    while(iterador.hasNext()) {
                                                                        String value = iterador.next().getLocalName();
                                                                        String option = value.toUpperCase().substring(0,
1).concat(value.substring(1));
                                                                        out.println("<option value=\"" + value + "\">" + option + "</option>");
                                                                    }
                                                                %>
                                                            </select>
                                                            <strong>Listar por:</strong>
                                                            <label>
                                                                <input name="rbClasificacion" id="rbClasificacion" type="radio"
                                                                onClick="cambioBusqueda(this)" value="clasificacion" checked>
                                                                Clasificación

```

```

</label>
<label>
  <input      name="rbClasificacion"      id="rbMarca"      type="radio"
onClick="cambioBusqueda(this)" value="marca">
  Marca
</label>
</p>
<table width="80%" border="0" bordercolor="#000066" id="tBusqueda"
align="center">
  <tr>
    <th scope="col" id="thSegmento" align="center">Segmento</th>
    <th scope="col" id="thClase" align="center">Clase</th>
    <th scope="col" id="thMarca" align="center" style="display:none"
>Marca</th>
    <th scope="col" align="center">Modelo</th>
  </tr>
  <tr>
    <td id="tdSegmento" align="center">
      <select      name="segmento"      id="segmento"
onchange="cambioSegmento(this)" disabled>
        <option value="0">¿Segmento?</option>
      </select>
    </td>
    <td name="select" id="tdClase" align="center">
      <select id="clase" name="clase" disabled>
        <option value="0">¿Clase?</option>
      </select>
    </td>
    <td id="tdMarca" align="center" style="display:none" >
      <select name="marca" id="marca" onchange="cambioMarca(this)"
disabled>
        <option value="0">¿Marca?</option>
      </select>
    </td>
    <td align="center" >
      <select name="modelo" id="modelo" style="width:200px;" disabled>
        <option value="0">¿Modelo?</option>
      </select>
    </td>
  </tr>
</table>
<br>
<table>
  <tbody>
    <tr>
      <td style="width:200px;">
        <strong>¿Qué tan importante es para usted este criterio?</strong>
      </td>
      <td>
        <div class="slider" id="slider-clase">
          <input name="slider-input-clase" class="slider-input" id="slider-input-
clase"/>
        </div>
      </td>
      <td id="color-picker-clase" style="width:30px;height: 10px; border:1px
solid rgb(90,97,90); background-color: rgb(178,255,28);"></td>
      <td style="width:100px;">
        <p id="importancia-clase">Muy importante</p>

```

```

<script type="text/javascript">
    var sclase = new Slider(document.getElementById("slider-
class"),document.getElementById("slider-input-class"));
    sclase.setValue(100);
    sclase.onChange=function () {
        var cp = document.getElementById("color-picker-class");
        var ic = document.getElementById("importancia-class");

        var value = sclase.getValue();
        if (value>=75) {
            cp.style.backgroundColor = "rgb(178,255,28)";
            ic.innerHTML="<p id='importancia-class'>Muy
importante</p>";

        } else if(value>=50) {
            cp.style.backgroundColor = "rgb(255,242,0)";
            ic.innerHTML="<p id='importancia-class'>Relevante</p>";
        } else if(value>=25) {
            cp.style.backgroundColor = "rgb(255,170,0)";
            ic.innerHTML="<p id='importancia-class'>Secundario</p>";
        } else {
            cp.style.backgroundColor = "rgb(255,0,0)";
            ic.innerHTML="<p id='importancia-class'>Poco
importante</p>";
        }
    };
</script>
</td>
</tr>
</tbody>
</table>
</fieldset>
</td>
</tr>
<tr>
<td>
<p>
<strong>Tipo de búsqueda:</strong>
<label>
<input id="rbSencilla" type="radio" name="rbAvanzada"
onClick="cambioTipoDeBusqueda(this)" value="rbSencilla" checked>
Sencilla
</label>
<label>
<input id="rbAvanzada" type="radio" name="rbAvanzada" value="rbAvanzada"
onClick="cambioTipoDeBusqueda(this)">
Avanzada
</label>
</p>
</td>
</tr>
<tr>
<td>
<table id="tavanzadas" border=0 cellspacing=0 cellpadding=0 style="display:none;">
<tbody>
<%
String sql = "SELECT id, descriptor, jerarquia, nodoraiz FROM descriptores WHERE
'idTipoProducto\'=1 AND id<>1 AND cualitativo=true";
PostgresSQLWrapper psql = new PostgresSQLWrapper();

```

```

ResultSet rs = psql.getQuery(sql);
String owl,raiz;
Jerarquia j;
while(rs.next()) {
    out.println("<tr>");
    out.println("<td>");
    out.println("<fieldset>");
    out.println("<p>");
    out.println("<strong>"+rs.getString("descriptor")+":</strong> ");
    out.println("<select                                name='select_" + rs.getString("id")+""
id='select_" + rs.getString("id")+"">");
        owl = "file:/CatalogoDeAutomoviles/";
        owl = owl.concat(rs.getString("jerarquia"));
        raiz = rs.getString("nodoraiz");
        try {
            j = new Jerarquia(owl,raiz);

            if(rs.getString("descriptor").compareTo("Equipamiento")==0           ||
rs.getString("descriptor").compareTo("Cilindrada del motor")==0) {
                iterador = j.orderByOrden(j.getChildrenOf(j.getRoot())).listIterator();
            } else {
                iterador = j.getHojas(j.getRoot()).listIterator();
            }
            out.println("<option
value="" + j.getRoot().getLocalName()+"">¿"+rs.getString("descriptor")+?"</option>");
            while(iterador.hasNext()) {
                String value = iterador.next().getLocalName();
                String          option          =          value.toUpperCase().substring(0,
1).concat(value.substring(1));

                option = option.replace("_", " ").trim();
                out.println("<option value='"+value+"'>"+option+"</option>");
            }
        } catch(Exception e) {
            out.println("Error: "+e.getMessage());
        }
        out.println("</select>");

        out.println("</p>");
        out.println("<table><tbody><tr>");

        out.println("<td style='width:200px;'>");
        out.println("<strong>¿Qué tan importante es para usted este
criterio?</strong>");

        out.println("</td>");

        out.println("<td>");
        out.println("<div class='slider' id='slider-"+rs.getInt("id")+"">");
        out.println("<input  name='slider-input-"+rs.getInt("id")+""  class='slider-input'
id='slider-input-"+rs.getInt("id")+""/>");

        out.println("</div>");
        out.println("</td>");

        out.println("<td                                id='color-picker-"+rs.getInt("id")+""
style='width:30px;height:10px;border:1px solid rgb(90,97,90);background-color:rgb(178,255,28);'></td>");

        out.println("<td style='width:100px;'>");
        out.println("<p id='importancia-"+rs.getInt("id")+"">Muy importante</p>");
        out.println("<script type='text/javascript'>");

```



```

</td>
<td>
  <div class="slider" id="slider-mapa" >
    <input name="slider-input-mapa" class="slider-input" id="slider-input-mapa"/>
  </div>
</td>
<td id="color-picker-mapa" style="width:30px;height: 10px; border:1px solid rgb(90,97,90);
background-color: rgb(178,255,28);"></td>
<td style="width:100px;">
  <p id="importancia-mapa">Muy importante</p>
  <script type="text/javascript">
    var smapa = new Slider(document.getElementById("slider-
mapa"),document.getElementById("slider-input-mapa"));
    smapa.setValue(100);
    smapa.onChange=function () {
      var cp = document.getElementById("color-picker-mapa");
      var ic = document.getElementById("importancia-mapa");

      var value = smapa.getValue();
      if (value>=75) {
        cp.style.backgroundColor = "rgb(178,255,28)";
        ic.innerHTML="<p id='importancia-mapa'>Muy importante</p>";
      } else if(value>=50) {
        cp.style.backgroundColor = "rgb(255,242,0)";
        ic.innerHTML="<p id='importancia-mapa'>Relevante</p>";
      } else if(value>=25) {
        cp.style.backgroundColor = "rgb(255,170,0)";
        ic.innerHTML="<p id='importancia-mapa'>Secundario</p>";
      } else {
        cp.style.backgroundColor = "rgb(255,0,0)";
        ic.innerHTML="<p id='importancia-mapa'>Poco importante</p>";
      }
    };
  </script>
</td>
</tr>
</tbody>
</table>
</div>
</td>
</tr>
</tbody>
</table>
</div>
<div id="controles del mapa" style="position: absolute;top:100px;right: 15%">
  <input type="radio" name="type" value="none" id="noneToggle" onclick="toggleControlNavegar(this);"
checked="checked" />
  <label for="noneToggle">Navegar</label>
  <input type="radio" name="type" value="point" id="pointToggle" onclick="toggleControl(this);" />
  <label for="pointToggle">Indicar posición</label>
</div>
</form>
</body>
</html>

```

Archivo “*script.js*” para la interacción del lado del cliente con la aplicación:

```
function inicializar() {
    cambioBusqueda(document.getElementById("rbClasificacion"));
}
```

```
function cambioTipoDeBusqueda(rButton) {
    tabla = document.getElementById("tavanzadas");
    if(rButton.id=="rbAvanzada") {
        tabla.style.display='block';
        x=document.getElementById("select_2");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_3");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_4");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_5");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_6");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_7");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_8");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_9");
        x.selectedIndex=0;
        x.disabled=false;
        x=document.getElementById("select_10");
        x.selectedIndex=0;
        x.disabled=false;
        s2.recalculate();
        s3.recalculate();
        s4.recalculate();
        s5.recalculate();
        s6.recalculate();
        s7.recalculate();
        s8.recalculate();
        s9.recalculate();
        s10.recalculate();
        s2.setValue(100);
        s3.setValue(100);
        s4.setValue(100);
        s5.setValue(100);
        s6.setValue(100);
        s7.setValue(100);
        s8.setValue(100);
        s9.setValue(100);
        s10.setValue(100);
    } else {
        tabla.style.display='none';
    }
}
```

```

    }
}

function cambioBusqueda(rButton) {
    cambioVehiculo(document.getElementById("vehiculo"));
    vehiculo = document.getElementById("vehiculo");
    selects = document.getElementsByTagName("select");

    for(i=0;i<selects.length;i++) {
        if(selects[i].id!="vehiculo") {
            selects[i].selectedIndex=0;
            if(vehiculo.selectedIndex==0) {
                selects[i].disabled=true;
            } else
                selects[i].disabled=false;
        }
    }
}

if (rButton.id=="rbMarca") {

    document.getElementById("thSegmento").style.display='none';
    document.getElementById("tdSegmento").style.display='none';

    document.getElementById("thClase").style.display='none';
    document.getElementById("tdClase").style.display='none';

    document.getElementById("thMarca").style.display="";
    document.getElementById("tdMarca").style.display="";

} else {

    document.getElementById("thSegmento").style.display="";
    document.getElementById("tdSegmento").style.display="";

    document.getElementById("thClase").style.display="";
    document.getElementById("tdClase").style.display="";

    document.getElementById("thMarca").style.display='none';
    document.getElementById("tdMarca").style.display='none';

    if(vehiculo.selectedIndex!=0) {
        document.getElementById("clase").disabled=true;
    }
}

}

function cambioVehiculo(vehiculo) {
    //document.getElementById("submit").disabled=vehiculo.selectedIndex==0;
    if(vehiculo.selectedIndex==0) {
        selects=document.getElementsByTagName("select");
        for (sel in selects) {
            if (selects[sel].id!=vehiculo.id) {
                obj=selects[sel];
                obj.selectedIndex=0;
                obj.disabled=true;
            }
        }
    }
}

```

```

} else {
    if (document.getElementById("rbMarca").checked == true) {
        document.getElementById("marca").selectedIndex=0;
        document.getElementById("modelo").selectedIndex=0;
        document.getElementById("marca").disabled=false;
        document.getElementById("modelo").disabled=false;
        updateMarca();
        updateModelo("vehiculo");
    } else {
        updateModelo("vehiculo");
        document.getElementById("segmento").disabled=false;
        document.getElementById("clase").disabled=true;
        document.getElementById("modelo").disabled=false;
        document.getElementById("clase").selectedIndex=0;
        updateSegmento();
    }
}
}

function cambioSegmento(segmento) {
    if(segmento.selectedIndex==0) {
        updateModelo("vehiculo");
        document.getElementById("clase").disabled=true;
        document.getElementById("clase").selectedIndex=0;
    } else {
        updateClase();
        updateModelo("segmento");
    }
}

function cambioClase(clase) {
    if (clase.selectedIndex==0)
        updateModelo("segmento");
    else
        updateModelo("clase");
}

function cambioMarca(marca) {
    if(marca.selectedIndex==0){
        updateModelo("vehiculo");
    } else {
        updateModelo("marca");
    }
}

function updateSegmento() {
    var ajaxSegmento;
    var segmento = document.getElementById("segmento");
    var value = document.getElementById("vehiculo").value;
    if( window.XMLHttpRequest)
        ajaxSegmento = new XMLHttpRequest(); // No Internet Explorer
    else
        ajaxSegmento = new ActiveXObject("Microsoft.XMLHTTP");

    ajaxSegmento.onreadystatechange = function() {
        if( ajaxSegmento.readyState == 4 ) {
            // Comprobamos si la respuesta ha sido correcta (resultado HTTP 200)

```

```

        if( ajaxSegmento.status == 200 ) {
            segmento.parentNode.innerHTML=ajaxSegmento.responseText;
        }
    }
}
ajaxSegmento.open( "GET", "ajaxresponse.jsp?vehiculo="+value+"&combo=segmento", true );
ajaxSegmento.send( "" );
}

function updateClase() {
    var ajaxClase;
    var clase = document.getElementById("clase");
    var value = document.getElementById("segmento").value;
    if( window.XMLHttpRequest )
        ajaxClase = new XMLHttpRequest(); // No Internet Explorer
    else
        ajaxClase = new ActiveXObject("Microsoft.XMLHTTP");

    ajaxClase.onreadystatechange = function() {
        if( ajaxClase.readyState == 4 ) {
            // Comprobamos si la respuesta ha sido correcta (resultado HTTP 200)
            if( ajaxClase.status == 200 ) {
                clase.parentNode.innerHTML=ajaxClase.responseText
            }
        }
    }
}
ajaxClase.open( "GET", "ajaxresponse.jsp?segmento="+value+"&combo=clase", true );
ajaxClase.send( "" );
}

function updateMarca() {
    var ajaxMarca;
    var marca = document.getElementById("marca");
    var value = document.getElementById("vehiculo").value;
    if( window.XMLHttpRequest )
        ajaxMarca = new XMLHttpRequest(); // No Internet Explorer
    else
        ajaxMarca = new ActiveXObject("Microsoft.XMLHTTP");

    ajaxMarca.onreadystatechange = function() {
        if( ajaxMarca.readyState == 4 ) {
            // Comprobamos si la respuesta ha sido correcta (resultado HTTP 200)
            if( ajaxMarca.status == 200 ) {
                marca.parentNode.innerHTML=ajaxMarca.responseText;
            }
        }
    }
}
ajaxMarca.open( "GET", "ajaxresponse.jsp?vehiculo="+value+"&combo=marca", true );
ajaxMarca.send( "" );
}

function updateModelo(request) {
    var ajaxModelo;
    var modelo = document.getElementById("modelo");
    var value = document.getElementById(request).value;
    if( window.XMLHttpRequest )
        ajaxModelo = new XMLHttpRequest(); // No Internet Explorer
    else

```

```

ajaxModelo = new XMLHttpRequest("Microsoft.XMLHTTP");

ajaxModelo.onreadystatechange = function() {
  if( ajaxModelo.readyState == 4 ) {
    // Comprobamos si la respuesta ha sido correcta (resultado HTTP 200)
    if( ajaxModelo.status == 200 ) {
      modelo.parentNode.innerHTML=ajaxModelo.responseText;
    }
  }
}
if(request=="marca") {
  ajaxModelo.open(
    "GET",
    "ajaxresponse.jsp?"+request+"="+value+"&combo=modelo&vehiculo="+document.getElementById("vehiculo").value, true );
} else {
  ajaxModelo.open( "GET", "ajaxresponse.jsp?"+request+"="+value+"&combo=modelo", true );
}
ajaxModelo.send( "" );
}

```

Archivo “*ajaxresponse.jsp*” para responder a la carga dinámica de datos utilizando AJAX:

```

<%-- Document : ajaxresponse Author : Nahun --%>

<%@page contentType="text/html" pageEncoding="UTF-8"%>
<%@ page import = "java.util.*" %>
<%@ page import = "confusion.*" %>
<%@ page import = "com.hp.hpl.jena.ontology.*" %>
<%@ page import = "com.hp.hpl.jena.rdf.model.*" %>
<%@ page import = "postgressdb.*" %>
<%@ page import = "java.sql.*" %>

<%
// Obtener el parametro con el nombre de la persona
Jerarquia jerarquia =null;
String combo = request.getParameter("combo");
String value, strValue,strOpcion;
if (combo.compareTo("segmento")==0) {
  jerarquia = new Jerarquia("file:/CatalogoDeAutomoviles/clasificacion.owl","automovil");
  value = request.getParameter("vehiculo");
  out.println("<select name=\\"segmento\\" id=\\""+combo+"\" onchange=\\"cambioSegmento(this)\\">");
  out.println("\t<option value=" + 0 + ">¿Segmento?</option>");
  Iterator<OntClass> iterador = jerarquia.getChildrenOf(jerarquia.getNode(value)).listIterator();
  while(iterador.hasNext()) {
    strValue=iterador.next().getLocalName();
    strOpcion=strValue.toUpperCase().substring(0, 1).concat(strValue.substring(1));
    out.println("\t<option value=" + strValue + ">" + strOpcion + "</option>");
  }
  out.println("</select>");
} else if(combo.compareTo("clase")==0) {
  jerarquia = new Jerarquia("file:/CatalogoDeAutomoviles/clasificacion.owl","automovil");
  value = request.getParameter("segmento");
  out.println("<select name=\\"clase\\" id=\\""+combo+"\" onchange=\\"cambioClase(this)\\">");
  out.println("\t<option value=\\"0\\">¿Clase?</option>");
}

```

```

Iterator<OntClass>                                iterador                                =
jerarquia.orderByOrden(jerarquia.getChildrenOf(jerarquia.getNode(value))).listIterator();
ArrayList array = new ArrayList();
while(iterador.hasNext()) {
    OntClass nodo = (OntClass)iterador.next();
    strValue=nodo.getLocalName();
    strOpcion=strValue.substring(strValue.indexOf("_")+1);
    strOpcion=strOpcion.toUpperCase().substring(0, 1).concat(strOpcion.substring(1));
    out.println("\t<option value=" + strValue + "> " + strOpcion + "</option>");
}
out.println("</select>");
} else if(combo.compareTo("modelo")==0) {
    boolean flag = true;
    jerarquia = new Jerarquia("file:/CatalogoDeAutomoviles/clasificacion.owl", "automovil");
    value = request.getParameter("clase");
    out.println("<select name=\"modelo\" id=\""+combo+"\" style=\"width:200px;\" >");
    out.println("\t<option value=" + 0 + ">¿Modelo?</option>");
    String sql = "SELECT p.id, p.descripcion";
    sql = sql.concat(" FROM (SELECT DISTINCT ON (p.producto) producto, p.id, p.descripcion FROM productos p) p,");
    sql = sql.concat(" (SELECT * FROM detallesproductos d WHERE d.\"idDescriptor\"=1 AND (");
    if (value!= null && value.compareTo("")!=0) { //Clase
        sql = sql.concat("d.nodo=");
        sql = sql.concat(value);
        sql = sql.concat(")");
    } else {
        value = request.getParameter("segmento");
        if (value!= null && value.compareTo("")!=0) { //Segmento
            OntClass nodo = jerarquia.getNode(value);
            Iterator<OntClass> hijos = jerarquia.getHojas(nodo).iterator();
            int count=0;
            while(hijos.hasNext()) {
                OntClass temp = (OntClass)hijos.next();
                if(count>0)
                    sql = sql.concat(" OR ");
                sql = sql.concat("d.nodo=");
                sql = sql.concat(temp.getLocalName());
                sql = sql.concat(")");
                count++;
            }
        } else {
            String strMarca = request.getParameter("marca");
            if (strMarca != null && strMarca.compareTo("")!=0) { // Marca
                flag=false;
                sql = "SELECT r.id, r.descripcion FROM (";
                sql = sql.concat("SELECT DISTINCT ON (p.producto) producto, p.id, p.descripcion FROM detallesproductos d,
productos p ");
                sql = sql.concat("WHERE p.id=d.\"idProducto\" AND d.\"idDescriptor\"=1 AND marca=");
                sql = sql.concat(strMarca);
                sql = sql.concat(" AND (");
                value = request.getParameter("vehiculo");
                OntClass nodo = jerarquia.getNode(value);
                Iterator<OntClass> hijos = jerarquia.getHojas(nodo).iterator();
                int count=0;
                while(hijos.hasNext()) {
                    OntClass temp = (OntClass)hijos.next();
                    if(count>0)
                        sql = sql.concat(" OR ");
                    sql = sql.concat("d.nodo=");

```

```

        sql = sql.concat(temp.getLocalName());
        sql = sql.concat("");
        count++;
    }
    sql = sql.concat(") r ORDER BY r.descripcion");
    PostgresSQLWrapper psql = new PostgresSQLWrapper();
    ResultSet rs = psql.getQuery(sql);
    while (rs.next())
        out.println("\t<option value=" + rs.getInt(1) + ">" + rs.getString(2) + "</option>");
    out.println("</select>");
    rs.close();
} else { // Vehículo
    value = request.getParameter("vehiculo");
    OntClass nodo = jerarquia.getNode(value);
    Iterator<OntClass> hijos = jerarquia.getHojas(nodo).iterator();
    int count=0;
    while(hijos.hasNext()) {
        OntClass temp = (OntClass)hijos.next();
        if(count>0)
            sql = sql.concat(" OR ");
        sql = sql.concat("d.nodo=");
        sql = sql.concat(temp.getLocalName());
        sql = sql.concat("");
        count++;
    }
}
}
}
}
if(flag) {
    sql = sql.concat(") d WHERE p.id=d.\"idProducto\" ORDER BY p.descripcion");
    PostgresSQLWrapper psql = new PostgresSQLWrapper();
    ResultSet rs = psql.getQuery(sql);
    while (rs.next())
        out.println("\t<option value=" + rs.getInt(1) + ">" + rs.getString(2) + "</option>");
    out.println("</select>");
    rs.close();
}
} else if(combo.compareTo("marca")==0) {
    jerarquia = new Jerarquia("file:/CatalogoDeAutomoviles/clasificacion.owl", "automovil");

    out.println("<select name=\\\"marca\\\" id=\\\""+combo+"\\\" onchange=\\\"cambioMarca(this)\\\">");
    out.println("\t<option value=" + 0 + ">¿Marca?</option>");

    String sql = "SELECT DISTINCT ON (p.marca) marca";
    sql = sql.concat(" FROM (SELECT d.*,p.marca FROM detallesproductos d, productos p WHERE p.id=d.\"idProducto\"
");
    sql = sql.concat(" AND d. \"idDescriptor\"=1 AND (");

    value = request.getParameter("vehiculo");
    OntClass nodo = jerarquia.getNode(value);
    Iterator<OntClass> hijos = jerarquia.getHojas(nodo).iterator();
    int count=0;
    while(hijos.hasNext()) {
        OntClass temp = (OntClass)hijos.next();
        if(count>0)
            sql = sql.concat(" OR ");
        sql = sql.concat("d.nodo=");
        sql = sql.concat(temp.getLocalName());

```

```

        sql = sql.concat("");
        count++;
    }
    sql = sql.concat(") p ORDER BY p.marca");
    PostgreSQLWrapper psql = new PostgreSQLWrapper();
    ResultSet rs = psql.getQuery(sql);
    while (rs.next()) {
        strValue=rs.getString(1);
        strOpcion=strValue.toUpperCase().substring(0, 1).concat(strValue.substring(1));
        out.println("\t<option value=" + strValue + "> " + strOpcion + "</option>");
    }
    out.println("</select>");
    rs.close();
}
%>

```

## Archivo “*detalles.jsp*” para consultar los detalles de un producto (automóvil):

```

<!-- Document: detalles Author: Nahun Enrique --%>

<%@page contentType="text/html" pageEncoding="UTF-8"%>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
    "http://www.w3.org/TR/html4/loose.dtd">
<%@ page import = "postgressdb.*" %>
<%@ page import = "java.sql.*" %>
<%@ page import = "java.util.*" %>
<html>
    <head>
        <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
        <title>GSPS</title>
    </head>
    <body style="width:100%; height:100%;background-color:#FFFFFF;border-width:0px;margin: 0px;" >
        <div id="banner" align="center">
            <table>
                <tr>
                    <td></td>
                    <td><h2 style="color:#0066FF;" align="center">Geo-Semantic Product Search</h2></td>
                    <td><h2 align="center">-</h2></td>
                    <td><h2 style="color:#CC3333;" align="center">GPSP</h2></td>
                    <td></td>
                </tr>
            </table>
        </div>
        <div id="black_banner" style="background-color:black;height:5px;"/></div>
        <div id="containt" style="background-color:#FFFFFF;overflow:auto;width: 100%" align="center">
            <h3 align="center">Descripción del Producto</h3>
            <table>
                <tr>
                    <td>
                        <fieldset>
                            <legend><strong>General:</strong></legend>
                            <%
                                ResultSet rs;
                                PostgreSQLWrapper psql = new PostgreSQLWrapper();
                                String sql ="SELECT p.producto, p.descripcion, p.precio, p.marca FROM productos p WHERE id=";
                                sql = sql.concat(request.getParameter("ID"));

```

```

rs = psql.getQuery(sql);
rs.next();
out.println("<table border=0 cellspacing=0 cellpadding=2 bordercolor='666633' align='center'>");
out.println("<tbody>");
out.println("<tr><td align='left'><strong>Producto: </strong>"+rs.getString("producto")+ "</td></tr>");
out.println("<tr><td align='left'><strong>Descripción:
</strong>"+rs.getString("descripcion")+ "</td></tr>");
String marca = rs.getString("marca");
marca=marca.toUpperCase().substring(0, 1).concat(marca.substring(1));
out.println("<tr><td align='left'><strong>Marca: </strong>"+marca+ "</td></tr>");
Formatter formatter = new Formatter();
formatter.format("%9.2f", ((Double)rs.getDouble("precio")).doubleValue());
out.println("<tr><td align='left'><strong>Precio: </strong>$"+formatter.toString()+ "</td></tr>");
out.println("</tbody>");
out.println("</table>");
%>
</fieldset>
</td>
</tr>
<tr>
<td>
<fieldset>
<legend><strong>Atributos: </strong></legend>
<%
sql ="SELECT d.descriptor, dp.nodo FROM detallesproductos dp ";
sql = sql.concat("INNER JOIN descriptores d ON dp.\"idDescriptor\"=d.id ");
sql = sql.concat("WHERE dp.\"idProducto\"=");
sql = sql.concat(request.getParameter("ID"));

out.println("<table border=0 cellspacing=0 cellpadding=2 bordercolor='666633' align='center'>");
out.println("<tbody>");

String valor;
rs = psql.getQuery(sql);
while(rs.next()) {
out.print("<tr><td align='left'><strong>");
valor = rs.getString("nodo");
valor = valor.replaceAll("_", " ");
out.print(rs.getString("descriptor")+": </strong>"+valor);
out.print("</td></tr>");
}
out.println("</tbody>");
out.println("</table>");
%>
</fieldset>
</td>
</tr>
</table>
</div>
</body>
</html>

```