



INSTITUTO POLITÉCNICO NACIONAL
Centro de Investigación en Computación

**Sistema de recomendación para la generación de
itinerarios turísticos utilizando datos enlazados**

Tesis

que para obtener el grado de

Maestría en Ciencias de la Computación

presenta

Luis Cabrera Rivera

Asesores:

Dr. Marco Antonio Moreno Ibarra

Dr. Miguel Jesús Torres Ruiz

México, D.F., Junio del 2014





INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 10:00 horas del día 06 del mes de junio de 2014 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

“Sistema de recomendación para la generación de itinerarios turísticos utilizando datos enlazados”

Presentada por el alumno:

CABRERA
Apellido paterno

RIVERA
Apellido materno

LUIS
Nombre(s)

Con registro:

A	1	2	0	3	7	2
---	---	---	---	---	---	---

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

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

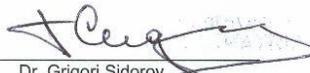
LA COMISIÓN REVISORA Directores de Tesis



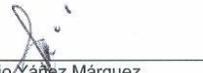
Dr. Miguel Jesús Torres Ruiz



Dr. Marco Antonio Moreno Ibarra



Dr. Grigori Sidorov



Dr. Cornelio Yáñez Márquez



Dr. Luis Manuel Vilches Blázquez



Dra. Nareli Cruz Cortés

PRESIDENTE DEL COLEGIO DE PROFESORES



Dr. Luis Alfonso Villa Vargas
DIRECCIÓN





INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día _____ del mes Junio del año 2014, el (la) que suscribe Luis Cabrera Rivera alumno (a) del Programa de Maestría en Ciencias de la Computación con número de registro A120372, adscrito a Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. María Antonia Morera Ibarra y Dr. Miguel Jesús Torres Ruiz y cede los derechos del trabajo intitulado Sistema de recomendación para la generación de itinerarios turísticos utilizando redes enlazadas, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección gir250@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.


Luis Cabrera Rivera

Nombre y firma

Resumen

La falta de personalización de los recorridos turísticos que se ofrecen a los turistas provocan que visiten únicamente los sitios más conocidos de la zona, sin tomar en cuenta las actividades que le interesaría al usuario realizar o incluso llevándolos a lugares que no son de su agrado.

Por todo lo anterior, en este trabajo se propone una metodología capaz de tomar en cuenta las preferencias de un usuario, como la distancia que gusta recorrer, las opiniones de otros usuarios por medio de redes sociales además de los sitios de interés o actividades que desea visitar durante su estancia.

Esta metodología está conformada por cuatro bloques: 1) *Análisis y tratamiento de los datos originales*, 2) *Caracterización del perfil de usuario*, 3) *Análisis y Geoprocesamiento* y 4) *Visualización*. El primer bloque realiza los cambios necesarios a los datos originales, así como el establecimiento de las conexiones necesarias para recuperar información de redes sociales y *Linked Data*. El segundo bloque procesa las preferencias del usuario para generar un vector de preferencias y posición geográfica del usuario con los datos proporcionados. El tercer bloque realiza el análisis del vector de preferencias y las operaciones espaciales que se requieran realizar, con el fin de generar una ruta acorde a los requerimientos del usuario. Finalmente, el cuarto bloque visualiza la ruta que se desplegará para que el usuario la siga y llegue a sus destinos.

Como resultado se obtiene una ruta ponderada que pasa por los sitios de interés a recomendar así como los especificados por el usuario. Como caso de estudio se tomó el primer cuadro del centro histórico de la Ciudad de México, por la diversidad de los sitios que posee.

Abstract

The lack of personalization of the touristic tours offered to the visitants provoke that they only visit the most important points of interest in the city, without take in consideration the activities that would like them to the tourist or even takes them to places that not interested.

Because of the previously, in this work propose a methodology capable of take in consideration the user preferences, like distance that the user wants to go, opinions of other users by social networks also the points of interest or activities that wants to doing during him stay.

This methodology is composed by four blocks: 1) *Analysis and treatment of originals data*, 2) *Characterization of user profile*, 3) *Analysis and Geoprocessing* and 4) *Visualization*. The first block focuses in make the necessary changes in the original data; As well establish the necessary connections for retrieve information from the social networks and Linked Data. The second block focuses in process the user's preferences in order to generate a vector of preferences and geographic position with the user data. The third block focuses in the analysis of preferences vector and the special operations required to make, in order to generate a route with the requirements of the user. Finally, the fourth block focuses in the visualization of the route will be displayed for the user who will follow it and arrive to their destinations.

As result is obtained a weighted route, which passed through all points of interest to recommended and specified by the user. As case of study is consider the first square of the historic center of the Mexico City, because the diversity of point which found in the area.

Agradecimientos

A mi familia por soportarme y darme apoyo y estabilidad necesaria a pesar de mi temperamento aun inestable y explosivos, pero sin los cuales no estaría aquí.

A los amigos y compañeros del CIC por los momentos que pasamos durante estos 2 años.

A mi amigo Vladimir Luna Soto que se convirtió en una persona importante y de la cual he aprendido muchas cosas a lo largo de nuestra amistad y q me ha ayudado en los momentos justos.

A Imelda Escamilla Bouchán por todo su apoyo y ánimos a lo largo de este tiempo y que aun a la fecha me sigue regañando para mejorar no únicamente como investigador sino como persona.

A Marco Moreno Ibarra que más que asesor ha sido un amigo a lo largo de mi maestría y que estuvo a la altura en los momentos necesarios.

A Miguel Jesús Torres Ruiz que además de ser asesor es un gran amigo y que estuvo en los momentos fuertes de la tesis cuando Marco se fue a CONACYT y que gran parte de la misma se realizó con su colaboración

A Luis Manuel Vilches-Blázquez que me brindo su amistad durante mi estancia en Madrid al igual que Elena y sus anécdotas y preguntas, que sirvieron para aliviar un poco el ambiente que había en aquellos lares del mal, aparte de obligarme y alentarme a rediseñar toda mi tesis y a mostrarme que con poco tiempo puedo hacer algo de magia.

Y los demás mmmm ñe jodance XD

ATT Luis Cabrera Rivera.

Índice

CAPÍTULO 1	INTRODUCCIÓN	1
1.1	DESCRIPCIÓN DEL PROBLEMA	3
1.2	SOLUCIÓN PROPUESTA	4
1.3	OBJETIVOS	5
1.3.1	<i>Objetivos Generales</i>	5
1.3.2	<i>Objetivos Particulares</i>	5
1.4	JUSTIFICACIÓN DEL TRABAJO	6
1.5	ALCANCES Y LIMITACIONES	8
1.6	ORGANIZACIÓN DE LA TESIS	9
CAPÍTULO 2	ESTADO DEL ARTE	11
2.1	SISTEMAS DE RECOMENDACIÓN Y FILTRADO	11
2.2	ONTOLOGÍAS GEOGRÁFICAS	15
2.2.1	<i>Ontologías geoespaciales</i>	16
2.2.2	<i>Ontologías Turísticas</i>	18
2.3	LINKED DATA	20
2.3.1	<i>Linked Data Turístico</i>	23
2.4	SISTEMAS DE RECOMENDACIÓN SEMÁNTICOS	26
2.5	CONCLUSIONES	36
CAPÍTULO 3	MARCO TEÓRICO	37
3.1	ONTOLOGÍA	37
3.2	METODOLOGÍAS	41
3.3	LINKED DATA	51
3.4	<i>RDF</i>	54
3.5	SPARQL	55
3.6	GEOSPARQL	57
3.7	SERVICIOS WEB	58
CAPÍTULO 4	METODOLOGÍA	61
4.1	DESCRIPCIÓN DE LA METODOLOGÍA	61
4.1.1	<i>Análisis y tratamiento de los datos originales</i>	63
4.1.2	<i>Caracterización del perfil de usuario</i>	71
4.1.3	<i>Análisis, Geoprocesamiento, Ruteo</i>	73
4.1.4	<i>Visualización</i>	75
CAPÍTULO 5	PRUEBAS Y RESULTADOS	77
5.1	CASO DE ESTUDIO	77
5.2	ANÁLISIS Y TRATAMIENTO DE LOS DATOS ORIGINALES	78
5.2.1	<i>Diseño de la red de ontologías turísticas</i>	78
5.2.2	<i>Transformación de información original a RDF</i>	87
5.2.3	<i>Enriquecimiento con datos de la Web 2.0 (Redes sociales)</i>	90
5.2.4	<i>Enriquecimiento con datos de la Web 3.0</i>	91
5.3	CARACTERIZACIÓN DEL PERFIL DE USUARIO	92
5.4	ANÁLISIS, GEOPROCESAMIENTO Y RUTEO	93
5.5	VISUALIZACIÓN	94
5.6	EXPERIMENTOS	95
5.6.1	<i>Búsqueda de Sitios</i>	95
5.6.2	<i>Geoprocesamiento</i>	100
5.6.3	<i>Generación de itinerario</i>	109
CAPÍTULO 6	CONCLUSIONES Y TRABAJO A FUTURO	119
6.1	TRABAJO A FUTURO	121
BIBLIOGRAFÍA		123

Índice de Figuras

Figura 1.1 Inversión en telecomunicaciones en 2010	7
Figura 1.2 Participación porcentual de suscripciones en 2010	7
Figura 2.1. Arquitectura de un filtro híbrido (Ahmad, Mahnaz, 2013).....	13
Figura 2.2 Sistema Web SLOTD.....	24
Figura 2.3. Diseño de la página SigTur/E-Destination (Moreno et al., 2013).....	27
Figura 2.4 Mapa de <i>DBpedia Mobile</i> (Becker, C.; Bizer, C. 2008).....	28
Figura 2.5. Diseño Web de SMARTMUSEUM (Ruotsalo et al., 2013)	29
Figura 2.6. Diseño del Sistema Sem-Fit (Garcia et al., 2011)	30
Figura 2.7. Arquitectura del sistema de recomendación SPETA (García et al., 2009)	33
Figura 2.8 Vista general de la aplicación More! (Parra, G.; Klerkx, J.; Duval, E. 2010).....	34
Figura 2.9 Arquitectura del sistema wayUO (Zablith, F.; Fernandez, M.; Rowe.M, 2011)	35
Figura 3.1.- Tareas que componen la metodología METHONTOLOGY(Corcho et al., 2005)...	44
Figura 3.2.- Escenarios para construcción de ontologías y ontologías de red (Suárez et al., 2012)	49
Figura 3.3.- Último diagrama del estado de <i>Linked Data</i>	51
Figura 3.4.- Tripletas en <i>RDF</i>	55
Figura 3.5.- Esquema general de un servicio Web	59
Figura 4.1 Diagrama general de la metodología propuesta.....	62
Figura 4.2 Fuentes de datos del <i>Triple Store</i>	64
Figura 4.3.- Ontologías que integran la red de ontologías	65
Figura 4.4 Procesamiento de <i>shapefiles</i> y <i>POIs</i> a formato <i>RDF</i>	67
Figura 4.5 Diagrama de integración de evaluaciones a <i>RDF</i>	68
Figura 4.6 Integración de con fuentes de la Web 3.0.	70
Figura 4.7 Diagrama entrada y salida del bloque de Caracterización.	71
Figura 4.8 Diagrama general del bloque de Análisis, Geoprocesamiento y Ruteo	73
Figura 5.1 Zona 1 primer cuadro del Centro Histórico de la Ciudad de México.	77
Figura 5.2 Taxonomía de la ontología	80
Figura 5.5 <i>Tweet</i> de la librería Gandhi	91
Figura 5.6 Búsqueda de sitios.	96
Figura 5.7 Lista de sitios.	97
Figura 5.8 Ubicación de marcador representando un sitio.	98
Figura 5.8 Categorías de la ontología.	99
Figura 5.9 Despliegue de marcadores por categoría.	100
Figura 5.10 Ventana de Geoprocesamiento	101

Figura 5.11 Marcadores de posición.	102
Figura 5.12 Despliegue de barra para tamaño de buffer.	103
Figura 5.13 Resultados Buffer.	104
Figura 5.14 <i>Intersect</i> sobre calle del usuario.	105
Figura 5.15 Resultado <i>Intersect</i>	106
Figura 5.16 Visualización del <i>Routing</i>	107
Figura 5.17 Preparando información para <i>Routing</i>	108
Figura 5.18 Resultado Routing	109
Figura 5.19 Ventana principal Itinerario	110
Figura 5.20 Llenado de Actividades	111
Figura 5.21 Menú principal con Actividad	112
Figura 5.22 Ventana de para agregar sitios.....	113
Figura 5.23 Ventana para preferencias.	114
Figura 5.24 Perfil de usuario completo.	115
Figura 5.25 Resultado del itinerario turístico personalizado.....	117

Índice de Tablas

Tabla 2.1 Características de diferentes Linked Data geospaciales (Vilches-Blázquez, L. M. et al 2013).....	22
Tabla 5.1 Ejemplo del Glosario de Términos	79
Tabla 5.2 Ejemplo del Diccionario de Conceptos	81
Tabla 5.3 Ejemplo del Diccionario de Conceptos	82
Tabla 5.4 Ejemplo de Descripción de atributos de clase	83
Tabla 5.5 Ejemplo de Descripción de constantes.....	84
Tabla 5.6 Ejemplo de Descripción de reglas.....	85
Tabla 5.7 Ejemplo de Descripción instancias	86

Capítulo 1 Introducción

En la actualidad existen varias agencias turísticas, las cuales diseñan recorridos por sitios de concurrencia turística en un determinado orden y con tiempos estimados de recorridos, también llamados itinerarios turísticos. Estos itinerarios no toman en consideración el tiempo del que dispone el turista para visitar, debido a que éstos ya están prediseñados con los sitios de interés más concurridos como museos, restaurantes, sitios arqueológicos, entre otros.

Sin embargo, existen usuarios que desean conocer varios lugares, y no uno en específico, un ejemplo de ello es cuando un usuario desea conocer algún museo de la ciudad, pero solo le interesan los museos de arte barroco, o desea ir a comer a algún restaurante de comida japonesa cercano o visitar el salón de baile más popular de la ciudad.

Para tomar en cuenta las preferencias de un usuario, se emplea un perfil de usuario; un perfil es una representación de información esencial acerca de un usuario (Schiaffino et al., 2009), donde la información que pueden recopilar depende del dominio de la aplicación donde se empleará, en este caso se recuperan las actividades y sitios que desea visitar a lo largo de un día, así como preferencias diversas como pesos para determinar la importancia de los criterios a considerar.

Debido a estas necesidades de personalización, existe un gran auge de sistemas que permiten sugerir ya sean objetos para compras por Internet, páginas para que el usuario visite o la visita a sitios de interés, todo esto conforme a un perfil de usuario; estos sistemas son llamados *sistemas de recomendación* (Bobadilla et al., 2013). Estos sistemas permiten por medio de un análisis el poder

recomendar cualquier cosa con base en las preferencias de los usuarios, además de utilizar filtros para mejorar la precisión de las sugerencias. Existen varios tipos de filtros que se emplean en estos sistemas y que otorgan mejores resultados dependiendo de cada uno, en este caso para la incorporación de opiniones de usuarios se emplean los filtros *colaborativos* (Sarwar, B. et al., 2001)., éstos filtros permiten a los sistemas usar las evaluaciones realizadas por otros usuarios y así poder crear una recomendación con mayor precisión para el usuario sobre algún objeto conforme a las preferencias de usuarios similares a él.

Los sistemas de recomendación están muy ligados al uso de servicios basado en localización, los cuales ayudan a determinar la posición de un usuario para después ser usada junto con información externa para proveer aplicaciones y servicios personalizados (Zibuschka, J. et al., 2011). En este caso, para el dominio turístico los servicios basados en localización ayudan a recuperar la posición de un usuario, además de recuperar la posición de los sitios de interés a recomendar.

Actualmente, existen varias aplicaciones que proporcionan servicios de localización, así como breves reseñas de los mismos, tal es el caso de la red social *Foursquare* (Mao, Peifeng, Wang-Chien, 2010), la cual le permite ver que usuarios han visitado el sitio en el que se encuentran, así como breves opiniones, ya sean positivas o negativas, del sitio en cuestión. Además, esta red social permite agregar nuevos sitios o buscar un determinado tipo de lugar de interés.

Con respecto a la implementación de estos sistemas en dispositivos móviles, existen diversas propuestas debido a que este tipo de dispositivos forman parte del uso diario de las personas. Así, por ejemplo, esto permite realizar recomendaciones de sitios utilizando servicios con base en su localización; de tal manera que no importa que no se conozca la región que visita, estas aplicaciones basadas en la tecnología *GPS* o *GLONASS* (Virrataus et al., 2001) permiten encontrar los sitios que se desean.

Debido a esto, en este trabajo se propone un sistema de recomendación que explota información conforme a los principios de *Linked Data*. Este sistema permitirá el uso de un método semántico para sugerir sitios de interés en un contexto turístico. Este método considera las opiniones de otros usuarios extraídas de algunas redes sociales, aspectos semánticos (como la categoría a la que pertenecen los sitios) y aspectos geográficos. Adicionalmente, se propone la inclusión de un perfil de usuario para recuperar las preferencias del usuario y procesarlas en el sistema, el cual estará basado en una lista de actividades y sitios que desean visitar. Con lo anterior, se recuperará las ponderaciones que proporcione el usuario, junto con la información recuperada de las redes sociales (opiniones y *ratings*), de tal modo que se puedan obtener resultados más precisos y así satisfacer las necesidades de los usuarios.

Para tal fin, se realiza un procesamiento semántico sobre los datos definidos en una ontología de aplicación en el dominio turístico, con la finalidad de recuperar información de manera semántica. El sistema se implementará en un dispositivo móvil, bajo el lenguaje *Objective-C* para el sistema operativo *iOS*. Como caso de estudio se propone aplicar la metodología y el sistema sobre el primer cuadro del Centro Histórico de la Ciudad de México, debido a la diversidad de lugares de interés que existen en el área mencionada.

1.1 Descripción del Problema

Por lo regular, las personas suelen viajar a lugares que les son desconocidos, donde frecuentemente existe una diversidad de atractivos o sitios de interés; sin embargo, generalmente cuentan con tiempo restringido para visitar los lugares que son de su interés; para esto existen aplicaciones móviles como *Foursquare* (Mao, Peifeng, Wang-Chien, 2010), *iTravel* (Wan-Shiou, San-Yih, 2013), *Turist@* (Batet, Moreno, Sánchez, 2012) o sitios web como *SigTur* (Moreno et al., 2013), *SemFit* (Garcia et al., 2011) para sugerir los lugares a visitar.

Desafortunadamente, estas recomendaciones se realizan sin considerar las preferencias particulares o perfiles de usuario de cada visitante, lo que conlleva a tener resultados inesperados o no deseados por los usuarios, ya que éstos pueden estar muy alejados de sus preferencias o pueden coincidir con las mismas.

Estos servicios o aplicaciones no se encargan de recolectar las opiniones de los usuarios acerca de los lugares visitados, y/o solo utilizan o toman en cuenta los sitios en un determinado radio, dejando en su itinerario sitios que no cumplen con las expectativas de los usuarios o con las características basadas en su interés, excluyendo por tanto, otros sitios que podrían ser de su interés. Este tipo de sistemas operan bajo un esquema global de consultas a nivel de coincidencia exacta; en otras palabras, bajo un enfoque totalmente léxico, lo cual impide procesar semánticamente y medir conceptualmente los perfiles de los usuarios.

1.2 Solución Propuesta

Con base en el planteamiento del problema, se propone diseñar e implementar una metodología que permita analizar un perfil del usuario para determinar cuáles son los mejores sitios a los que un usuario puede ir. Además, se le proporciona al usuario una ruta que contenga estos lugares así como información relacionada de los mismos. Esta ruta se adaptará a las actividades que el usuario desea realizar.

Como se mencionó con anterioridad, el caso de estudio propone implementar un sistema de recomendación que explote información conforme a los principios de *Linked Data* permitiendo procesar semánticamente los perfiles de

usuario y proporcionar los resultados con mayor similitud a las características de cada usuario y con base en su interés.

1.3 Objetivos

En esta sección se describirán de manera exhaustiva los objetivos tanto generales como particulares que son planteados para este trabajo.

1.3.1 Objetivos Generales

Diseñar e implementar una metodología para proporcionar sitios de interés e itinerarios turísticos con base en un perfil de usuario y sus características asociadas siguiendo los principios de *Linked Data*.

1.3.2 Objetivos Particulares

- Diseñar y construir una ontología turística en el dominio turístico con las consideraciones particulares del área de estudio.
- Desarrollar una arquitectura funcional para implementar las diferentes etapas de la metodología.
- Generar y publicar *Linked Data* sobre información turística de la zona de estudio.
- Diseñar e implementar un sistema de recomendación para el contexto turístico, como caso de estudio para aplicar la metodología propuesta.

- Considerar la opinión e intereses de los usuarios a través de información proveniente de redes sociales.

1.4 Justificación del Trabajo

Este trabajo brinda una opción enfocada al fortalecimiento del área *e-Tourism*, ofreciendo al usuario una herramienta de cómputo que permite planear sus vacaciones o salidas, a partir del análisis de sus preferencias. Así mismo, permite la difusión de lugares no tan famosos que otros usuarios pueden preferir y que son considerados de interés para su visita; más allá de los sitios turísticos ya establecidos.

El valor agregado es que esta aplicación está enfocada hacia su uso y explotación en dispositivos móviles (ejemplos de otras aplicaciones Mobile GIS [Steiniger, Hunter, 2013]), ya que de acuerdo con datos estadísticos de la COFETEL (Referencia COFETEL), a junio de 2010 existían en México alrededor de 89.8 millones de teléfonos celulares, de los cuales aproximadamente el 10% se pueden clasificar como teléfonos inteligentes (*smartphones*). Esto significa que actualmente hay más de 9 millones de teléfonos móviles inteligentes que interconectados se convierten en una fuente de información importante. En la Figura 1.1, se muestra la inversión en telecomunicaciones del año 2010 y se aprecia cómo la telefonía móvil tiene el 52.1% sobre las demás.

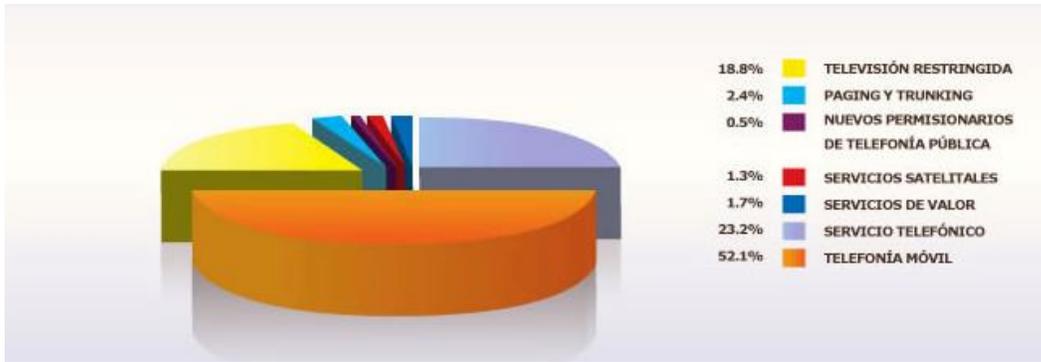


Figura 1.1 Inversión en telecomunicaciones en 2010

De igual manera, en la Figura 2 se presenta la participación porcentual de suscripciones en las principales ciudades en el año 2010, en donde se aprecia que la participación de la Ciudad de México destaca con un 22.59%. Con base en estos datos justificamos la importancia del desarrollo de aplicaciones orientadas a dispositivos móviles.

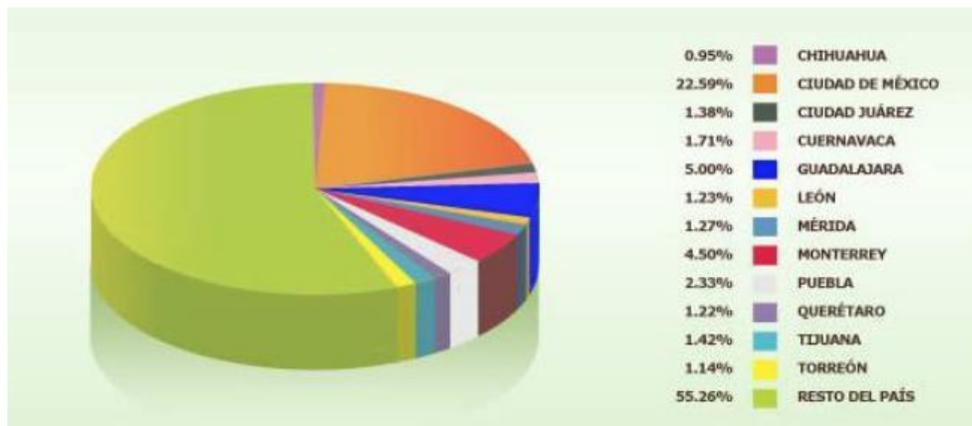


Figura 1.2 Participación porcentual de suscripciones en 2010

Como caso de estudio para esta tesis, se ha considerado el primer cuadro de la Ciudad de México, ya que presenta una alta concurrencia de visitantes extranjeros y foráneos, tomando en cuenta además, la gran variedad de

establecimientos existentes en el mismo, así como la vasta cantidad de museos y sitios arqueológicos que posee esta zona en particular.

1.5 Alcances y Limitaciones

La metodología será capaz de poder realizar operaciones de geoprocésamiento, tales como: *buffer*, *intersect*, *contains* y *touch* sobre la ontología que se construirá. Además, se utilizará el perfil de usuario para capturar las preferencias y ponderaciones del usuario.

El sistema también será capaz de recuperar información de la Web y enriquecer los datos que se tienen utilizando información de la Web de *Linked Data*, así como de diferentes redes sociales en este caso *Twitter* y *Foursquare*.

Como limitantes del trabajo se encuentran los siguientes:

- La falta de datos públicos para tener un buen enriquecimiento de datos.
- El uso de internet para realizar las peticiones al servidor.
- El sistema no considera el tiempo para realizar recomendaciones, ya que se utiliza un tiempo promedio únicamente en sitios culturales como museos, iglesias, etc.
- El sistema mostrará únicamente un segmento a la vez de la ruta que debe seguir del usuario, sin embargo no le proporcionará las indicaciones para llegar al siguiente sitio de interés.

1.6 Organización de la Tesis

Este trabajo de tesis está compuesto por los siguientes capítulos:

- **Capítulo 2.** Este capítulo se centra en el estado del arte, esta compuesto por trabajos relacionados en un sentido científico y tecnológico, acerca de sistemas de recomendación, medidas de similitud semántica y ontologías turísticas.
- **Capítulo 3.** Este capítulo trata sobre el marco teórico de la tesis, donde se mencionan los conceptos relevantes como ontologías, sistemas de recomendación, entre otros.
- **Capítulo 4.** Este capítulo describe la metodología propuesta con base en los conceptos descritos en el marco teórico, así como la arquitectura del sistema de recomendación.
- **Capítulo 5.** Este capítulo presenta los experimentos realizados y los resultados obtenidos, los cuales servirán para un análisis de la metodología propuesta.
- **Capítulo 6.** En este capítulo se presentan las principales contribuciones y conclusiones obtenidas en el proceso de investigación desarrollado en este trabajo, así como algunas propuestas como relacionadas con trabajos futuros.

Capítulo 2 Estado del Arte

En este capítulo se describen algunos trabajos relacionados con los sistemas de recomendación, así como su uso en dispositivos móviles e incorporación de *Linked Data*. Por otra parte, se considera un análisis enfocado en trabajos relacionados con recuperación semántica de información y fundamentos sobre el perfil de usuario.

2.1 Sistemas de recomendación y filtrado

Los *sistemas de recomendación* fueron desarrollados con la finalidad de proporcionar recomendaciones sobre algún objeto, algunos ejemplos de recomendaciones serían películas, revistas, videojuegos, sitios turísticos, entre otros. Estos sistemas recolectan información sobre las preferencias de los usuarios como un conjunto de objetos (libros, tipos de comida, aplicaciones etc.) (Bobadilla et al., 2013). Esta información puede ser adquirida de manera explícita [(Rich,1983), (Golemati et al., 2007)], cuando un usuario llena un registro de manera implícita [(Seok Lee, Yoon Ho, Soung Hie, 2010), (Rich,1983), (Golemati et al., 2007)], usualmente obtenida desde redes sociales, sitios web visitados, películas vistas o música escuchada entre otros.

Los sistemas de recomendación son una técnica ampliamente utilizada en el comercio electrónico (*e-Commerce*) (Sarwar et al., 2002), debido a que existe una alta precisión en las recomendaciones generadas para los usuarios entre los millones de objetos que se venden a través de la Web. Un ejemplo de ello, es el sitio de comercio electrónico *Amazon* (Linden, Smith, York, 2003), el cual utiliza filtros para comparar objeto por objeto, con la finalidad de recomendar aquellos que son más adecuados a sus usuarios.

En los sistemas de recomendación se utiliza un algoritmo de filtrado para llevar a cabo el análisis del perfil de usuario y de acuerdo con [(Adomavicius, Tuzhilin, 2005), (Schafer et al., 2007), (Candiller, Meyer, Boullé, 2007)] están clasificados de la siguiente manera:

- **Basado en contenido.** Este filtro utiliza las recomendaciones realizadas, basado en las decisiones tomadas por el usuario previamente, así como las que obtiene de analizar los objetos que fueran recomendados con anterioridad. Al utilizar una métrica de similitud sobre éstos, se puede determinar qué otros objetos se pueden recomendar al usuario.
- **Basado en el sector demográfico.** Este tipo de filtro se utiliza en ciertas características como el género, la edad o el país para determinar las preferencias en común con otros usuarios que comparten características similares.
- **Basado en colaboración.** Este filtro utiliza las opiniones de los usuarios sobre un conjunto de objetos (películas, sitios de interés, videojuegos, etc.), para almacenarlo y recopilar la cantidad de información necesaria para poder generar de mejor manera las recomendaciones para los usuarios, basados en las evaluaciones realizadas por otros usuarios con preferencias similares a las de ellos.

De los filtros anteriormente mencionados, los filtros colaborativos han sido parte esencial en el desempeño de los sistemas de recomendación, debido a que utilizan la experiencia de otros usuarios para hacer recomendaciones, o incluso el uso de redes sociales para determinar de manera implícita sus preferencias.

Los filtros híbridos suelen usar la combinación de filtros (Burke, 2002), tales como colaborativos, basados en contenido y en el sector demográfico. Estos filtros proporcionan las ventajas de los diferentes tipos de filtros antes mencionados, dando una recomendación más precisa para el usuario. Comúnmente se suelen combinar los filtros colaborativos con otro tipo de filtro, como los que están basados en el sector demográfico o con los que se basan en contenido. Estos filtros suelen estar basados en modelos probabilísticos o en bio-inspirados como los algoritmos genéticos, genética difusa, redes neuronales, redes bayesianas, *clustering* y propiedades latentes. La arquitectura de estos filtros se muestra en la Figura 2.1.

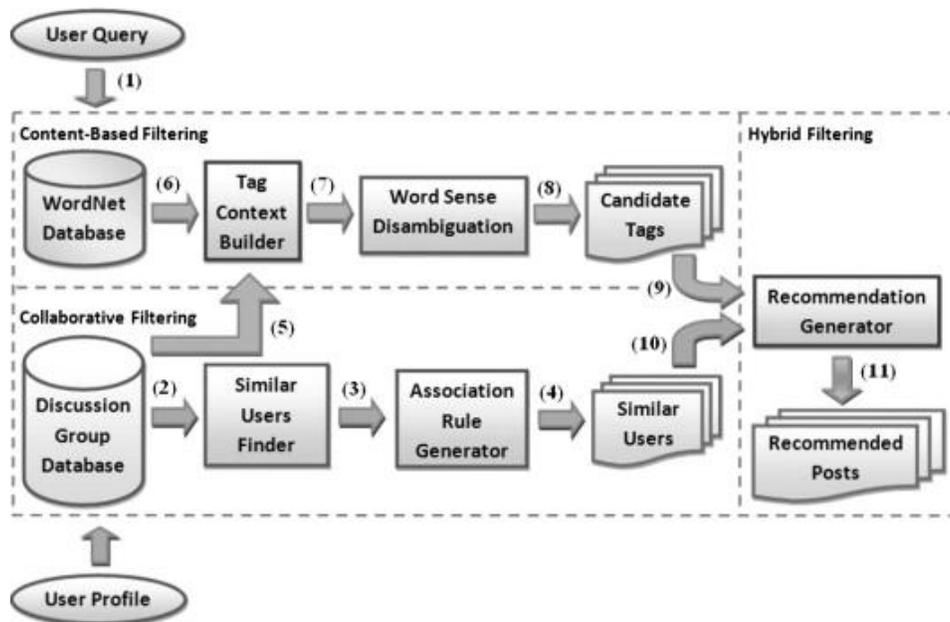


Figura 2.1. Arquitectura de un filtro híbrido (Ahmad, Mahnaz, 2013)

En la figura anterior se muestra como están constituidos los filtros híbridos, en este caso se utilizan dos filtros, uno basado en contenido el cual utiliza la base de datos de *WordNet* para realizar un etiquetado de contexto y una desambiguación para recuperar las etiquetas más apropiadas, en el segundo filtro se emplea uno de tipo colaborativo, el cual nos recupera los usuarios similares

conforme al perfil de usuario. Después se juntan las salidas de ambos filtros para determinar lo que se le recomendará al usuario conforme a las etiquetas que concuerden o que hayan sido escogidas por usuario con preferencias similares al de nuestro usuario.

Para este trabajo se ocuparán los filtros de tipo colaborativo, esto debido a que se implementará una recuperación de información de las redes sociales para determinar las evaluaciones hechas por los usuarios de las mismas, es por eso que a continuación se describen con mayor detalle estos filtros.

Los filtros colaborativos utilizan algoritmos para realizar el filtrado de preferencias de los usuarios o de propiedades de objetos, estos algoritmos se dividen en dos categorías, según [(Adomavicius, Tuzhilin, 2005), (Candiller, Meyer, Boullé, 2007), (Bobadilla et al., 2013)].

- **Algoritmos basados en memoria.** Estos algoritmos son esencialmente heurísticas que realizan una predicción de las evaluaciones, basado en una colección de evaluaciones previas a los objetos. Estos algoritmos utilizan una medida de similitud para obtener la distancia, ya sea entre dos usuarios o dos objetos.
- **Algoritmos basados en modelos.** Este tipo de métodos utilizan una colección de evaluaciones para entrenarse y aprender un modelo, del cual se realizan predicciones de evaluaciones de objetos. Un ejemplo de ellos son las aproximaciones probabilísticas donde la calificación de un objeto es calculada con base en las evaluaciones previas.

2.2 Ontologías geográficas

Las ontologías geográficas se pueden definir en una doble perspectivas según (Vilches-Blázquez, L. M. 2011), una visión general y técnica que surge del ámbito de ISO/TC211 y, por otro lado, una perspectiva más específica y vinculada con el dominio geoespacial.

Desde la visión general, ISO/TC211 afirma que una ontología en el dominio geográfico se refiere a la representación formal de un fenómeno con un vocabulario subyacente. Esta representación incluye definiciones y axiomas que hacen explícito el significado pretendido y que sirven para describir al fenómeno y a aquellas interrelaciones que puedan ser utilizadas por las aplicaciones software para dar soporte a la compartición, reutilización e integración de información geográfica con cualquier otra fuente de información dentro de un dominio de conocimiento, así como entre varios dominios de conocimiento.

Desde una perspectiva más relacionada con el dominio, (Fonseca *et al.*, 2006) afirma que el concepto geo-ontología tiene dos tipos básicos de conceptos. Por un lado, los conceptos correspondientes a los fenómenos físicos en el mundo real. Por otro, los conceptos que corresponden a fenómenos del mundo que han sido creados para representar construcciones institucionales y sociales. De esta manera, los conceptos en una geo-ontología se dirigen hacia los objetos espaciales presentes en el mundo.

Estas visiones se pueden resumir afirmando que una ontología geográfica se diseña para ofrecer un mejor entendimiento de la estructura del mundo geográfico (Smith *et al.*, 1998).

En definitiva, estas definiciones hacen entender que las ontologías no están formadas únicamente por meros conceptos con una cierta organización, lo que no distaría de las formas más habituales de organización de la IG (catálogos de fenómenos, diccionarios de datos y/o tesauros), sino que también se van a definir relaciones, atributos, reglas y axiomas entre conceptos que enriquecen y contribuyen, entre otros aspectos, a formalizar el vocabulario y conocimiento del área, dejando a un lado el simple almacenamiento realizado sobre la parte más específica de la información, es decir, los datos (Vilches *et al.*, 2008).

En la actualidad parece reconocerse ampliamente que las ontologías tienen significativos beneficios para capturar y compartir la semántica de la información geoespacial (Klien *et al.*, 2005) y servir para el desarrollo de sofisticados SIG (por ejemplo, Fonseca *et al.*, 2002) y/o IDE (por ejemplo, Lacasta *et al.*, 2005; Lutz *et al.*, 2009). A continuación, se recogen algunos datos (grupos, proyecto, líneas de investigación y ontologías desarrolladas) que reflejan este reconocimiento.

2.2.1 Ontologías geoespaciales

En (Ressler *et al.*, 2007) se recogen un estudio sobre ontologías relevantes para su utilización en anotación razonamiento cualitativo e interoperabilidad en la comunidad de inteligencia geoespacial. En este trabajo se recogen 45 ontologías espaciales y temporales divididas en 11 categorías, compuestas por las 7 categorías del *W3C Geospatial Incubator Group* (descritas con anterioridad) y 4 categorías adicionales para representar completamente el rango de datos geoespaciales y temporales (*Geometric ontology*, *Coverage ontology*, *Geopolitical ontology* y *Temporal ontology*). A continuación se describen trabajos, de manera no exhaustiva, acerca de ontologías geoespaciales.

Ontology for Geography Markup Language (GML3.0) es la representación en OWL del estándar Geography Markup Language versión 3.0 (GML) del Open Geospatial Consortium. GML es el estándar OGC para el transporte y almacenamiento de información geográfica, incluyendo las propiedades y las geometrías de los fenómenos geográficos. Esta ontología surgió en la Universidad de Drexel en 2004 sobre una versión antigua del GML (la 3.0, cuando la estándar para los WFS es la 3.1). La idea de implementar el GML como ontología es que sirviera como ontología base para el desarrollo de otras ontologías que extendiesen las clases y propiedades de estas clases definidas en ella (perfiles de aplicación de GML) para ser utilizadas en el ámbito de sistemas de información que necesiten explotar tanto la componente geográfica o espacial como las relaciones semánticas de sus datos.

Geospatial Resource Description Framework (GRDF) (Alam, Khan, Thuraisingham, 2008) es una ontología en OWL desarrollada por la Universidad de Texas y construida sobre las clases y propiedades definidos en GML. El fin de esta ontología es proponer un lenguaje para el dominio geoespacial, amplio, expresivo y que haga uso de las bondades de la Web semántica. La ventaja más importante que tiene sobre otros lenguajes geoespaciales es la posibilidad de usar inferencia lógica y agregación dinámica de contenidos.

OntoSensor (Russomanno *et al.*, 2005) es una ontología en OWL que referencia y extiende la ontología IEEE Suggested Upper Merged Ontology (SUMO) que define conceptos y asociaciones generales. Se basa en parte en SensorML, el estándar OGC para la temática de sensores y que a su vez está basado en GML.

2.2.2 Ontologías Turísticas

Dentro de las ontologías geoespaciales, se deriva el dominio turístico, el cual está compuesto objetos geográficos de carácter turístico, ideales para realizar una interpretación de sitios y actividades que se pueden realizar en un área de estudio por ejemplo visitas a sitios históricos, recorridos por museos, etc., en la actualidad existen varias ontologías en este dominio, éstas están basadas en estándares globales como el tesoro de la Organización Mundial del Turismo o *Gazetteer* que fungen como vocabularios para su construcción, estas ontologías son utilizadas en los sistemas de recomendación, para complementar los filtros implementados, a continuación se describen algunas de las ontologías turísticas que se revisaron y se mencionan en el trabajo de (Gutiérrez 2010):

Harmonise Ontology (Höpken, W. 2006). El objetivo de esta ontología es proporcionar a las organizaciones turísticas la capacidad de intercambiar información sin necesidad de modificar sus estructuras internas ni sus sistemas de información. Así pues, se puede entender *Harmonise* como una ontología que actúa como mediadora entre diferentes ontologías turísticas, actuando de enlace semántico entre sistemas, permitiendo al receptor interpretar la fuente de datos como una extensión de su propia base de datos sin preocuparse de cómo están esos datos representados.

Está implementada en *RDF*, y contiene alrededor de 200 conceptos y propiedades para describir entidades turísticas, centrados fundamentalmente en el campo del alojamiento (hoteles, *bed & breakfast*, turismo rural, campings), de los eventos y actividades (festivales, conferencias, eventos deportivos), de la gastronomía, y de los monumentos y lugares de interés.

Mondeca Tourism Ontology¹.- La ontología turística desarrollada por *Mondeca* introduce conceptos fundamentales en el ámbito del turismo definidos en el tesoro gestionado por la Organización Mundial de Turismo (*WTO*), los cuales incluyen información y definiciones del dominio turístico y de las actividades de ocio. La cobertura de la ontología de *Mondeca* abarca entidades turísticas, culturales, paquetes turísticos y contenido multimedia. Está desarrollada en lenguaje *OWL* y contiene alrededor de 1000 conceptos.

Hi-Touch Ontology (Hi-Touch Working Group, 2003). La ontología *Hi-Touch*, se desarrolló dentro del programa europeo *IST/CRAFT Hi-Touch*, destinado a establecer metodologías de Web Semántica y al desarrollo de herramientas para colaborar en el campo del turismo sostenible europeo. El objetivo de este proyecto era modelar el conocimiento sobre las expectativas de los viajeros, así como ofrecer productos turísticos personalizados.

La ontología, esta codificada en *OWL*, fue desarrollada principalmente por *Mondeca* y el núcleo de la misma está construido a partir del repositorio de datos semánticos de esta compañía. Esta ontología clasifica entidades turísticas y las relaciona a partir de una red de relaciones semánticas proporcionada por un mapa de temas. Las clases de nivel superior de la ontología son Documentos, Objetos y Publicación.

- Documentos Es cualquier tipo de documentación o anuncio sobre un producto turístico.
- Objetos se refiere a las propias entidades turísticas.
- Publicación Es el documento generado a partir del resultado de una consulta.

¹ <http://www.mondeca.com>

Los objetos turísticos pueden ser clasificados además por palabras clave utilizando el tesoro de sinónimos de conceptos turísticos y de actividades de de la Organización Mundial de Turismo.

OnTour Ontology².- Es una ontología creada específicamente para el dominio turístico y fue desarrollada por DERI (Digital Enterprise Research Institute), esta ontología está diseñada en el lenguaje OWL-DL. Aparte de tomar en cuenta los conceptos tradicionales del dominio turístico, este incluye nuevos conceptos como actividades de esparcimiento e información geográfica.

2.3 Linked Data

Dentro de *Linked Data* existen varios trabajos que proporcionan un contexto mínimo necesario para comprender el principio, un ejemplo de estos es el trabajo de (Auer, S. et al. 2013) el cual proporciona una breve introducción sobre los principios de *Linked Data*, conceptos básicos para su manejo como los lenguajes que existen para consulta, almacenamiento, métodos existentes, etc., además de hablar acerca del ciclo de vida de *Linked Data*.

Otro trabajo interesante es el de (Heath y Bizer 2011) el cual ofrece una visión general de los principios de *Linked Data*, además de presentar un conjunto de patrones para la publicación *Linked Data*. Sin embargo, este no hace uso de información geoespacial además de no contar con una implementación.

En cuanto al ámbito geográfico en el trabajo de (Vilches-Blázquez, L. M. et al 2013) se mencionan varios trabajos como se ve en la Tabla 2.1, donde se han realizado varias propuestas para la publicación de datos geográficos, ejemplo de

² <http://e-tourism.deri.at/ont/index.html>

ellos son las iniciativas *Ordnance Survey* y *OpenStreetMap*, las cuales surgen como los principales proveedores de información de *Linked GeoData* a la fecha. Estas iniciativas utilizan las primitivas de punto, línea y polígono, las cuales ofrecen información geográfica conforme a *Geography Markup Language (GML)* (*Ordnance Survey*) o *Well-Known Text (WKT)*.

Con respecto al vocabulario de estas iniciativas, podemos ver que las ontologías que se utilizan son diversas, usando su propio vocabulario en algunos casos aunque la mayoría de los enfoques de estas ontologías emplean el vocabulario *WGS84*. Por último, los conjuntos de datos que se utilizan son diferentes para cada enfoque, donde el idioma inglés es predominante. Sin embargo, *DBpedia*³ y *GeoNames*⁴ proporcionan soporte para diferentes idiomas como español, italiano, francés, entre otros.

³ <http://dbpedia.org/About>

⁴ <http://www.geonames.org/>

Tabla 2.1 Características de diferentes *Linked Data* geoespaciales (Vilches-Blázquez, L. M. et al 2013)

Proveedor	Procedencia de datos	Triple Store	Tópico	Conjunto de datos	Generación de RDF	Patrón de RDF	Vocabulario
Ordnance Survey	Ordnance Survey	Talis	Unidades Administrativas	Un Gazetteer (1:50,000) y el gazetteer administrativo par Gran Bretaña	-	http://data.ordnance-survey.co.uk/id/ . +	Ontología de relaciones espaciales, ontología de administración geográfica, WGS84 Geo Positioning, FOAF, OWL, y ontología Gazetteer
Linked GeoData	OpenStreet Map	Virtuoso	Puntos de Interés	Base de datos de OpenStreet Map	Triplificación y propietario del framework de mapeo	http://linked-geodata.org/triplify/ . +	Ontología LGD, WGS84 Geo Positioning
Geonames	Geonames	RDF dump	Topónimos	Fuentes de datos usados para Geonames	-	http://sws.geonames.org/id/	Ontología GeoNames, WGS84 Geo Positioning
DBpedia	Wikipedia	Virtuoso	Conocimiento general	Wikipedia	Framework de extracción de DBpedia	http://dbpedia.org/resource/ +	Ontología DBpedia, WGS84 Geo Positioning

2.3.1 Linked Data Turístico

Dentro del concepto de *Linked Data* se han realizado varios trabajos enfocados en el dominio turístico como sistemas de recomendación, soporte para sistemas de decisión, entre otros, a continuación se describirán algunos de ellos.

TourExp (Buján, D., et al 2013) presenta un nuevo enfoque para parametrizar, modelar y compartir datos relacionado a un contexto para un visitante o turista en movimiento, de igual forma que una arquitectura para una plataforma para el manejo de información contextual basado en el modelo de datos contextual. Los dominios principales que maneja son el humano, ambiental y factores de sistema, los cuales han sido tomados en cuenta en orden de representar la información contextual acerca del turista en una localización y tiempo específico.

Los *checkins* del turista son usados para crear un mejor sistema de recomendación basado en datos contextuales. Todas las entidades de datos contextuales representados han sido modelados en una base de datos relacional que han sido mapeados a diferentes vocabularios. Estos datos contextuales pueden ser consultados por otros módulos del sistema y servicios de terceros por medio de servicios *Web RESTful* que regresan archivos *JSON* ó *XML*, o incluso usando un *SPARQL Endpoint* que regresan tripletas *RDF*, el cual permite incorporar la base de datos de TourExp a *Linked Data*.

SLOTD (*Swiss Linked Open Tourism Data*) La idea de este trabajo es convertir los datos estadísticos de turismo en Suiza y convertirla a formato *RDF* para realizar la conexión de los datos con otros repositorios de datos. Para los usuarios se les muestra una representación gráfica de los departamentos, los cuales poseen información adicional (población, numero de festivales musicales

entre otros), la cual sirve como base para futuras consultas. Esta información adicional es recuperada de diferentes fuentes de datos.

Los datos estadísticos están conectados a otros conjuntos de *Linked Data* como *DBpedia* y *GeoNames* para ser consultado, para la transformación de estos datos se basaron el esquema RDF en un RDF Data Cube además de diseñar otras propiedades que representan las dimensiones acerca del tiempo/posición/ocupamiento nocturno, para después realizar las conexiones hacia los repositorios y mostrar la información en una página web desarrollada en el lenguaje PHP y usando la cartografía de Google Maps para mostrar la información así como sus conexiones como se ve en la Figura 2.2.

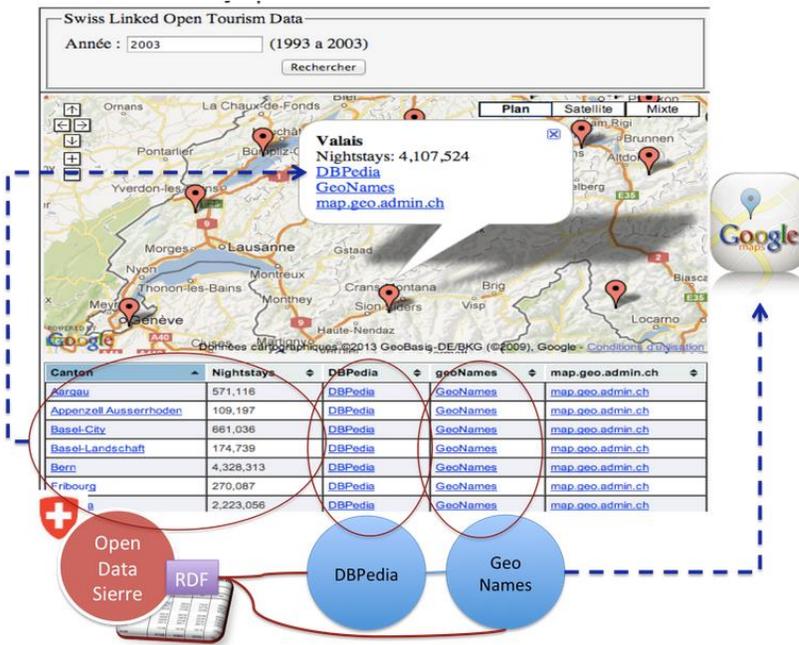


Figura 2.2 Sistema Web SLOTD

TouriMIS⁵.- El sistema TourMIS es una base de datos en la Web que consiste en datos sobre el estudio de mercado turístico como llegadas, ocupación hotelera y capacidades en varios países y ciudades (Wöber, K. 2003). El objetivo es tener datos comparables que ayuden a los gestores turísticos a realizar sus procesos de decisión (Wöber, K. 2003). Este trabajo utiliza los datos de la *National Tourism Statistics Austria*, *European Travel Commission (ETC)*, *European Cities Marketing (ECM)* y *Austrian National Tourist Office*, los cuales se aseguran que continúan el desarrollo y la población del sistema.

El sistema toma en cuenta cuatro grupos de usuarios: primero los representativos nacionales, provinciales, regionales y organizaciones turísticas de la ciudad, que están involucradas a largo plazo, planeamiento estratégico de desarrollos turísticos de la región; segundo proveedores turísticos como proveedores de alojamiento, comida, viaje, cultura, deportes así como agencias turísticas y operadores turísticos; tercero instituciones educativas activas en la investigación turística y por último consultores y autoridades e instituciones públicas. *TourMIS* contiene datos de tres indicadores importantes:

- Llegadas.- Número de turistas que llegan a varios tipos de alojamiento como hoteles, *bed & breakfast*, sitios de acampar.
- Noches.- Número de noches gastadas en varios tipos de alojamiento
- Capacidad.- Total de capacidad de camas en los alojamientos en el destino.

⁵ http://www.tourmis.info/index_e.html

2.4 Sistemas de recomendación semánticos

Existen trabajos [(Moreno et al.,2013), (Middleton, Roure, Shadbolt, 2009)], donde se propone utilizar el análisis semántico para dar una descripción formal y semántica sobre un objeto que se recomendará, permitiendo la desambiguación de la recomendación hacia una sola recomendación que puede provenir de diferentes sistemas de recomendación. Para esto se propone la incorporación de una ontología (una ontología es una especificación explícita de una conceptualización, de acuerdo con (Gruber, 1995)); las cuales se emplean actualmente en muchos trabajos como los de [(Moreno et al., 2013), (Golemati et al., 2007)] , para la modelación de un perfil de usuario.

Las ontologías también son utilizadas para enriquecer la descripción de los objetos a recomendar (Buriani et al., 2006), ya que les otorgan propiedades como precio, ubicación, nombre, etc., dejando de ser solo un artículo con un identificador en una base de datos a pasar a ser una instancia del concepto respetivo al que pertenece y del cual se busca sobre esas propiedades para permitir realizar una recomendación a los usuarios que busquen ese tipo de objetos.

Una limitante de la mayoría de este tipo de trabajos es que se queda únicamente con los datos, conceptos e instancias preestablecidos sin poder obtener mayor información aparte de la recopilada, descartando bastante información que existe en la Web que pudieran dale mayor detalle y a la vez enriquecer la información que se le va a proporciona al usuario. A continuación se describen algunos trabajo que emplean sistemas de recomendación y ontologías turísticas y *Linked Data*.

SigTur/E-Destination (Moreno et al., 2013). Este sistema está desarrollado para el funcionamiento en la Web, como se muestra en la Figura 2.3 y provee

recomendaciones de actividades en la región de Tarragona. Dentro del trabajo, las actividades están categorizadas en una ontología turística, la cual está basada en el tesoro de la Organización Mundial de Turismo (OWT), lo que proporciona un sentido semántico a la aplicación.

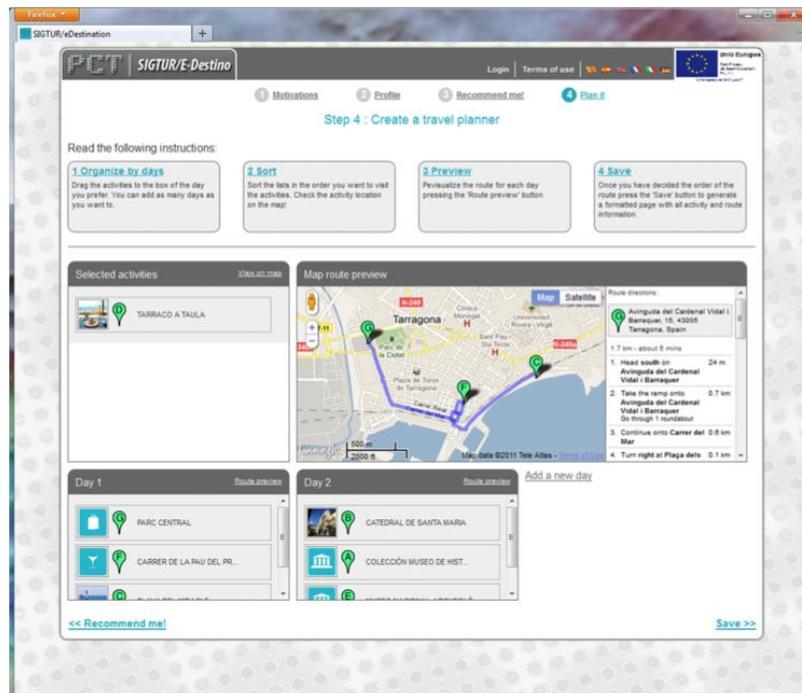


Figura 2.3. Diseño de la página SigTur/E-Destination (Moreno et al., 2013)

Para realizar recomendaciones, el sistema emplea un filtro colaborativo que recopila las evaluaciones de otros usuarios, esto aunado a varios métodos más ayudan a determinar qué actividades le gustarían al usuario, para ello se considera los siguientes tópicos que utiliza para desarrollar un plan de actividades para el usuario:

- La motivación del usuario.
- Las acciones del usuario acerca de las actividades disponibles.

- Las evaluaciones explícitas de las actividades realizadas.
- Una propagación *bottom-up* de interés y confianza.
- La similitud entre el usuario y un perfil predefinido.
- La similitud del usuario y un clúster de usuarios con características demográficas similares.
- La del usuario y un clúster de usuarios con evaluaciones similares.

DBpedia Mobile (Becker, C.; Bizer, C. 2008) Es un cliente de localización céntrica de *DBpedia* para dispositivos móviles como se ve en la Figura 2.4, la cual consiste de una cartografía base y un browser de *Linked Data* basado en Fresnel. El sistema permite al usuario acceder a la información acerca de los recursos de *DBpedia* en su proximidad, desde los cuales pueden acceder a otros recursos de la Web semántica mediante vínculos.



Figura 2.4 Mapa de *DBpedia Mobile* (Becker, C.; Bizer, C. 2008)

El mapa que muestra está formado por tripletas *RDF* obtenidas de enviar el área actual al igual que el idioma y los ponderaciones del filtrado al servidor, las cuales son reescritas como una consulta de *SPARQL*, la cual le llega a un servidor Virtuoso que contiene la información geoespacial de *DBpedia*.

SMARTMUSEUM (Ruotsalo et al., 2013).- Es un sistema de recomendación móvil y de plataforma Web, como se muestra en la Figura 2.5, el cual sugiere sitios culturales y artísticos para los usuarios, aparte de proporcionar una breve descripción de los mismos y contenido multimedia como videos e imágenes.

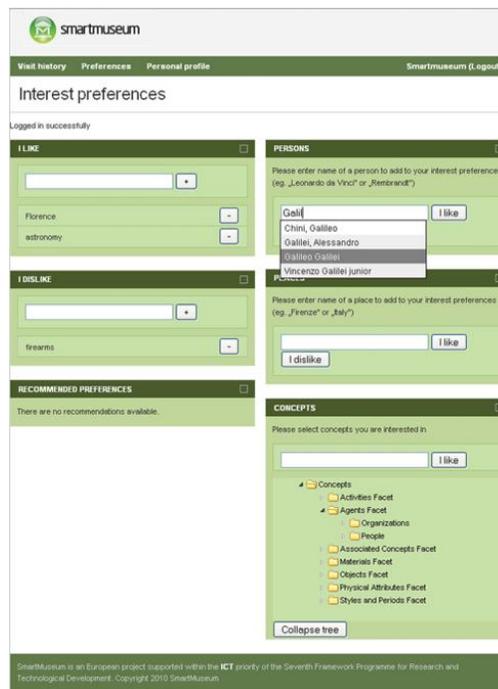


Figura 2.5. Diseño Web de SMARTMUSEUM (Ruotsalo et al., 2013)

El sistema emplea un filtro basado en contenido, realiza cómputo ubicuo y utiliza lenguajes de Web Semántica para la representación de los datos, en este caso se emplea la ontología para que las descripciones de contenido heterogéneo de los sitios converjan, así como entradas de GPS para ubicar al usuario en exteriores y RFID para obtener la posición del usuario que se encuentren dentro de algún edificio. Se considera además la implementación de perfiles de usuario,

utiliza *frameworks* para la recuperación de información en el que los datos contextuales y la búsqueda de datos agrupados y clasificados son utilizados para llevar a cabo la generación de recomendaciones con contenido útil para los usuarios.

Sem-Fit (García et al., 2011). Éste es un sistema de recomendación semántico para la industria hotelera. La aplicación hace uso de la lógica difusa para realizar las sugerencias. El *kernel* del trabajo es considerar el paradigma de la lógica difusa, el cual define un conjunto de teorías que proveen marcos de trabajo para la representación de muchos aspectos inciertos del conocimiento humano (Zadeh, 1965). De esta manera, el sistema recopila la información en las recomendaciones de hoteles, a partir de:

- La descripción semántica de los hoteles.
- Las relaciones difusas entre las características de los comensales y las características de los hoteles.
- Los sentimientos del comensal acerca de las recomendaciones realizadas.

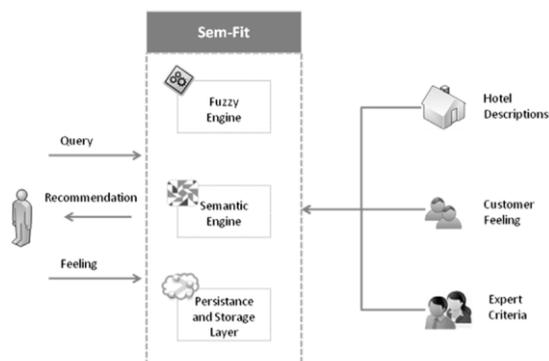


Figura 2.6. Diseño del Sistema Sem-Fit (García et al., 2011)

Como se muestra en la Figura 2.6 el proceso general de funcionamiento del sistema se divide en los siguientes pasos:

1. Obtener las características del usuario (información personal, estado de ánimo, gustos, etc.).
2. Fusificación los valores proporcionados por el usuario se convierten en conjuntos difusos definidos según un experto para describir las características del usuario.
3. Cuando ya se cuenta con el perfil de usuario difuso, se procede a evaluar las reglas difusas, esto con el fin de obtener el conjunto de valores difusos de las características del hotel, a la vez que se define el nivel de adecuación basado en un *grid* de afecto.
4. Los resultados obtenidos son defusificados para obtener el conjunto de características para los hoteles.
5. Con los resultados del paso 4, los hoteles que tengan las mismas características del conjunto obtenido, son recuperados basado en la ontología de hoteles.
6. La siguiente recomendación del sistema calculará los pesos de cada hotel, basado en los valores de idoneidad obtenidos desde la matriz de decisión.
7. El hotel que obtenga el mayor peso se recomendará como principal; además de proporcionar otros sitios, los cuales serán mostrados al usuario como hoteles sugeridos.

SPETA (García et al., 2009). Es un sistema de recomendación diseñado para dispositivos móviles enfocado a servicios turísticos (ver Figura 2.7). Éste implementa el uso de un filtro híbrido, así como el uso de una ontología turística para dar un contexto semántico a las recomendaciones. También emplea el uso de redes sociales para recopilar información implícita del perfil del usuario y poder recabar la información necesaria para su posterior análisis, tomando en cuenta sus gustos y lugares visitados, así como los de sus amigos. El filtro híbrido se divide en las siguientes partes:

- La primera parte del filtro utiliza la información contextual como el tiempo, el clima y la localización del usuario para determinar qué servicios se pueden ofrecer a los usuarios.
- La segunda parte del filtro utiliza técnicas basadas en conocimiento, con esto podemos dar paso al uso de una ontología turística, la cual modelará al usuario y los servicios que se puedan ofrecer, con lo que proporciona el uso de una medida de similitud, basada en características para conocer el grado de similitud entre las preferencias del usuario y los servicios.
- La tercera parte del filtro recopila preferencias de los usuarios, basándose en su red social, determinando gustos como libros, o tipo de comida que suele visitar, además de tomar en cuenta si estos lugares fueron o serán visitados por los amigos del usuario.

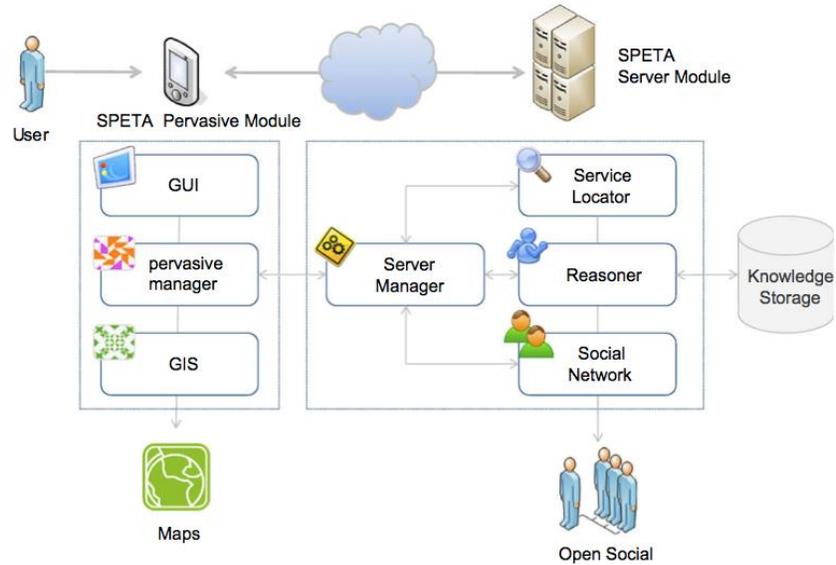


Figura 2.7. Arquitectura del sistema de recomendación SPETA (García et al., 2009)

More! (Parra, G.; Klerkx, J.; Duval, E. 2010) Es una aplicación móvil la cual tiene como propósito dar a conocer los artículos y trabajos publicados, compartir material que se encuentre en la Web y hacer uso de las redes sociales para contactar a los investigadores o expositores durante un congreso como se ve en la Figura 2.8, esto se logra mediante un código QR⁶, este código estará expuesto durante la presentación y los asistentes podrán decodificarlo con cualquier aplicación comercial y recuperar la URL que redirige al usuario al sitio web, la cual desplegará la información del expositor a la audiencia.

⁶ <http://www.qrcode.com/en/>

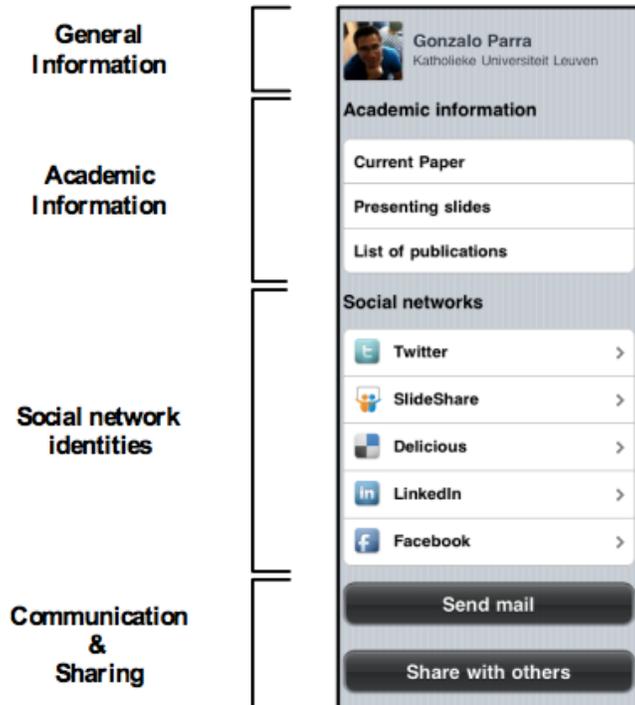


Figura 2.8 Vista general de la aplicación More! (Parra, G.; Klerkx, J.; Duval, E. 2010)

Para poder recuperar información de la Web 2.0 se desarrolló una *API* llamada “Reserch.fm” la cual almacena información sobre los investigadores y expositores, en orden de interconectar las fuentes de información por medio de *Linked Data*, todo esto basado en la ontología *SWRC (Semántica Web for Research Communities)* (Sure, Y., et al 2005), la cual permite la manipulación y vinculación de los recursos, en orden para recuperar toda la información útil se extendieron las ontologías *SWRC* con la *FOAF*⁷ (Friend of a Friend), *SIOC*⁸ (*Semantic-Interlinked Online Communities*) y *vCard*⁹.

⁷ <http://www.foaf-project.org/>

⁸ <http://sioc-project.org/>

⁹ <http://www.w3.org/Submission/vcard-rdf>

wayUO (Zablith, F.; Fernandez, M.; Rowe.M, 2011) Es una aplicación para dispositivos móviles con sistema operativo *Android*, el cual propone no solo emplear *Linked Data* para hacer uso de datos de la Web, sino para crear nuevas conexiones para nutrir la información disponible, entre ellas incluye las relaciones entre personas y lugares, algo similar a lo q realiza la red social *Foursquare*¹⁰.

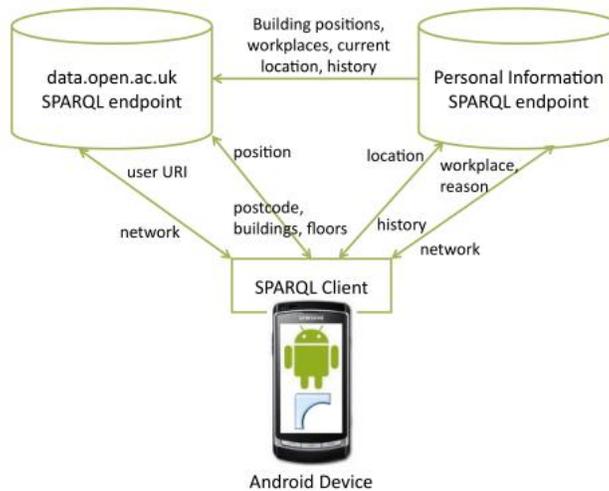


Figura 2.9 Arquitectura del sistema wayUO (Zablith, F.; Fernandez, M.; Rowe.M, 2011)

Como se ve en la Figura 2.9, para realizar todo el proceso, se asigna una *URL* única por usuario por medio de la red de la universidad, e implementando un pequeño formulario para asociar la ubicación del usuario con la actividad que realiza, se genera un historial el cual es usado para conocer las ubicaciones frecuentadas por los usuario o la ubicación del despacho de algún investigador; El repositorio consta de dos fuentes de datos, la primera es la información que provee la universidad desde su *SPARQL Endpoint* , el cual contiene información sobre lugares, edificios, pisos de la universidad, así como información básica de los miembros de la universidad; la segunda es un servicio el cual actualiza un *RDF* que contiene las actividades e información de los usuarios.

¹⁰ <http://www.w3.org/Submission/vcard-rdf>

2.5 Conclusiones

De los trabajos revisados para formar el estado del arte, son pocos los que integran un bloque o componente semántico para generar las recomendaciones. La mayoría de los trabajos sólo se enfocan en utilizar filtros, ya sean basados en contenido, en aspectos demográficos o colaborativos para discriminar y saber si el objeto a sugerir realmente es útil al usuario. Varios trabajos solo incorporan una jerarquía para poder clasificar los objetos que ofertan y tener un mayor control en ellos.

No obstante, existen sistemas de recomendación que incorporan como base las ontologías de aplicación, del dominio turístico y de e-Commerce. Estos sistemas permiten brindar una mayor información semántica y controlada a los objetos, más que un simple identificador común de una base de datos. Sin embargo, no todos estos trabajos implementan los principios de *Linked Data* para recuperar información de la Web y poder ofrecer al usuario mayor información la cual este localizada en otras fuentes de datos, e incluso poder incorporar nuevos sitios que no estén en el repositorio original de la aplicación.

A la vez no todos los trabajo implementan una comunicación a las redes sociales para poder extraer una evaluación conforme a la opinión de los usuarios, de los trabajos que usan filtros colaborativos, la mayoría requiere que exista un repositorio de decisiones y perfiles de usuarios almacenados para realizar las comparaciones necearías entre el usuario actual y el histórico, con tal de dar una mejor recomendación.

Capítulo 3 Marco Teórico

En este capítulo se describirá los conceptos y definiciones necesarias para dar sustento a la metodología que se propone en este trabajo.

3.1 Ontología

Las ontologías se han empleado desde hace varios siglos en el campo de la filosofía y del conocimiento, el primero en dar una definición fue Aristóteles diciendo que una ontología era una explicación sistemática de la existencia. En tiempos recientes el campo de la inteligencia artificial ha proporcionado varias definiciones de una ontología como:

- "Una ontología define los términos y relaciones básicas que componen el vocabulario de un área temática, así como las reglas para combinar términos y relaciones para definir la extensión del vocabulario".(Neches, 1991).
- "Una ontología es una especificación explícita de una conceptualización". (Gruber, 1995).
- "Una ontología es una especificación formal de una conceptualización compartida". (Borst, 1997).

Las ontologías están desarrolladas bajo un dominio de discurso y un marco teórico que define un vocabulario del cual se derivan los conceptos, relaciones,

instancias, constantes, atributos, axiomas, y reglas estos conceptos se describirán a continuación:

- **Conceptos.**- Son objetos o entidades, considerados desde un punto de vista amplio. Los conceptos de una ontología están normalmente organizados en taxonomías en las cuales se pueden aplicar mecanismos de herencia.
- **Relaciones.**- Las relaciones representan un tipo de asociación entre conceptos del dominio. Si la relación une dos conceptos se denomina relación binaria. Una relación binaria relevante es Subclase-de, que se utiliza para construir taxonomías de clase, como se ha especificado anteriormente.
- **Instancia.**- Se utilizan para representar individuos en la ontología. Las relaciones también se pueden instanciar.
- **Constantes.**- Son valores numéricos que no cambian en un largo período de tiempo.
- **Atributos.**- Los atributos describen propiedades, se pueden distinguir dos tipos de atributos de instancia y de clase.
 - Los atributos de instancia describen propiedades de las instancias de los conceptos, en las cuales toman su valor, éstos se definen en un concepto y se heredan a sus subconceptos e instancias.
 - Los atributos de clase describen conceptos y toman su valor en el concepto en el cual se definen. Estos atributos no se heredan ni a los subconceptos ni a las instancias.
- **Axiomas.**- Los axiomas son expresiones lógicas siempre verdaderas que suelen utilizarse para definir restricciones en la ontología.
- **Reglas.**- Las reglas se utilizan normalmente para inferir conocimientos en la

ontología, tales como valores de atributos, instancias de relaciones, etc. [(Gruber, 1995), (Corcho et al., 2005)].

Por otro lado las ontologías se pueden categorizar con base en el grado de formalidad que exista en ella:

- Alta informalidad.- Las ontologías están expresadas en lenguajes natural, un ejemplo de ello son los glosarios.
- Semi-informalidad.- Estas ontologías estas estructuradas y restringidas por el lenguaje natural y son usadas en orden de reducir la ambigüedad de éstas.
- Semi-formal.- Estas ontologías están expresadas en un lenguaje artificial definido formalmente como los lenguajes de marco.
- Formalmente riguroso.- Son las ontologías que están precisamente definidas con semántica formal, ejemplo de ello son los lenguajes basados en lógica [(Uschold, Gruninger, 1996), (Roche, 2003)].

También éstas pueden ser clasificadas de acuerdo al tipo de conocimiento que será transmitido por la ontología:

- **Ontologías genéricas.**-Estas ontologías cubren conceptos generales definidos independientemente del dominio de la aplicación y que puede ser usada en varios dominios.
- **Ontologías de dominio.**- Estas ontologías están especificadas para un dominio en particular y cubren conceptos genéricos del mismo, lo que permite que sean

reutilizadas en tareas diferentes que están relacionadas con el dominio de discurso.

- **Ontologías de aplicación.**- Estas ontologías usan el conocimiento específico para una tarea en particular, que incluye conocimiento específico de expertos para la aplicación, por lo general estas ontologías no se pueden reutilizar.
- **Meta-ontología.**- Esta ontología especifica la representación del conocimiento usada para definir los conceptos del dominio y ontologías genéricas [(Roche, 2003), (Guarino, 1995)].

Para el diseño de una ontología existen varios criterios que se deben tener en cuenta:

- **Claridad y objetividad.**- La ontología debe proporcionar el significado de términos definidos proporcionando definiciones objetivas y documentadas en el lenguaje natural.
- **Complejidad.**- Una definición expresada por una condición necesaria y suficiente es preferida por una definición parcial.
- **Coherencia.**- Permite que se puedan realizar inferencias y que éstas sean consistentes con las definiciones ya preestablecidas.
- **Maximiza la extensibilidad monotónica.**- Los términos generales nuevos o especializados deben incluirse en la ontología de tal forma que no requiera revisión de definiciones existentes.

- **Mínimo compromiso ontológico.**- Hace pocas afirmaciones acerca del mundo a ser modelado, lo cual significa que la ontología debe ser específica tanto como sea posible el significado de sus términos , dando libertad a la ontología para especializar e instanciar.
- **Principio de distinción ontológica.**- Las clases de una ontología deben ser disjuntas. El criterio utilizado para aislar las propiedades principales consideradas a ser invariantes para una instancia de una clase se llama criterio de identidad.
- **Diversificación de jerarquías.**- Si el conocimiento es suficiente es representado en la ontología y con muchas formas o criterios de clasificación, para que facilite el introducir nuevos conceptos y heredad propiedades de diferentes puntos de vista.
- **Modularidad.**- Minimiza el acoplamiento entre módulos.
- **Minimizar la distancia semántica entre conceptos hermanos.**- Los conceptos similares son agrupados y representados como subclases de 1 clase y deben definirse utilizando las mismas primitivas, mientras que los conceptos menos similares son separados en la jerarquía.
- Estandarización de nombres para evitar inconsistencias en la ontología, así como la confusión al momento de realizar inferencias en la misma (Buriano et al., 2006).

3.2. Metodologías

Después de describir los componentes de una ontología, ahora se hablará acerca de las diferentes metodologías que existen para su diseño e implementación como se describe en el trabajo de (Guzmán, López, Torres, 2012).

CYC.- Publicada por Lenat y Guha desde 1990, en la que divulgaron algunos pasos generales para la construcción de ontologías; el primero consiste en extraer manualmente el conocimiento común que está implícito en diferentes fuentes para después cuando se tenga suficiente conocimiento en la ontología adquirir nuevo conocimiento común usando herramientas del procesamiento de lenguaje natural o aprendizaje computacional.

Uschold y King.- Publicada por Uschold y King en 1995, empleado en el Modelo Enterprise, donde recrean una serie de pasos que permiten plasmar y especificar los conocimientos que se tiene sobre un dominio específico, centrando sus esfuerzos en la forma en la cual representar conocimientos. Los pasos que propone la metodología son:

- Identificar propósito.
- Capturar los conceptos y relaciones entre estos conceptos y los términos utilizados para referirse a estos conceptos y relaciones.
- Codificar la ontología.

Grüninger y Fox.- Publicada por Grüninger y Fox en 1995, fue desarrollada paralelamente de la metodología del Uschold y King y se usó para construir las ontologías del proyecto *TOVE* (Toronto Virtual Enterprise); EL primer paso consiste en identificar intuitivamente las aplicaciones posibles en la que se usara la ontología. Luego se utilizan un conjunto de preguntas en lenguaje natural, llamadas cuestiones de competencia, para determinar el ámbito de la ontología, Se usan estas preguntas para extraer los conceptos principales, sus propiedades, relaciones y axiomas, los cuales se definen formalmente como *Prolog*.

Kactus.- Surge en 1996 usando el dominio de las redes eléctricas para desarrollar ontologías como parte del proyecto *Spirit KACTUS*. Esta metodología usa una base de conocimiento por medio de un proceso de abstracción. Los pasos son los siguientes:

- Especificación de la aplicación
- Diseño preliminar basado en categorías ontológicas top-level relevantes
- Refinamiento u estructuración de la ontología.

Methontology.- Esta metodología es desarrollada por (Fernández, Gómez, Juristo, 1997), siendo la propuesta más completa debido a que toma la creación de ontologías como un proyecto informático. Así, además de las actividades propias de la generación de la ontología esta metodología abarca actividades para la planificación del proyecto, la calidad del resulta, la documentación, etc. La metodología está basada en once tareas como se ve en la Figura 3.1, a continuación se describen como en el trabajo de (Corcho et al., 2005):

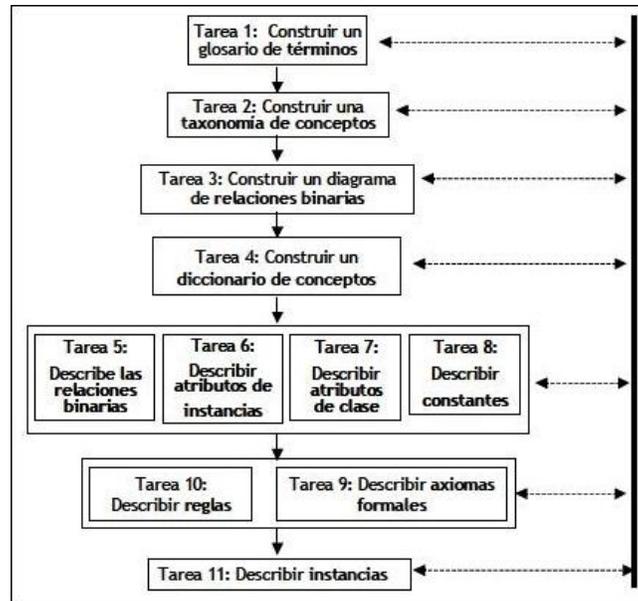


Figura 3.1.- Tareas que componen la metodología METHONTOLOGY(Corcho et al., 2005).

Tarea 1: Construir el glosario de términos. En primer lugar, el desarrollador de la ontología construye un glosario de términos que incluye todos los términos relevantes del dominio (conceptos, instancias, atributos, relaciones entre conceptos, etc.), sus descripciones en lenguaje natural, y sus sinónimos y acrónimos.

Tarea 2: Construir taxonomías de conceptos. Una vez que el glosario de términos contiene suficientes términos, el desarrollador de la ontología construye las taxonomías de conceptos que definen su jerarquía. Para construir taxonomías de conceptos, se seleccionan del glosario de términos aquellos términos que son conceptos.

Tarea 3: Construir diagramas de relaciones binarias ad hoc. Una vez construida y evaluada la taxonomía, la actividad de conceptualización propone construir diagramas de relaciones binarias ad hoc. El objetivo de este diagrama es

establecer las relaciones ad hoc existentes entre conceptos de la misma o de distintas taxonomías de conceptos.

Tarea 4: Construir el diccionario de conceptos. Una vez que las taxonomías de conceptos y los diagramas de relaciones binarias ad hoc han sido generados, el desarrollador de la ontología debe especificar cuáles son las propiedades que describen cada concepto de la taxonomía, así como las relaciones identificadas en el diagrama del paso anterior y las instancias de cada uno de los conceptos. El diccionario de conceptos contiene todos los conceptos del dominio, sus relaciones, sus instancias, y sus atributos de clase y de instancia.

Las relaciones especificadas para cada concepto son aquellas en las que el concepto es el origen de la misma.

Tarea 5: Describir las relaciones binarias ad hoc. El objetivo de esta tarea es describir en detalle todas las relaciones binarias ad hoc identificadas en el diagrama de relaciones binarias e incluidas en el diccionario de conceptos. Para cada relación binaria ad hoc, el desarrollador de la ontología debe especificar su nombre, los nombres de sus conceptos origen y destino, su cardinalidad y su relación inversa, si existe.

Tarea 6: Describir los atributos de instancia. El objetivo de esta tarea es describir en detalle todos los atributos de instancia incluidos en el diccionario de conceptos. Cada fila de la tabla de atributos de instancia contiene la descripción detallada de un atributo de instancia.

Por cada atributo de instancia, el desarrollador de la ontología debe especificar su nombre, el concepto al que pertenece el atributo (los atributos son locales a los conceptos), su tipo de valor, su rango de valores (en el caso de atributos numéricos) y sus cardinalidades mínima y máxima.

Tarea 7: Describir los atributos de clase. El objetivo de esta tarea es describir en detalle todos los atributos de clase incluidos en el diccionario de conceptos. Para cada atributo de clase, el desarrollador de la ontología debe rellenar la siguiente información: nombre del atributo, nombre del concepto donde el atributo se define, tipo de valor, cardinalidad y valor(es).

Tarea 8: Describir las constantes. El objetivo de esta tarea es describir en detalle cada una de las constantes identificadas en el glosario de términos. Para cada constante, el desarrollador de la ontología debe especificar su nombre, tipo de valor, valor y unidad de medida en el caso de constantes numéricas.

Tarea 9: Definir axiomas formales. Para realizar esta tarea, el desarrollador de ontologías debe identificar los axiomas formales que son necesarios en la ontología y describirlos de manera precisa. Para cada definición de axioma formal, se propone especificar la siguiente información: nombre, descripción en lenguaje natural, expresión lógica que define de manera formal el axioma usando lógica de primer orden, y los conceptos, atributos y relaciones ad hoc utilizadas en el axioma, así como las variables utilizadas.

Tarea 10: Definir reglas. De manera similar a la tarea previa, en esta tarea el desarrollador de la ontología debe identificar en primer lugar qué reglas se necesitan en la ontología, y entonces describirlas en la tabla de reglas. Para cada regla, se propone incluir la siguiente información: nombre, descripción en lenguaje

natural, expresión que describe formalmente la regla, y conceptos, atributos y relaciones ad hoc utilizados en la regla, así como las variables usadas.

METHONTOLOGY propone especificar las expresiones de las reglas utilizando el formato si <condiciones> entonces <consecuente>. La parte izquierda de la regla es una conjunción de condiciones simples, mientras que la parte derecha es una simple expresión de un valor de la ontología.

Tarea 11: Describir instancias. Una vez que el modelo conceptual de la ontología ha sido creado, se pueden definir las instancias que aparecen en el diccionario de conceptos. Para cada instancia se define: su nombre, el nombre del concepto al que pertenece y los valores de sus atributos de instancia, si se conocen.

Sensus.- Surge en 1997 como nuevo método para construir ontologías basado en la ontología de *SENSUS*, la cual constituye un enfoque *top-down* para derivar ontologías específicas del dominio a partir de grandes ontologías. En ésta se identifican un conjunto de términos semilla que son relevantes para el dominio particular.

Tales términos se enlazan manualmente a una ontología de amplia cobertura, con lo cual el usuario puede seleccionar los términos más relevante y así acotar la ontología *Sensus*. Después, el algoritmo devuelve un conjunto de términos estructurados jerárquicamente para describir un dominio, que puede ser usado como para la base de conocimiento.

ON-TO- KNOWLEDGE.- Desarrollada por (Sure, Staab, Studer, 2003), ésta aplica ontologías a la información disponible electrónicamente para mejorar la calidad de la gestión de conocimiento en organizaciones grandes y distribuidas. Además, incluye la identificación de metas que deberían ser conseguidas por herramientas de gestión de conocimiento y está basada en el análisis de escenarios de uso y en los diferentes papeles desempeñados por trabajadores de conocimiento y accionistas en las organizaciones.

NeOn.- Diseñada por (Suárez, Gómez, Fernández, 2012), esta metodología diseñada para la construcción de redes de ontologías está basada en escenarios que se apoya en los aspectos de colaboración de desarrollo de ontologías y la reutilización, así como en la evolución dinámica de las redes de ontologías en entornos distribuidos. Las claves de la Metodología *NeOn* son:

Un conjunto de nueve escenarios para la construcción de ontologías como se ve en la Figura 3.2 y redes de ontologías, haciendo hincapié en la reutilización de los recursos ontológicos y no ontológicos, la reingeniería y la fusión, y teniendo en cuenta la colaboración y el dinamismo. El Glosario de Procesos y Actividades identifica y define aquellos procesos y actividades involucrados en el desarrollo de redes de ontologías.

Directrices metodológicas para diferentes procesos y actividades del proceso de desarrollo de la ontología de la red, tales como la reutilización y la reingeniería de los recursos ontológicos y no ontológicos, la especificación de los requisitos de la ontología, la localización de la ontología, la programación, etc. Todos los procesos y actividades se describen con (a) una tarjeta llena, (b) un flujo de trabajo, y (c) ejemplo.

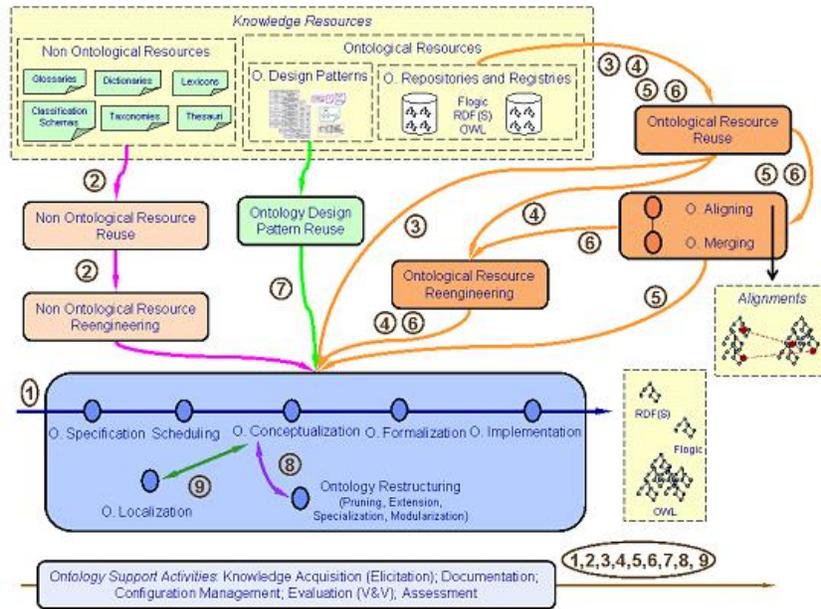


Figura 3.2.- Escenarios para construcción de ontologías y ontologías de red (Suárez et al., 2012)

Escenario 1: Desde la especificación de la aplicación. La red de ontologías se desarrolla a partir de cero (sin volver a utilizar los recursos existentes). Los desarrolladores deben especificar los requisitos de la ontología. Después de eso, se asesora para llevar a cabo una búsqueda de recursos potenciales para ser reutilizados. A continuación, la actividad de planificación se debe realizar, y los desarrolladores deben seguir el plan.

Escenario 2: La reutilización y reingeniería de los recursos no ontológicos (NOR). Los desarrolladores deben llevar a cabo el proceso de reutilización NOR para decidir, de acuerdo con los requisitos de la ontología, que NORs pueden ser reutilizados para construir la red de la ontología. A continuación, los NORs seleccionados deben ser volver al proceso de re-ingeniería ontológicas.

Escenario 3: La reutilización de los recursos ontológicos. Los desarrolladores utilizan recursos ontológicos (ontologías como un conjunto de módulos ontológicos, y/o declaraciones) para construir redes de ontologías.

Escenario 4: La reutilización y re-ingeniería de los recursos ontológicos. Los desarrolladores de ontologías reutilizan los recursos y reorganizar los recursos ontológicos.

Escenario 5: La reutilización y la fusión de los recursos ontológicos. Este escenario se produce cuando varios recursos ontológicos en el mismo dominio que se seleccionan para su reutilización, y los desarrolladores desean crear un nuevo recurso ontológico con los recursos seleccionados.

Escenario 6: Reutilización, la fusión y re-ingeniería de los recursos ontológicos. Los desarrolladores de ontologías reutilizan, combinan y reorganizan los recursos-ontológicos. Este escenario es similar al Escenario 5, pero en este caso los desarrolladores deciden reorganizar el conjunto de recursos combinados.

Escenario 7: Reutilización de los patrones de diseño de ontologías (ODPs). Los desarrolladores de ontologías acceden a repositorios de reutilización ODPs.

Escenario 8: Reestructuración de recursos ontológicos. Los desarrolladores de ontologías reestructuran (modularizan, podan, extienden y / o especializan) recursos ontológicos que deben integrarse posteriormente en la red de ontologías.

Escenario 9: Localización de recursos ontológicos. Los desarrolladores de ontologías adaptan una ontología a otras lenguas y la cultura las comunidades, obteniendo así una ontología multilingüe.

3.3 Linked Data

El termino *Linked Data* se refiere al conjunto de mejores prácticas para publicar e interconectar estructuras de datos en la Web. Este universo está constituido por varias temáticas y dominios como se ve en la Figura 3.3

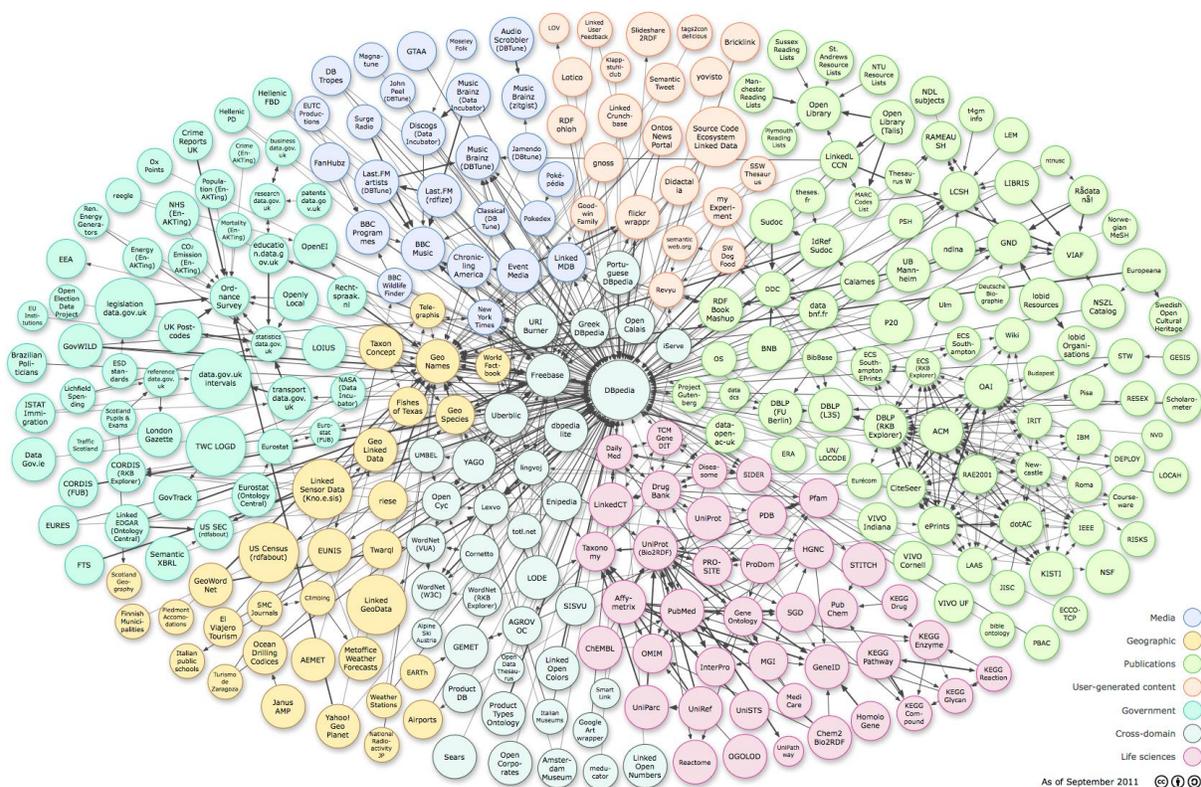


Figura 3.3.- Último diagrama del estado de *Linked Data*

A partir de estos dominios se pueden hacer relaciones y consultas sobre algún concepto y sabes si está vinculado en algún otro dominio; Por lo tanto a continuación se mencionan y describen los principios a seguir para incorporar *Linked Data*:

- Usar *URIs* como nombres para las cosas; Esto se refiere el usar las *URI* no solo como referencias para identificar únicamente en la Web sino también en el mundo real, estos conceptos pueden incluir objetos tangibles como personas o lugares y algunos pueden ser más abstractos como relaciones tipo alguien conoce a alguien más.

Este principio se puede ver como una extensión de la web desde los recursos en línea hacia cualquier objeto o concepto en este mundo.

- Usar *HTTP URIs* para que las personas puedan buscar esos nombres; Esto se refiere a usar *HTTP URIs* para identificar objetos y conceptos abstractos, habilitando estas *URIs* para ser dereferenciables sobre el protocolo *HTTP* para una descripción del objeto o concepto identificado.
- Cuando alguien busca una *URI*, proveer información útil usando los estándares (*RDF*, *SPARQL*); Esto se refiere al uso de un modelo simple de datos para publicar la estructura de datos en la Web en el Marco de Descripción de Recursos (*RDF*) el cual es un modelo de datos simple basado en grafos que está diseñado para el uso de contexto en la Web (Klyne, Carroll 2004).
- Incluir relaciones hacia otras *URIs*, para que las personas puedan obtener mayor información sobre alguna cosa. Esto se refiere al uso de hipervínculos para conectar no solamente documentos de la Web, sino cualquier tipo de cosa. Por ejemplo tener un hipervínculo que conecte personas con lugares ó lugares con compañías.

Estos hipervínculos son llamados *RDF links* de tal manera que se distinguen del enfoque clásico.

Estos principios nos permiten que a través de diferentes servidores se posible responder las peticiones de referencia *HTTP URIs* con diferentes nombres y regresar descripciones en *RDF* para la identificación de los recursos de las *URIs* proporcionadas, esto da la posibilidad que con el contexto de *Linked Data* si un *RDF link* conecta con algunas *URIs* diferentes, este se capaz de conectarse a esos conjuntos de datos remotos.

Linked Data está basado en la arquitectura general de la *WWW*, la cual realiza tareas para compartir estructuras de datos a escala mundial. Para esto se debe entender la arquitectura clásica de un documento Web. El documento Web está constituido en un conjunto de estándares como:

- Identificador de Recursos Uniforme (*URIs*) como un mecanismo global de identificación única (Berners et al 1998).
- Protocolo de Transferencia de Hipertexto (*HTTP*) como un mecanismo de acceso universal (Fielding 1999).
- Lenguaje de Marcado de Hipertexto (*HTML*) como un formato de contenido (Raggett 1999).

La Web está constituida en la idea de poner hipervínculos entre documentos web que pueden residir en diferentes servidores. Siendo éstos capaces de permitir a los usuario navegar entre diferentes servidores, con lo que permite el navegar

entre *URIs* para poder proporcionarle al usuario que consulta la capacidad de ver información que se encuentra en otro servidor.

3.4 RDF

*RDF*¹¹ es un modelo estándar para el intercambio de datos en la Web, el cual posee características que facilitan la fusión entre diferentes esquemas, y tiene soporte para la evolución de esquemas sobre el tiempo sin que se requiera que todos los datos del consumidor sean cambiados.

También extiende la estructura de vinculación de la Web para usar *URIs* en los nombres de las relaciones entre los objetos, que sirve para relacionar dos objetos y formar una tripleta. Usando este simple modelo nos permite estructurar y semi-estructurar datos para ser mezclados, expuestos o compartido desde diferentes aplicaciones.

La estructura de vinculación forma directa, por un grafo etiquetado donde los bordes representas los links nombrados entre dos recursos, representado por los grafos de un nodo. Donde cada nodo está conformado por tres cosas:

- **Sujeto.**- Este puede ser el nodo inicial, una instancia, una entidad o una característica.
- **Predicado.**- Este puede ser un verbo, una propiedad, un atributo, una relación, un miembro, un enlace o una referencia.

¹¹ <http://www.w3.org/RDF/>

- **Objeto.**- Este puede ser un valor, un nodo final o algún valor no literal que pueda ser usado como un sujeto.

Cualquiera de estos tres valores en una simple tripleta como se ve en la Figura 3.4, esta puede ser representada vía una *URI*. Los sujetos y objetos son llamados nodos y pueden ser representados como un nodo en blanco. Los objetos pueden también ser representado como un valor literal. También un mismo nodo puede jugar el rol de sujeto en una arista, y en otras ser un objeto.

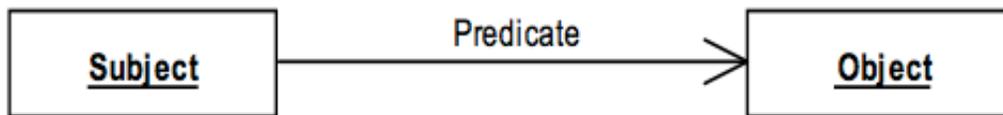


Figura 3.4.- Tripleta en *RDF*

3.5 SPARQL

SPARQL es un acrónimo recursivo de pie del *SPARQL Protocol* y *RDF Query Language*, por lo tanto este término es usado para hacer referencia al lenguaje de consultas para *RDF* así como un protocolo con especificaciones para realizar consultas remotas.

Para el primer uso *SPARQL Language*¹² es un lenguaje basado en sintácticamente que es utilizado para realizar consultas a grafos en *RDF* a través de patrones de coincidencias. Este lenguaje da posibilidad de incluir patrones básicos conjuntivos, filtros por valores, patrones opcionales y patrones de disyunción .

¹² <http://www.w3.org/TR/rdf-sparql-query/>

Para el segundo uso (Protocolo) se define como un método para invocar remotamente *consultas* de *SPARQL*. El protocolo *SPARQL*¹³ se puede describir de dos maneras: primero, como una interfaz independiente abstracta de cualquier realización concreta, aplicación, o unirse a otros protocolos; la segunda como *HTTP* o *SOAP* el cual se puede unir a dicha interfaz.

El protocolo *SPARQL* contiene una interfaz, un *SparqlQuery* que contiene una operación de *query*. Este protocolo descrito abstractamente con *WSDL 2.0* en términos de servicio web que implementa una interfaz, tipos, faltas y operaciones, a la vez como *HTTP* y *SOAP* relacionados.

Para realizar consultas al protocolo *SPARQL* se necesita implementar un *Endpoint*., este es un servicio del protocolo *SPARQL* el cual está definido conforme al estándar *SPROT*. Este permite al usuario realizar consultas a una base de conocimiento vía el lenguaje *SPARQL*. Resulta típico regresar en uno o más formatos la respuesta a la consulta realizada en el *Endpoint*, por lo tanto, un *SPARQL Endpoint* está concebido como una interfaz amigable hacia una base de conocimiento.

Ambas formulaciones de las consultas y la presentación de resultados humanamente legibles, debe ser implementada y llamada en software a la vez de no ser un proceso manual hecho por humanos, estas consultas utilizan como base para la consulta de archivos *RDF*, los cuales están almacenados en un *triple store*.

Un *triple store* es una base de datos especialmente diseñada para el almacenamiento y recuperación de tripletas, siendo una tripleta una entidad de datos compuesta de sujeto-predicado-objeto, como "Alejandro tiene 35" o "Alejandro conoce Oscar". Al igual que una base de datos relacional, uno

¹³ <http://www.w3.org/TR/rdf-sparql-protocol/>

almacena la información en un *triple store* y lo recupera a través de un lenguaje de consulta.

A diferencia de una base de datos relacional, *un triple store* está optimizado para el almacenamiento y recuperación de triples. Además de las consultas, las tripletas usualmente se pueden importar ó exportar utilizando el formato *RDF* ó en algún otro formato como *HTML* o *JSON* (Ilube, 2007).

3.6 GeoSPARQL

*GeoSPARQL*¹⁴ es un estándar de la OGC, el cual define una extensión de funciones para *SPARQL* [W3r], un conjunto de reglas *RIF*¹⁵ y un núcleo *RDF/OWL*¹⁶ vocabulario para información geográfica basada en el *General Feature Model* (ISO 19109), *Simple Features* [ISO 19125-1], *Feature Geometry* (ISO 19107) y *SQL MM* (ISO/IEC 13249-3).

El estándar *GeoSPARQL* de OGC representa soporte y consultas de datos geoespaciales en la Web semántica. *GeoSPARQL* define un vocabulario para representar datos en RDF, y define una extensión para el lenguaje de consulta SPARQL para poder procesar datos geoespaciales.

El estándar sigue un diseño modular; este comprende diferentes componentes, los cuales se enumeran y describen a continuación:

- Un *core* que define una clase *top-level* de *RDFS/OWL* para objetos espaciales.

¹⁴ <http://www.opengeospatial.org/standards/geosparql>

¹⁵ <http://www.w3.org/TR/rif-core/>

¹⁶ <http://www.w3.org/RDF/>

- Un vocabulario topológico que define propiedades *RDF* para afirmar y consultar las relaciones topológicas entre objetos espaciales.
- Una geometría que define tipos de datos *RDFS* para serializar datos geométricos, propiedades *RDF* geoméricamente relacionadas y funciones topológicas no espaciales para objetos geométricos.
- Una topología geométrica que define funciones topológicas de consulta.
- Una vinculación *RDFS* que define un mecanismo de coincidencias implícitas en tripletas *RDF* que son derivadas en un *RDF* o en un esquema *RDFS*.
- Un *query rewriter* que define reglas para transformar un simple patrón de tripletas a una relación topológica entre dos objetos en una *query* equivalente que envuelve concretamente funciones de consulta geométrica y topológica.

Cada una de esas componentes descritas forman parte de los requerimientos para *GeoSPARQL*, el cual está diseñado para acoplarse a sistemas cualitativos basados en razonamiento espacial y sistemas cuantitativos basados en cómputo espacial

3.7 Servicios Web

Según la W3r¹⁷ un servicio web es un sistema de software diseñado para dar soporte interoperable a la interacción máquina-a-máquina sobre una red. Posee una interfaz descrita en un formato procesable para una máquina

¹⁷ <http://www.w3.org/TR/2005/WD-ws-addr-core-20050215/>

(específicamente *WSDL*). Otros sistemas interactúan con el servicio Web en un amanaera prescrita por la descripción usando mensajes *SOAP*, usualmente transmitido usando *http* con una socialización de *XML* en conjunto de otros estándares de la Web.

Un servicio Web al ser una noción abstracta es necesario implementar un agente, el cual es una pieza de software o hardware que envía y recibe mensajes, mientras el servicio es un recurso caracterizado por un conjunto abstracto de funcionalidad que son provistas.

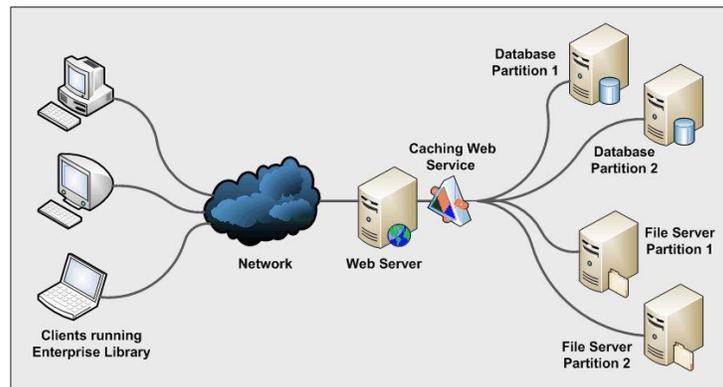


Figura 3.5.- Esquema general de un servicio Web

Como se ve en la Figura 3.5, las peticiones de los clientes, en este caso computadoras, viajan a través de la red hasta el servidor donde se encuentra implementado el servicio Web, el cual recupera la información de la petición y realiza un determinado proceso, para regresar un *XML* que contenga la respuesta obtenida al usuario que realizo la petición.

Capítulo 4 Metodología

En este capítulo se describirá, de manera general, la metodología propuesta, y de manera exhaustiva las etapas que la componen a la misma, describiendo cuales técnicas, estándares y protocolos fueron implementados en cada etapa.

4.1 Descripción de la metodología

La metodología propuesta en este trabajo es capaz de realizar recomendaciones acerca de sitios de interés, usando como base un perfil de usuario el cual contiene una lista de las actividades generales como comer o realizar compras, hasta realizar una visita a cierto sitio, incluyendo la posición actual del usuario y ponderaciones, para generar ruta con el itinerario turístico, esto por medio de consultas realizadas a un repositorio con la información turística disponible. A parte de poder llevar acabo ciertas operaciones como conocer que sitios existen alrededor de una posición, recuperar los sitios donde se realicen ciertas actividades, recuperar los sitios que están sobre una calle ó pedir una ruta entre dos o más sitios.

De manera más específica, la metodología es capaz de procesar una petición a un servicio web el cual recibe un perfil de usuario, una lista ordenada de lugares y/o actividades a realizar además de su posición actual, para caracterizarla y formar un vector el cual será usado para realizar un análisis, éste determinará las consultas que se realizarán al *triple store*¹⁸ que contiene la red de ontologías, así como las conexiones de *Linked Data*¹⁹ hacia otras fuentes de

¹⁸ <http://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/rusher.html>

¹⁹ <http://linkeddata.org/>

información, esto para recuperar mayores detalles y descripciones de algún sitio determinado, dependiendo de la operación que se realice se determina si se devuelve a la petición un sitio, información o una ruta. Después de recuperar los sitios o rutas, estos son mandados como respuesta para pasar a ser visualizados en una cartografía base.

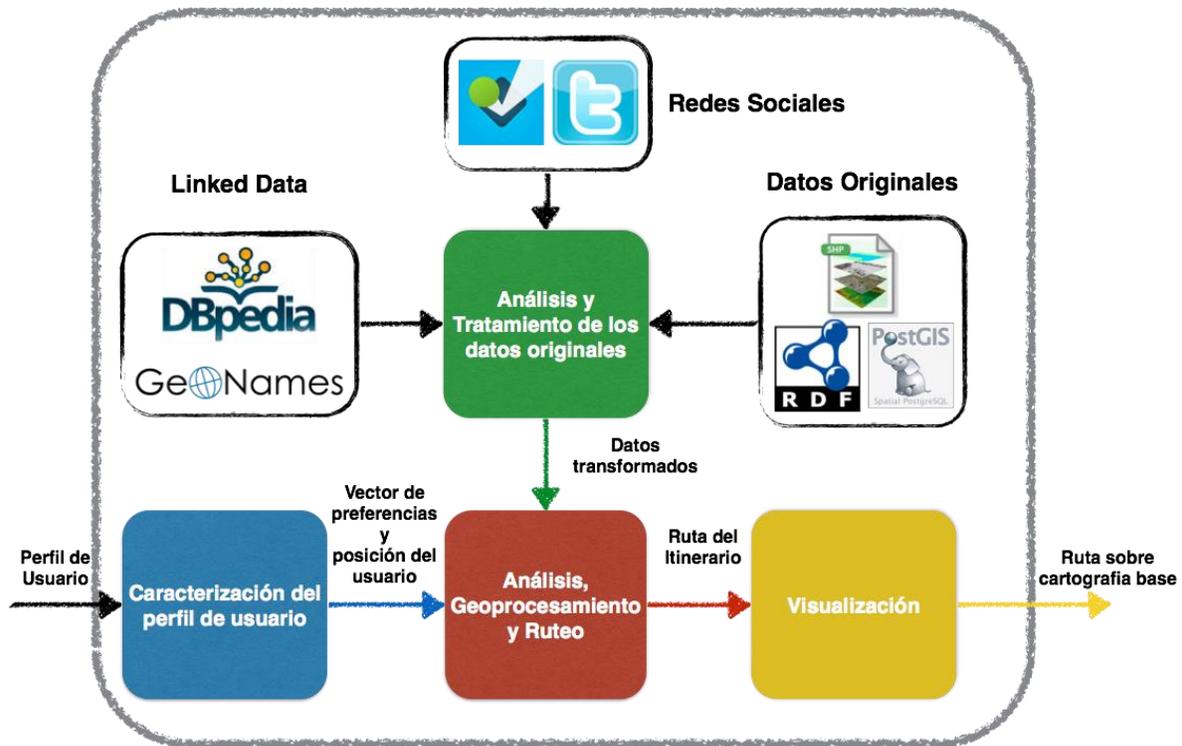


Figura 4.1 Diagrama general de la metodología propuesta

Como se puede ver en la Figura 4.1, la metodología está formada por cuatro bloques: el **primer bloque** es el análisis y tratamiento de los datos originales, dentro de este bloque se realizan los cambios necesarios a la información original, como por ejemplo la transformación del callejero y sitios de interés, así como el establecimiento de las conexiones necesarias para recuperar información de la Web 2.0 (Redes Sociales) y Web 3.0 (*Linked Data*). El **segundo bloque** es la caracterización del perfil de usuario con base en las preferencias del usuario para obtener un vector de preferencias y posición de usuario. El **tercer**

bloque es la realización del análisis del vector y geoprocesamiento de datos conforme a las operaciones espaciales que se requieran realizar, éste nos otorgará los sitios a recomendar para generar una ruta ponderada acorde a los requerimientos del usuario. Finalmente, el **cuarto bloque** se centra en la visualización de la ruta sobre una cartografía base que se desplegará para el usuario. A continuación se describirán con mayor detalle cada uno de los bloques de la metodología empezando por la creación del repositorio general.

4.1.1 Análisis y tratamiento de los datos originales

Este es el primer bloque es donde se lleva a cabo el proceso inicial de tratamiento y procesamiento de los datos originales de las vialidades que están en *shapefiles*²⁰ y sitios de interés del área de estudio, los cuales se encuentran en una BDE (Base de Datos Espacial). Así, primero se procede a crear un *SPARQL Endpoint*, el cual almacenará todas las fuentes de información necesarias para el sistema. En este trabajo se decidió usar *Parliament*²¹ como *triple store*, esto debido a que ya tiene soporte para realizar consultas con el estándar de *GeoSPARQL*. En el *triple store* se recogerán los siguientes componentes que permitirán dar funcionalidad al sistema desarrollado como se ve en la Figura 4.2:

- **Red de ontologías**
- **Transformación de información original a *RDF***
- **Enriquecimiento con datos de la Web 3.0**
- **Enriquecimiento con datos de la Web 2.0**

²⁰ <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>

²¹ <http://parliament.semwebcentral.org/>

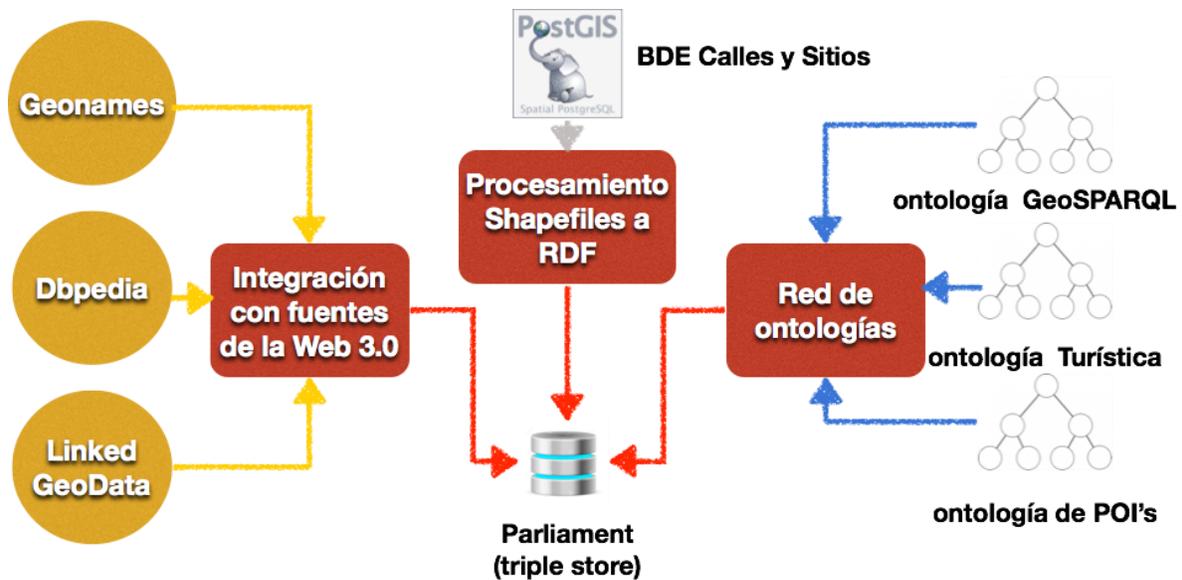


Figura 4.2 Fuentes de datos del *Triple Store*

Durante las siguientes secciones se explicará una por una las fuentes de datos de donde se recuperan los datos, para después ser almacenados en *triple store*.

4.1.1.1 Red de ontologías

Como primera fuente de información se tiene la red de ontologías, según (Allocca, C.; D'Aquin, M.; Motta, E. 2009) una red de ontologías puede ser definida como la colección de ontologías relacionadas entre una variedad de diferentes relaciones tales como mapeo, modularización y versionado entre otras, en nuestro caso la red de ontologías se encuentra en formato *RDF*, con el fin de ser almacenado en un *triple store*, las ontologías únicamente poseen siendo los conceptos, relaciones y tipos de datos sin instancia alguna. En la Figura 4.3 se mencionan las ontologías que la conforman la red de ontologías desarrollada.

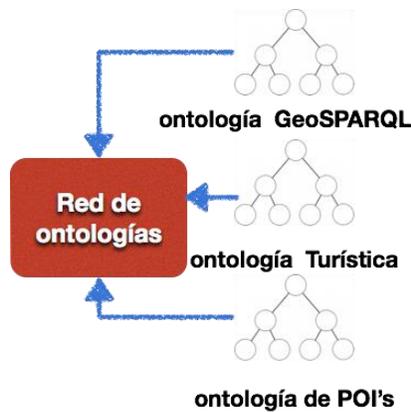


Figura 4.3.- Ontologías que integran la red de ontologías

- **Ontología turística.-** Esta ontología está basada en el tesoro creado por la Organización Mundial de Turismo junto con el gobierno francés, adaptando los conceptos que existen en el caso de estudio, la ontología está diseñada según la metodología de *METHONTOLOGY* y la herramienta *Protégé*.
- **Ontología GeoSPARQL.-** Ontología desarrollada por parte del Open Geospatial Consortium que permite darle a las instancias de la ontología turística una componente geográfica, la cual será usada para realizar consultas espaciales necesarias.
- **Ontología POIs.-** Ontología desarrollada que recoge puntos de interés (*POIs*), así como actividades y eventos relacionados con la actividad turística.

De tal manera que las ontologías están almacenadas en una red de ontologías, que actúan como una sola ontología guardada en el *SPARQL Endpoint* esperando ser poblada por el proceso de transformación de los datos originales (*shapefiles*) a RDF.

4.1.1.2 Transformación de información original a RDF

La segunda fuente de información son las vialidades y puntos de interés que pasarán a ser las instancias de la red de ontologías como se ve en la Figura 4.4. Las vialidades son tomadas del Directorio Estadístico Nacional de Unidades Económicas (DENUE) y se encuentran en el formato *shapefile*, para su almacenamiento en una base de datos espacial (BDE) se hace uso de la herramienta *QGIS*²², la cual tiene un *plug-in* llamado *SPLIT* que permite realizar conexiones base de datos para transformar los datos del *shapefile* a tuplas, en este caso se decidió utilizar el manejador de base de datos *PostgreSQL*²³ con su extensión espacial *PostGIS*²⁴ y realizar la transformación de la información del *shapefile* a una tabla que contiene sus atributos definido aparte de un atributo geométrico.

Para los puntos de interés se capturan los *data properties* especificados en la ontología como campos en la base, además de agregar dos campos, latitud y longitud, otra manera de realizar el procedimiento es generando una capa de puntos en *QGIS* la cual represente los puntos de interés, éstos poseerán atributos iguales a los *data properties* de la ontología turística, después se utiliza el *plug-in SPLIT* de *QGIS* para transformar los datos a una tabla de puntos en la base de datos espacial.

Una vez almacenados en la base de datos espacial las nuevas tablas (vialidades y puntos de interés), se empleará un programa desarrollado en el lenguaje *Java* con el ambiente de desarrollo *Eclipse*, el cual llevará acabo la transformación de las tuplas contenidas en las tablas de la base de datos, a un archivo *RDF*, éste contendrá las tripletas que describan cada instancia, a la vez se

²² <http://www.qgis.org/es/site/>

²³ <http://www.postgresql.org/>

²⁴ <http://postgis.net/>

generará una nueva relación entre la instancia y su componente geográfica por medio de una *URI*, este archivo contendrá los prefijos necesarios para el uso correcto de los estándares, también se definirán nuevos prefijos relacionados a la red de ontologías, los cuales contendrán la *URI* del concepto o clase a la que pertenece en la ontología.

De tal manera que al llevar acabo el almacenamiento de los archivos *RDF* de las vialidades y de puntos de interés en el *triple store*, estos pasen a poblar la ontología de manera automática, esto gracias a la relación generada por las *URI* únicas utilizadas al momento de transformar los datos.



Figura 4.4 Procesamiento de *shapefiles* y *POIs* a formato *RDF*

4.1.1.3 Enriquecimiento con datos de la Web 2.0 (Redes sociales)

Para realizar el filtrado colaborativo, se requiere de evaluaciones previas de usuarios sobre ciertos lugares a los que han asistido, en este caso se decidió usar la red social *Foursquare* ²⁵, debido a que dentro de los atributos que nos proporciona se encuentra una evaluación realizada a los sitios con base en criterios que utilizan para determinar la evaluación final. También se hace uso la

²⁵ <https://pypi.python.org/pypi/foursquare>

red social *Twitter*²⁶, esto con el fin de recuperar *tweets* acerca de los sitios de interés y mostrarlos al usuario como información complementaria, sin que participe en las evaluaciones para el filtrado.

Para este proceso se realizan consultas de la información disponible acerca del lugar mediante el *API* de *Foursquare* para este fin se desarrolló un programa en el lenguaje *Python*, el cual consulta el atributo *rating* de los sitios a recomendar del usuario y generar un archivo *RDF* con las tripletas necesarias, al mismo tiempo se define una nueva *data property* que se agregará a la ontología turística, con la cual se establecerá una relación entre *rating* con su respectiva instancia. El archivo *RDF* se carga en el *SPARQL Endpoint*, debido a las *URI* definidas, las tripletas del *RDF* pasan a ser instancias de la ontología relacionadas a la nueva *data property rating* junto con su valor como se ve en la Figura 4.5.



Figura 4.5 Diagrama de integración de evaluaciones a RDF

²⁶ <https://dev.twitter.com/docs/adding-social-framework>

4.1.1.4 Enriquecimiento con datos de la Web 3.0

La tercera fuente de datos es la integración de fuentes de datos de la Web 3.0 (la Web de Linked Data) como se ve en la Figura 4.6. En este caso, a la red de ontologías se le incluyen las sentencias necesarias para poder enriquecer nuestros datos con los que existen en otros conjuntos de datos ya disponibles en la mencionada Web, en este caso se utilizan los siguientes conjuntos de datos:

- **DBpedia**²⁷.- Se utiliza para recuperar las imágenes de los sitios de interés, así como una descripción del sitio en varios idiomas.
- **GeoNames**²⁸.- Se utiliza para comparar los sitios que se tienen con los que posee el conjunto de datos, si existen sitios que no se encuentran en los nuestros se pueden agregar y así incrementar los sitios de interés.
- **LinkedGeoData**²⁹.- Se utiliza para comparar los sitios que se tienen con los que este *SPARQL Endpoint* posee realizando una consulta al mismo, si existen sitios que no se encuentran en los nuestros se pueden agregar y así incrementar los sitios de interés.

²⁷ <http://dbpedia.org/About>

²⁸ <http://www.geonames.org/>

²⁹ <http://linkedgeo.org/About>

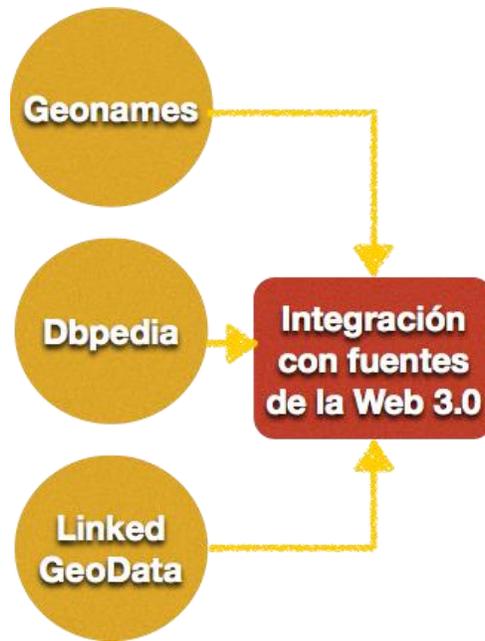


Figura 4.6 Integración de con fuentes de la Web 3.0.

Para realizar el enriquecimiento se establecen relaciones *owl:sameAs*,. Así, para cada recursos de información turística se realiza una consulta al *SPARQL Endpoint* de cada fuente de información, en nuestro caso se realiza consulta por los siguientes atributos: descripción del lugar en idioma español, pagina Web y una imagen del sitio turístico, con el objetivo de enriquecer la información original con las fuentes mencionadas.

Estas cuatro fuentes de datos integran la información que será almacenada en el *SPARQL Endpoint*, con lo cual se podrán realizar las consultas necesarias, además de poder adjuntar las evaluaciones de las redes sociales al mismo. Estando completo el *SPARQL Endpoint* se procederá a la explicación de los demás módulos que conforman la metodología conforme al diagrama de la Figura 4.1.

4.1.2 Caracterización del perfil de usuario

Este segundo bloque toma como entrada una petición que contiene un perfil de usuario, el cual contiene las actividades y sitios que el usuario desea realizar durante su estancia; la posición del usuario en coordenadas geográficas (latitud, longitud); las ponderaciones acerca de que tanto le interesa la cercanía de los lugares a visitar, así como la opinión de otros usuarios para formar un vector de preferencias y posición de usuario como se ve en la Figura 4.7.



Figura 4.7 Diagrama entrada y salida del bloque de Caracterización.

Con esas variables, se procede a generar un vector el cual contiene las siguientes variables para generar un itinerario, de tal manera que el vector es de esta forma:

$$v_i = \langle (\text{sitio/actividad})_1, (\text{sitio/actividad})_2, (\text{sitio/actividad})_3, \dots, (\text{sitio/actividad})_n, \lambda, \varphi, \text{distancia}, \text{opinión} \rangle$$

Donde :

- **(sitio/actividad)_n**. Se refiere al sitio o actividad que desea realizar el usuario.
- **λ**. Se refiere a la latitud de la posición inicial, esta puede ser de la posición del usuario, de un punto de partida sea determinado por el usuario ó un sitio.
- **φ**. Se refiere a la longitud de la posición inicial, esta puede ser de la posición del usuario, de un punto de partida sea determinado por el usuario ó un sitio.
- **distancia**. Se refiere a la ponderación por parte del usuario acerca de que tanto le interesa al usuario la distancia que hay que hay que recorrer entre lugares.
- **opinión**. Se refiere a la ponderación por parte del usuario acerca de que tanto le interesa al usuario las opiniones que otros usuarios realizaron acerca de un lugar.

El vector puede variar dependiendo de que sea la operación que desea realizar el usuario, debido que no todas las operaciones que se pueden realizar necesitan todas las variables, esto está determinado dependiendo del proceso.

Otras variables que pueden ser puestas en el vector, son las categorías o actividades a las que pertenecen los sitios de interés, la distancia para saber que puntos de interés existen alrededor de la posición de un usuario.

Al final este vector es mandado a un servidor el cual posee un servicio Web implementado en Java, que se encargara de procesarlos datos y determinar que procesos deben ser ejecutados dependiendo de las necesidades del usuario.

4.1.3 Análisis, Geoprocesamiento, Ruteo

Este bloque (tercer bloque de la metodología propuesta) utiliza del bloque anterior el vector de preferencias y posición de usuario, además de la petición del proceso a realizar. De tal forma que se realiza una petición SOAP al servidor que tiene implementado un servicio Web que procesa el vector, éste se manda embebido en un XML hacia el servicio adjuntando el proceso que se realizará como se ve en la Figura 4.8.

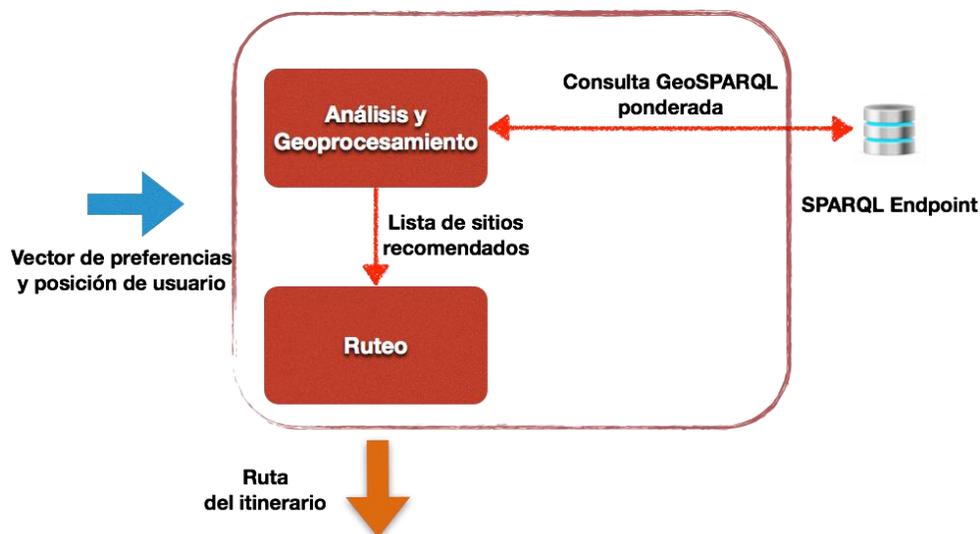


Figura 4.8 Diagrama general del bloque de Análisis, Geoprocesamiento y Ruteo

4.1.3.1 Análisis geoespacial

Dentro de este bloque se realiza un análisis geoespacial sobre los sitios de interés que se van a recomendar al usuario haciendo uso del estándar *GeoSPARQL*, este cuenta con varias funciones geoespaciales que son utilizadas en varios procesos que se describirán a continuación:

- **Buffer de sitios.** Este proceso nos permite implementar la operación *buffer* para recuperar los sitios existentes alrededor de una posición que proporciona el usuario.
- **Categorías de sitios.** Este proceso recupera las categorías que existen en la ontología con el fin de mostrarle al usuario las categorías existentes y consultar los sitios que pertenecen a las mismas.
- **Recuperación de sitios disponibles.** Este proceso recupera la lista de todas las instancias que existen dentro de la ontología.
- **Búsqueda de un solo sitio.** Este proceso realiza una búsqueda en la ontología para recuperar información acerca de un sitio y devolver la posición geográfica del mismo.
- **Contenido de una calle.** Este proceso recupera los sitios que existen sobre alguna calle de la cartografía que está almacenada en la ontología, por medio de un *buffer* sobre la posición que proporciona el usuario para conocer qué segmento de calle intersecta y recuperar los sitios sobre todos los segmentos.
- **Descripción de sitio.** Este proceso recupera toda la información disponible de la red de ontologías y verifica si existe una relación *owl:sameAs* que pueda ser consultada para recuperar información adicional de la Web de Linked Data acerca del sitio.

- **Itinerario turístico.** Este proceso nos devuelve un conjunto de sitios que se recomendarán al usuario dependiendo de las actividades/sitios junto con las ponderaciones que proporciono en su perfil.
- **Data properties de los conceptos.** Este proceso nos devuelve las propiedades dependiendo de la categoría que se defina, incluyendo las heredadas de conceptos principales.

Después de realizar alguno de estos procesos, se retorna la respuesta dependiendo del proceso, los valores que puede retornar son un lugar, un arreglo que contiene las posición de los sitios de interés o una descripción. En el caso que se requiera una ruta entre dos sitio o un itinerario, se solicita al servicio de ruteo que proporciona Apple mediante el *framework MapKit* que incluye en su *suite* de desarrollo *Xcode*, el trazado de ruta sobre su cartografía base, dependiendo de los lugares que retorne el servicio Web.

4.1.4 Visualización

En este bloque se despliega el resultado conforme a la petición que haya hecho el usuario. Así, si fue un itinerario, se procede a realizar el trazado de la ruta ponderada sobre la cartografía base de Apple, para indicarle al usuario por donde debe caminar para llegar a los lugares recomendados. Si fue una operación geoespacial se desplegará ya sea los lugares alrededor de una posición en forma de punto, los sitios que se encuentren sobre alguna calle en forma de punto o el trazo de la ruta entre dos lugares. Si fue una búsqueda de sitios se desplegará los sitio(s) buscado(s). Por último si la petición requería una descripción de algún sitio, se procederá a desplegar las características encontradas acerca del lugar en específico, donde se mostrará tanto la información original que se poseía como la que se recuperó de la Web 3.0.

Capítulo 5 Pruebas y Resultados

En este capítulo se describen las pruebas de la pertinencia y validez de la metodología desarrollada, descrita en el capítulo 4. Estas pruebas han permitido el desarrollo de un sistema sobre un dispositivo móvil, en este caso en un dispositivo iPhone con sistema operativo iOS, usando un caso de estudio con el fin de probar las diferentes funcionalidades implementadas, de tal manera que se pueda verificar que el sistema es funcional.

5.1 Caso de estudio

Para este trabajo se tomó como caso de estudio el centro histórico de la Ciudad de México, específicamente el primer cuadro como se ve en la Figura 5.1. Se escogió esta zona debido a la gran diversidad de puntos de interés turístico (por ejemplo, locales comerciales, restaurantes, teatros, antros, bares, etc.,) que existen en esta zona. Así, esta zona recoge uno de los principales atractivos turísticos, lo que la hace uno de los lugares más concurridos del Distrito Federal para los turistas ya sean foráneos o extranjeros.



Figura 5.1 Zona 1 primer cuadro del Centro Histórico de la Ciudad de México.

A continuación, tomando como referencia el mencionado caso de estudio, se demostrará la aplicabilidad de la metodología, describiendo de manera exhaustiva cada una de las etapas y sus procesos que lo componen.

5.2 Análisis y tratamiento de los datos originales

En esta sección se describe cómo se tratan los datos originales para formar los archivos RDF que conforman el SPARQL Endpoint, empezando por la red de ontologías, el análisis y tratamiento de los *shapefiles*, así como la recuperación de datos de las redes sociales y de la nube de *Linked Data*.

5.2.1 Diseño de la red de ontologías turísticas

Para la creación de la red de ontologías se utilizan la metodología *NeOn*³⁰, usando el escenario 2 para generar desde el principio la ontología turística que fue desarrollada conforme a la *METHONTOLOGY*³¹, el escenario 3 que habla sobre la reutilización de recursos ontológicos, en nuestro caso son las ontologías de *GeoSPARQL*³² y puntos de interés (*POIs*), y por último el escenario 2 donde se reutilizan recursos no ontológicos como los sitios turísticos que proporciona el Gobierno de la Ciudad de México a través de su guía del Centro Histórico. A continuación se detalla el proceso para la primera ontología con sus respectivas tareas, cabe recordar que los conceptos son extraídos del tesoro de la Organización Mundial de Turismo³³.

³⁰ <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/es/methodologies/59-neon-methodology>

³¹ http://oa.upm.es/5484/1/METHONTOLOGY_.pdf

³² https://portal.opengeospatial.org/files/?artifact_id=47664

³³ <http://www2.unwto.org/es>

Tarea 1.- Glosario de Términos

El glosario de términos está compuesto de las definiciones de los conceptos, instancias y relaciones que conformaran a la ontología turística como se ve en la tabla 5.1.

Tabla 5.1 Ejemplo del Glosario de Términos

Nombre	Sinónimos	Acrónimos	Descripción	Tipo
Turismo	-	-	"Es un fenómeno social, cultural y económico relacionado con el movimiento de las personas a lugares que se encuentran fuera de su lugar de residencia habitual por motivos personales o de negocios/profesionales."	Concepto
Vida nocturna	-	-	Es un término colectivo utilizado para cualquier entretenimiento que esté disponible y sea más popular desde la noche hasta las primeras horas de la mañana.	Concepto
Sitios Turísticos	-	-	"Sitio donde existe una atracción para los turistas ya sean locales o internacionales, estos pueden ser naturales o hechos por la civilización."	Concepto
Sitios Culturales	-	-	Es un espacio donde se pueden realizar actividades culturales y su principal objetivo es el de promover la cultura entre los turistas locales o extranjeros.	Concepto
Centro de Espectáculo	-	-	"Establecimiento donde se realizan funciones teatrales, conciertos, exposiciones, etc."	Concepto

Tarea 2.- Taxonomías de conceptos

Aquí se muestra el modelo taxonómico de los conceptos extraídos del tesauro de la Organización Mundial de Turismo³⁴, como se ve en la Figura 5.2 conforme al caso de estudio se discriminaron varios de los conceptos que contempla el tesauro, esto debido a el tipo de atracciones que existen en el área.



Figura 5.2 Taxonomía de la ontología

³⁴ <http://www2.unwto.org/es>

Tarea 3.- Relaciones Ad-Hoc

Para la ontología propuesta, se analizó las relaciones existentes conforme a los conceptos seleccionados conforme al caso de uso y no se lograron encontrar relaciones del tipo Ad-Hoc después del análisis.

Tarea 4.- Diccionario de Conceptos

A continuación se muestra un ejemplo del diccionario de conceptos, el cual contiene todas las instancias junto a la relación con sus respectivos conceptos, así como las relaciones existentes entre los conceptos del diccionario, como se muestra en la Tabla 5.2.

Tabla 5.2 Ejemplo del Diccionario de Conceptos

Nombre de Concepto	Instancias	Atributos de Clases	Atributos de Instancia	Relaciones
Sitios Turísticos	-	Dirección Nombre Horario Tiempo estimado	-	part_of compose_of is_a
Museo	Museo del Templo Mayor Museo Nacional de las Culturas Museo Archivo de la Fotografía	Tipo de Museo	-	compose_of is_a
Edificio Histórico	El Palacio de Hierro Sanborns (Casa de los Azulejos) Antiguo Palacio de Moctezuma / Palacio Nacional"	-	-	compose_of is_a
Hoteles	Chillout Flat Hotel Downtown Tulip Inn Ritz México City Hostel Hotel Roble	-	Ascensor Habitación Doble Habitación Sencilla Habitación Triple Climatización	part_of compose_of se_ofrece

Restaurantes	Salón Luz Salón España Salón Corona La ópera El Palacio de Hierro Sanborns (Casa de los Azulejos)	Tipo de Comida	Servicio de Bar	part_of compose_of se_ofrece
Salón de Baile	Oasis y Viena Club La Perla"			compose_of is_a
Local Comercial	Galerías Castillo Porrúa Librería Educal Suc. Templo Mayor Sposa Nova	TipodeLocalComercial	-	part_of compose_of is_a

Tarea 5.- Descripción de relaciones binarias ad hoc

Conforme a lo realizado en la Tarea 3, al no existir relaciones Ad-Hoc en el diseño de la ontología por lo tanto no se puede realizar esta tarea.

Tarea 6.- Descripción de los atributos de instancia

En la descripción de los atributos de instancia se muestran que tipo de valores puede tomar un atributo de instancia, así como su rango de valores y cardinalidad de los mismos, como se ve en la tabla 5.3.

Tabla 5.3 Ejemplo del Diccionario de Conceptos

Nombre de atributo de instancia	Concepto	Tipo de Valor	Rango de Valores	Cardinalidad
Funciones	Centro de Espectáculo	Cadena de Caracteres	-	"(1,1)"
Librería	Local Comercial	booleano	[false, true]	"(1,1)"
Tienda de Ropa	Local Comercial	booleano	[false, true]	"(1,1)"
Tienda de Cosméticos	Local Comercial	booleano	[false, true]"	"(1,1)"
Zapatería	Local Comercial	booleano	[false, true]	"(1,1)"
Ascensor	Hoteles	booleano	[false, true]	"(1,1)"

Habitación Doble	Hoteles	booleano	[false, true]	"(1,1)"
Jardín Privado	Hoteles	booleano	[false, true]	"(1,1)"
Estacionamiento	Hoteles	booleano	[false, true]	"(1,1)"
Piscina Privada	Hoteles	booleano	[false, true]	"(1,1)"
Playa privada	Hoteles	booleano	[false, true]	"(1,1)"
Comedor	Hoteles	booleano	[false, true]	"(1,1)"
Música en Vivo	Centro de Entretenimiento Nocturno	booleano	[false, true]	"(1,1)"
Show en Vivo	Centro de Entretenimiento Nocturno	booleano	[false, true]	"(1,1)"

Tarea 7.- Descripción de los atributos de clase

Al igual que en la sección anterior, en la tabla 5.4 se ven los atributos de clase junto a la clase a la que pertenece, tipo de valor que puede tener además de su cardinalidad definida en el análisis.

Tabla 5.4 Ejemplo de Descripción de atributos de clase

Nombre de atributo de clase	Concepto	Tipo de Valor	Cardinalidad	Valores
Dirección	Sitios Turísticos	Cadena de Caracteres	"(1,1)"	-
Nombre	Sitios Turísticos	Cadena de Caracteres	"(1,1)"	-
Horario	Sitios Turísticos	Cadena de Caracteres	"(1,1)"	-
Tiempo estimado	Sitios Turísticos	Cadena de Caracteres	"(1,1)"	-
Tipo de Museo	Museo	Arte, Temático, Historia, Ciencias, Antropología, Arqueología]	"(1,6)"	-
Tipo de Comida	Restaurantes	Mexicana, Japonesa, China, Tailandesa, Alemana, Española, Fast Food, Snacks	"(1,8)"	-

Tipo de Local Comercial	Local Comercial	Librería, Sombrerería, Bisutería, Tienda de Mochilas, Tienda de Ropa, Tienda de Ropa Deportiva, Tienda de Ropa Militar, Tienda de Música, Tienda de Arte, Tienda de Artículos de Piel, Tienda de Artículos Religiosos, Tienda de Decoración, Tienda Departamental, Tienda de Empeño, Perfumería, Ferretería, Tienda de Abarrotes	"(1,22)"	-
Cafetería	Restaurantes	Booleano	"(1,1)"	-

Tarea 8.- Descripción constantes

Dentro de esta tarea se describen las constantes que se utilizarán como auxiliar para generar axiomas y reglas para su posterior definición.

Tabla 5.5 Ejemplo de Descripción de constantes

Nombre	Tipo de Valor	Valor	Unidad de Medida
Mayoría de Edad	Entero	18	Años

Tarea 9.- Definición de axiomas formales

Dentro del diseño de la ontología turística , después de su análisis no se crearon axiomas formales que rijan a la ontología.

Tarea 10.- Definición de reglas

Para la ontología se definieron varas reglas, las cuales nos ayudan en a delimitar los casos que existen, proporcionando una norma a seguir si se da el caso como se ve en la tabla 5.6.

Tabla 5.6 Ejemplo de Descripción de reglas

Nombre de la Regla	Descripción	Expresión	Conceptos	Atributos	Relaciones	Variables
Regla Deportes No Acuáticos	Solo se puede practicar deportes no acuáticos en centros deportivos	"Si[DeportesNoAcuatico](?X) y CentrosDeportivos(?Y) entonces se_practica(?X,?Y)"	Deportes No Acuáticos Centros Deportivos	-	se_practica se_practica	?X ?Y
Regla Eventos Deportivos	Solo se pueden organizar eventos deportivos en centros deportivos	"Si[EventoDeportivo](?X) y CentrosDeportivos(?Y) entonces se_realiza(?X,?Y)"	Eventos Deportivos Centros Deportivos	-	se_realiza se_realiza	?X ?Y
Regla Eventos Culturales	Solo se pueden organizar eventos culturales en Sitios Culturales y en Centros de Espectáculos	"Si[EventoCultural](?X) y (CentrosdeEspectaculo(?Y) o SitiosCulturales(?Z)) entonces (se_realiza(?X,?Y) o se_realiza(?X,?Z))"	Eventos Culturales Centros de Espectáculos Sitios Culturales	-	se_realiza se_realiza se_realiza	?X ?Y ?Z
Regla Vida Nocturna	La vida nocturna solo se puede dar en Centros de Entretenimiento Nocturno y despues de las 19:00 hrs	"Si [CentrosdeEntretenimientoNocturno(?X)] y horario(?X,?Y) y ?Y>19:00 entonces [VidaNocturna(?Z)]"	Vida Nocturna Centros de Entretenimiento Nocturno	Horario	-	?X ?Y ?Z

Tarea 11 Definición de instancias

En la tabla 5.7 se pueden ver algunas instancias de la ontología definiendo su nombre de instancia, nombre del concepto al que pertenecen, atributo que se tiene y el valor que podrían tomar.

Tabla 5.7 Ejemplo de descripción instancias

Nombre de la instancia	Nombre de Concepto	Atributo	Valores
Limosneros	Restaurante	id	2
		nombre	Limosneros
		dirección	"Allende 3, colonia Centro, Ciudad de México."
		TipodeComida	Mexicana
		Cafetería	FALSE
		Servicio de bar	FALSE
		Horario	Lunes a viernes 13:30 a 23:00 horas; sábado de 9:00 a 23:00 y domingo de 9:00 a 19:00.
Antiguo Palacio de Moctezuma / Palacio Nacional	Edificio Histórico	id	3
		nombre	Antiguo Palacio de Moctezuma / Palacio Nacional
		Horario	Martes a Domingo de 10 a 17 horas.
		dirección	Costado este de la Plaza Mayor.
Fundación Herdez	Museo	id	4
		nombre	Fundación Herdez
		horario	Martes a Sábado de 9 a 17 horas.
		dirección	"Seminario 18, colonia Centro, Ciudad de México."
		TipodeMuseo	Gastronómico
		Tiempo estimado	50 minutos

5.2.2 Transformación de información original a *RDF*

Conforme al diseño de la red de ontologías ya mencionado en la primera etapa es necesario realizar un tratamiento de los datos originales, para poder llevar a cabo una población de nuestra red almacenada en el *SPARQL Endpoint*. En este sentido, se procede a transformar a *RDF* los datos originales de las vialidades del área de estudio, que se encuentran en formato *shapefile*, de los sitios de interés, recuperados del *DENUE*³⁵ (Directorio Estadístico Nacional de Unidades Económicas) y de la Guía del Centro Histórico³⁶, capturados como instancias en una hoja de cálculo Excel.

Una vez capturados los datos se procede a guardar el archivo Excel con formato *CSV (Comma-Separated Values)* y utilizarlo para crear una capa de puntos, esto usando la herramienta *QGIS*, la cual nos permite crear la capa y realizar correcciones en los puntos desfasados si existen y se hace que coincidan en sus coordenadas correctas usando como base la capa de vialidades que ya se tiene.

Una vez realizado ese proceso creamos una base de datos de nombre “bdtesis” en *PostgreSQL* usando el *template “template postgis”* que proporciona las funciones espaciales necesarias que incluye la extensión espacial *PostGIS*. Ahora usamos el *plug-in SPLIT* de *QGIS* para realizar una conexión a la base de datos “bdtesis” como se ven la Figura 5.3

³⁵ <http://www3.inegi.org.mx/sistemas/mapa/denue/default.aspx>

³⁶ <http://www.guiadelcentrohistorico.mx/>

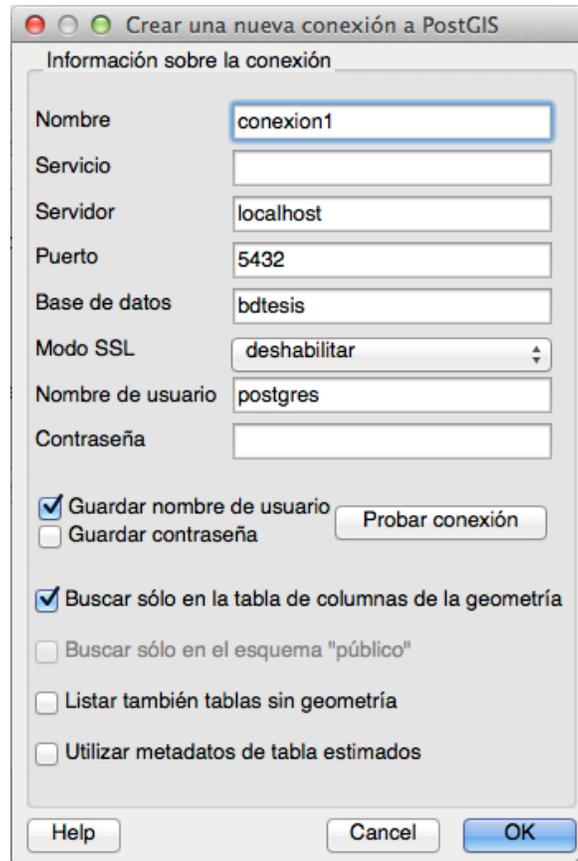


Figura 5.3 *Plug-in SPLIT* haciendo conexión a la base de datos

Una vez creada la conexión se realiza una prueba para verificar si realmente se está haciendo la conexión a la base de datos, procedemos a exportar tanto la capa de vialidades del área de estudio así como los sitios de interés que se tienen, esto con la finalidad de obtener dos tablas que cuenten con los atributos y su respectiva componente geométrica de cada uno de los objetos que existen en las capas, dentro de la base de datos como se ve en la Figura 5.4.

	gid [PK] serial	STREET character	FROMLEFT integer	TOLEFT integer	FROMRIGHT integer	TORIGHT integer	TYPE integer	SENTIDO character	COLONIA character	NIVELSE character	OBSERVAC character	the_geom geometry
1	1	11 DE ABRIL	28	36	27	37	-5	b	GUERRERO			0102000000
2	2	11 DE ABRIL	4	12	0	0	-5	b	GUERRERO			0102000000
3	3	16 DE SEPT	26	26	23	23	-5	b	CENTRO			0102000000

Figura 5.4 Tabla de vialidades del área de estudio

Teniendo ya los datos en la base de datos, se desarrolló un programa en java que realiza consultas a la base de datos “bdtesis” con el fin de generar las tripletas del archivo *RDF* resultante, definiendo las siguientes *URI* :

- **<http://datos.itinerariosTuristicos.mx/recurso/geometry/>**
para la geometría de los objetos.
- **<http://datos.itinerariosTuristicos.mx/ontologia/>**
para los conceptos definidos en la ontología.
- **<http://datos.itinerariosTuristicos.mx/recurso/>**
para los recursos (instancias) que se asignarán a la ontología.

Estas *URI* son usadas tanto para los sitios de interés como para las vialidades, esto para asignar la componente geométrica y hacer uso del estándar *GeoSPARQL*, se decidió usar ambos formatos *WKT* y *GML* para definir las geometrías de las instancias. El estándar de *GeoSPARQL* se emplea para definir como serán representadas las primitivas que serán usadas para formar el archivo *RDF*, en nuestro caso usamos los puntos (*Point*) como sitios de interés y líneas (*LineString*) para las vialidades como se ve en los dos siguientes ejemplo:

Para Puntos:

```
<geo:asWKT rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">
  <![CDATA[ <http://www.opengis.net/def/crs/OGC/1.3/CRS84> Point(-83.5 34.2)]]>
</geo:asWKT>
```

Para Líneas:

```
<geo:asWKT rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">
  <![CDATA[<http://www.opengis.net/def/crs/OGC/1.3/CRS84>LineString((-83.4 34.0, -83.3
  34.3) )]]>
</geo:asWKT>
```

Una vez que el programa recorre todas las tuplas de la base de datos espacial se genera el archivo RDF con ayuda de la biblioteca Jena³⁷ para Java, que usa los prefijos que se definieron anteriormente y con ayuda de las funciones espaciales que proporciona *PostGIS* extraemos coordenadas de inicio y fin de objeto espacial y su información que se conforma de los siguientes campos: nombre, latitud, longitud, concepto, horario, dirección.

5.2.3 Enriquecimiento con datos de la Web 2.0 (*Redes sociales*)

Para recuperar datos de las redes sociales se hace uso de las *APIs* de *Foursquare* y *Twitter*, como se mencionó en el capítulo 4. En el caso de la primera se desarrolló un programa en *Python* que usa la biblioteca *Foursquare* para recuperar un objeto *JSON* que corresponde a un sitio de interés, para esto previamente se buscaron los sitios de interés y se capturaron los identificadores de cada sitio, de tal manera que al momento de hacer la petición al servicio de *Foursquare* se recuperarán todas las características que posee el sitio y de las cuales se extrae mediante un filtro la calificación de cada sitio, conforme se obtiene las calificaciones por sitio se genera un archivo *RDF* donde se define una nueva *data property* llamada *Rating*, la cual contendrá las calificaciones recuperadas de cada instancia como se ve en el siguiente ejemplo.

```
<rdf:Description rdf:about="http://datos.itinerariosTuristicos.mx/recurso/Palacio_Postal">  
  <ontologia3:Rating rdf:datatype="&xsd:string">9.08</ontologia3:Rating>  
</rdf:Description>
```

En el caso de *Twitter* no influye en las calificaciones, ni aporta al *SPARQL Endpoint* información alguna, esto debido a que se emplea la información recuperada del sitio de interés en la ventana de descripción, esto mediante la búsqueda del nombre de sitio de interés y se muestran los *tweets* relacionados como se muestra en la Figura 5.5.

³⁷ <http://jena.apache.org/>



Figura 5.5 Tweet de la librería Gandhi

5.2.4 Enriquecimiento con datos de la Web 3.0

Para el enriquecimiento de los datos *RDF* generados se decidió hacer uso de diferentes conjuntos de datos presentes en *Linked Data*, tales como: *DBpedia*, *Geonames* y *Linked GeoData*. Estos conjuntos de datos forman parte del dominio geográfico de la nube de *Linked Data*, de tal forma que para poder conectar nuestra red de ontologías con estos conjuntos de datos se utilizará la relación *owl:sameAs* en las instancias que tengan alguna coincidencia o semejanza con alguna otra instancia de los conjuntos como se ve en el siguiente ejemplo.

```
<rdf:Description rdf:about="http://datos.itinerariosTuristicos.mx/recursoMuseo_del_Templo_Mayor">
  <rdf:type rdf:resource="http://datos.itinerariosTuristicos.mx/ontologia Museo"/>
  <geo:hasGeometry rdf:resource="http://datos.itinerariosTuristicos.mx/recurso/geometry/gml/Id5"/>
  <geo:hasGeometry rdf:resource="http://datos.itinerariosTuristicos.mx/recurso/geometry/wkt/Id5"/>
  <owl:sameAs rdf:resource="http://es.dbpedia.org/resource/Museo_del_Templo_Mayor"/>
  <rdfs:label xml:lang="es">Museo del Templo Mayor</rdfs:label>
</rdf:Description>
```

Después de realizar las conexiones necesarias se crea el archivo *RDF* que será almacena en el *SPARQL Endpoint* junto con la demás información, a partir de esto ya se podrán recuperar descripciones e imágenes de otros conjuntos de datos sin la necesidad de tener almacena esa información, aparte de proporcionar vínculos para que el usuario siga buscando más información por su cuenta.

5.3 Caracterización del perfil de usuario

Como se describió en el capítulo 4, este bloque usa como entrada una petición del sistema y se realiza una vectorización del mismo. A continuación se proporcionan más detalles sobre este proceso.

En el caso de la entrada para el itinerario se recibe un archivo *XML* el cual contiene el perfil del usuario que consiste de las actividades que el usuario desea realizar durante su estancia, aparte se anexa la latitud, longitud y preferencias (distancia y evaluaciones), las cuales se especifican por medio de las etiquetas del XML que se recibe.

```
<soapenv:Envelope xmlns:q0="http://Patito.pkg" xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <soapenv:Header></soapenv:Header>
  <soapenv:Body>
    <q0:itinerario>
      <q0:ActGenearl>Hotel</q0:ActGenearl>
      <q0:ActProp>
        <q0:array>HabitacionSencilla_true</q0:array>
      </q0:ActProp>
      <q0:latitud>19.4337927</q0:latitud>
      <q0:longitud>-99.138633</q0:longitud>
      <q0:Preferencias>3</q0:Preferencias>
      <q0:Preferencias>4</q0:Preferencias>
    </q0:itinerario>
  </soapenv:Header>
</soapenv:Body>
```

Este XML es mandado al servidor, el cual toma cada uno de los atributos por medio de su etiqueta y se asignan los valores a variables para vectorizarlas y mandar a llamar el proceso correspondiente como se ve en el siguiente ejemplo:

$v_1 = \langle [\text{Hotel}], [\text{HabitacionSencilla_true}], 19.4337927, -99.138633, 3, 4 \rangle$

5.4 Análisis, Geoprocesamiento y Ruteo

Esta sección corresponde al tercer bloque de la metodología, usando la salida del bloque anterior (vector de perfil de usuario y preferencias), el servidor realiza la acción indicada en la petición SOAP, usando consultas SPARQL se realizan las operaciones disponibles que fueron descritas en el capítulo 4 y que como resultado generan una respuesta XML. Esta respuesta contiene en su mayoría sitios que se recomendarán o que fueron consultados de manera explícita y en otros casos se regresarán categorías, *dataproperties* o valores de instancias requeridos. Un ejemplo de respuesta es el siguiente:

```
<?xml version='1.0' encoding='utf-8'?>
  <soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/">
    <soapenv:Body>
      <ns:buffersitiosResponse xmlns:ns="http://Patito.pkg">
        <ns:return>Shasa</ns:return>
        <ns:return>-99.1389769790748261</ns:return>
        <ns:return>19.433844360969541</ns:return>
        <ns:return>General</ns:return>
        <ns:return>The Body Shop</ns:return>
        <ns:return>-99.1390377749155078</ns:return>
        <ns:return>19.4338511148312847</ns:return>
        <ns:return>General</ns:return>
        <ns:return>Salon Corona</ns:return>
        <ns:return>-99.1384151887328784</ns:return>
        <ns:return>19.4333828168944507</ns:return>
        <ns:return>General</ns:return>
        <ns:return>Palacio de Cultura Banamex</ns:return>
        <ns:return>-99.1388179007237369</ns:return>
        <ns:return>19.433797444947956</ns:return>
        <ns:return>General</ns:return>
      </ns:buffersitiosResponse>
    </soapenv:Body>
  </soapenv:Envelope>
```

Una vez que la respuesta llega al sistema, este por medio de un *parser* de XML guarda en un arreglo general el nombre del lugar, longitud, latitud y tipo de marcador, con el fin de proporcionar a la clase que crea los marcadores la información necesaria. En el caso del ruteo, como se mencionó en el capítulo 4 se utiliza el servicio de *Apple Maps* que viene incluido en el *framework* de *Mapkit*, como se puede ver en el siguiente segmento de código.

```

- (IBAction)MostrarRuta:(id)sender {
    NSMutableArray *news = [[NSMutableArray alloc] initWithObjects:inicio,fin, nil];
    [map addAnnotations:news];
    [map showAnnotations:news animated:YES];
    MKDirectionsRequest *directionsRequest = [[MKDirectionsRequest alloc] init];
    [directionsRequest setSource:[[MKMapItem alloc] initWithPlacemark:_inicio]];
    [directionsRequest setDestination:[[MKMapItem alloc] initWithPlacemark:_destino]];
    directionsRequest.transportType = MKDirectionsTransportTypeWalking;
    MKDirections *directions = [[MKDirections alloc] initWithRequest:directionsRequest];
    [directions calculateDirectionsWithCompletionHandler:^(MKDirectionsResponse *response, NSError
*error) {
        if (error) {
            NSLog(@"Error %@", error.description);}
        else {
            routeDetails = response.routes.lastObject;
            [self.map addOverlay:routeDetails.polyline;]}
    }];
}

```

En este código se le indica el inicio y fin de la ruta así como el tipo de transporte (en nuestro caso es peatonal) y se realiza la petición, la cual nos devuelve una línea que representa la ruta y que forma la salida de nuestro bloque.

5.5 Visualización

Este es el último bloque en el cual con la línea que se genera por la petición al servicio de ruteo de Apple se procede a dibujarla en la propia cartografía base del sistema. Este mismo caso se realiza en los marcadores que representan las posiciones de los sitios de interés, pero a diferencia de las rutas, estos no requieren petición alguna, simplemente se agregan a un arreglo y se mandan a dibujar como se ve en el siguiente código.

```

for (int u=0; u<(contador1/4); ++u) {
    cl.latitude=[[strings objectAtIndex:(2+(u*4))] doubleValue];
    cl.longitude=[[strings objectAtIndex:(1+(u*4))]doubleValue];
    annotation=[[MyAnnotation alloc] initWithCoordinate:cl title:[strings objectAtIndex:(0+(u*4))]
categoria:[strings objectAtIndex:(3+(u*4))]];
    [annotationList addObject: annotation];
}
[map addAnnotations:annotationList];
[map showAnnotations:annotationList animated:YES];

```

Una vez descrito en detalle lo que se realiza en cada uno de los bloques pertenecientes a la metodología propuesta en este trabajo, a continuación se mostrarán algunos experimentos realizados sobre el sistema, mostrando las operaciones implementadas y que son disponibles para el usuario.

5.6 Experimentos

En esta sección se procede a describir de manera exhaustiva cada uno de los experimentos realizados en el sistema, donde algunos experimentos requieren hacer uso de las operaciones espaciales que se implementaron en el sistema.

5.6.1 Búsqueda de Sitios

Dentro de este experimento se llevan a cabo los procesos de búsqueda, estos pueden ser mediante el nombre del sitio proporcionado por el usuario, de igual forma se pueden buscar sitios por medio de la categoría al que pertenezca y ser desplegados en la cartografía base usando marcadores que representan la posición que tiene cada sitio. Para este experimento, el usuario seleccionará dentro del menú principal la opción de Búsqueda, la cual desplegará una nueva ventana que contendrá un *tab control* con dos opciones, la primera son de Puntos de Interés y la segunda son Actividades. Para nuestro experimento usaremos la opción de Sitios de Interés como se ve en la Figura 5.6.



Figura 5.6 Búsqueda de sitios.

5.6.1.1 Búsqueda de sitio por nombre

Para buscar un sitio por nombre, el sistema realiza una consulta de SPARQL donde se recuperará todos los nombres (*labels*) de las instancias que contenga la ontología, las cuales servirán para realizar un autocompletado de los lugares que pueda pedir el usuario.

El usuario procede a teclear el nombre del lugar que busca, usando expresiones regulares se busca dentro de la lista de sitios, las coincidencias con los sitios y se desplegarán en una lista para que el usuario seleccione su sitio como se ve en la Figura 5.7.

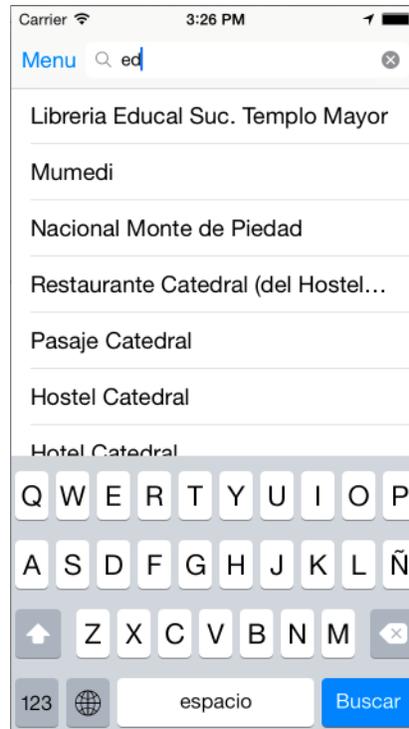


Figura 5.7 Lista de sitios.

Después de seleccionar un sitio, el sistema realiza una petición al servicio web para consultar al *SPARQL Endpoint* acerca de la posición del sitio seleccionado por medio de su nombre, el servidor regresa la posición (latitud, longitud) y nombre del sitio, para poner un marcador sobre la posición recibida, junto al marcador se le proporciona un botón de detalles como se ve en la Figura 5.8.

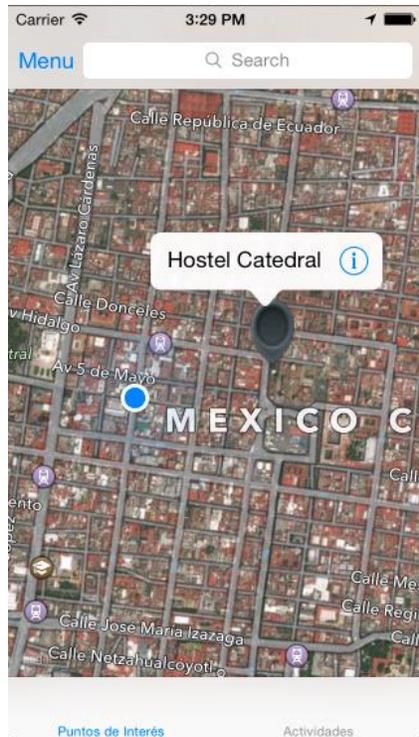


Figura 5.8 Ubicación de marcador representando un sitio.

5.6.1.2 Búsqueda por categoría de sitio

Para realizar una búsqueda por categoría, el usuario tiene que realizar un *tap* sobre la pantalla del dispositivo para desplegar un menú donde se visualizarán las categorías como se ve en la Figura 5.8, estas categorías son recuperadas de la red de ontologías por medio de una consulta al *SPARQL Endpoint*.



Figura 5.8 Categorías de la ontología.

El usuario tiene la posibilidad de buscar por una o varias categorías, seleccionando las que le interesan que sean desplegadas en la cartografía; Cuando ya han sido seleccionada(s) la(s) categoría(s), se realiza una petición al servicio Web, el cual realiza la consulta al SPARQL *Endpoint*, la cual pregunta por todas las instancias de una categoría, en el caso que sean más de una categoría, el servicio web realiza la misma consulta variando la categoría requerida y va almacenando en un vector los sitios que pertenecen a cada categoría, usando como separador el nombre de la categoría a la que pertenecen.

Después de terminar todas las consultas, se regresa un vector con todos los sitios con sus respectivas coordenadas para poder ser visualizados con un marcador personalizado por categoría sobre la cartografía como se ve en la Figura 5.9.



Figura 5.9 Despliegue de marcadores por categoría.

5.6.2 Geoprocesamiento

Para estos experimentos se hace uso del estándar *GeoSPARQL* para realizar las operaciones espaciales sobre coordenadas proporcionadas por el usuario, para esto el usuario en la menú principal debe seleccionar la opción Geoprocesamiento, para acceder a la sección que se ve en la Figura 5.10, a continuación se describen las operaciones implementadas en el sistema:

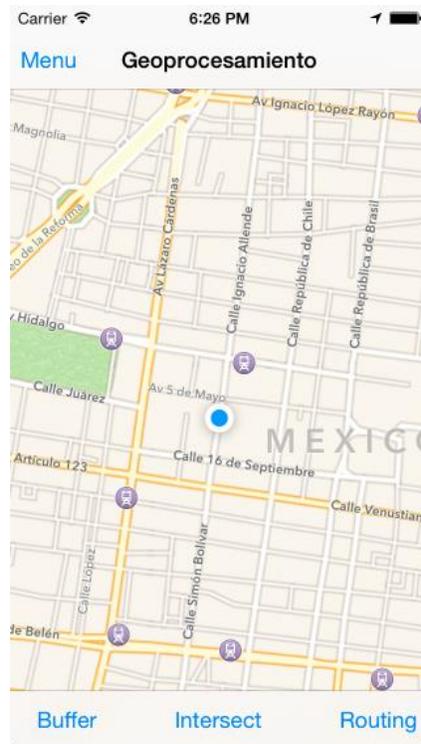


Figura 5.10 Ventana de Geoprocesamiento

5.6.2.1 Búsqueda de sitios alrededor de un lugar (Buffer)

Para realizar una búsqueda de sitios alrededor de una posición, el usuario tiene tres opciones: la primera es usar la posición actual del usuario obtenida por medio del *GPS* del dispositivo móvil, este posee un marcado circular azul, la segunda utilizar un punto de interés como posición inicial para la búsqueda el cual posee un marcado color negro y la tercera opción es crear un marcado rojo en la posición que es proporcionada por el usuario esta se creará mediante un gesto en la pantalla del dispositivo, en este caso se emplea un gesto *long press* el cual después de dos segundos creará el marcado que indique posición deseada, las tres opciones se pueden ver en la Figura 5.11.

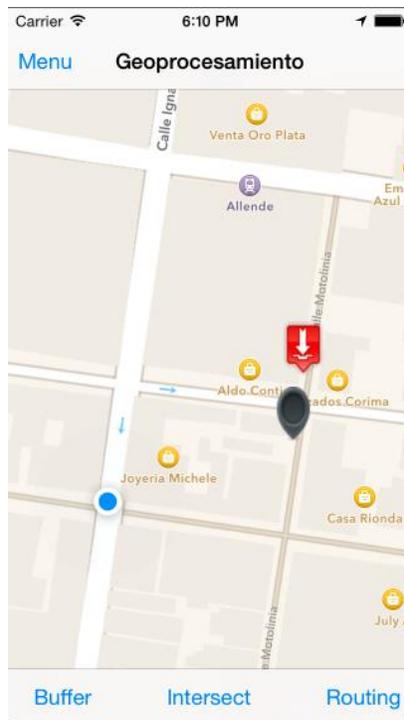


Figura 5.11 Marcadores de posición.

Después de que el usuario determina cuál de los tres marcadores (posición actual de usuario, posición de sitio de interés, posición proporcionada por el usuario) que se desplegaron en la cartografía base decide utilizar como posición para realizar la operación que el usuario *buffer*, para este ejemplo es necesario que el usuario seleccione el botón *Buffer* ubicado en la parte inferior izquierda, a continuación al ser presionado se habilita una barra deslizable donde el usuario podrá proporcionar el tamaño en metros del radio de búsqueda de sitios como se puede observar en la Figura 5.12.

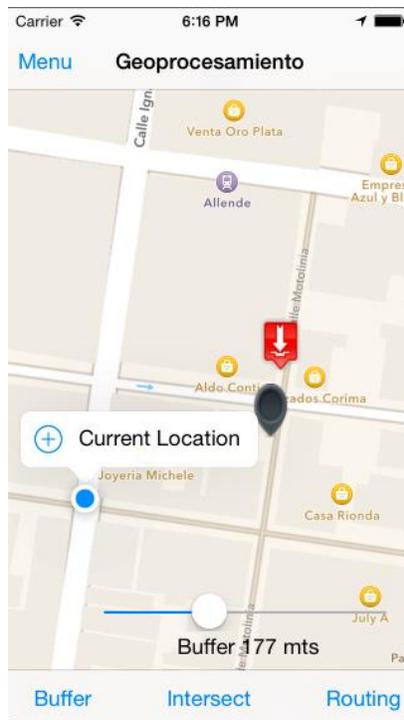


Figura 5.12 Despliegue de barra para tamaño de buffer.

Como se ve en la Figura 5.12 se decide utilizar un buffer de 177 metros, utilizando como posición inicial la posición actual del usuario, a continuación el usuario presiona el botón que se encuentra del lado derecho del nombre de la posición inicial y el sistema manda una petición al servicio Web, el cual realiza una consulta *GeoSPARQL* donde pregunta a la ontología acerca de los sitios que se encuentran dentro del buffer, de esa lista se pide recuperar el nombre así como sus coordenadas y las envía como respuesta.

Para que el sistema use un *parser* de *XML* y recupera las coordenadas para pintarlas sobre la cartografía, a la vez de realizar un auto zoom a la zona que contenga todos los marcadores recuperados y mostrarlos al usuario, como se ve en la Figura 5.13.

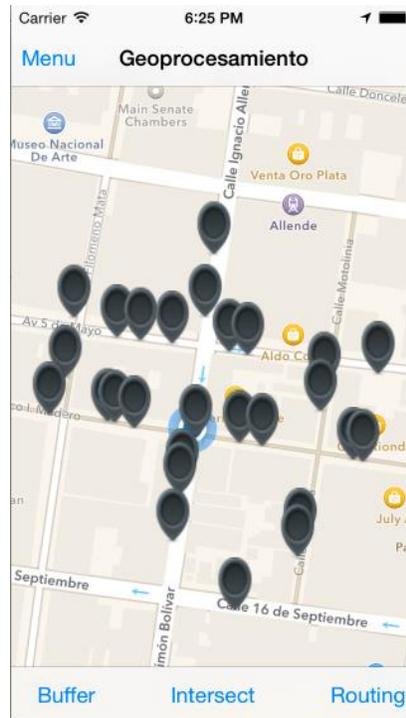


Figura 5.13 Resultados Buffer.

5.6.2.2 Búsqueda de sitios sobre una vialidad

Al igual que el experimento anterior el usuario cuenta con las tres maneras de determinar qué calle es la que se desea recuperar los sitios posicionados sobre ella. En este caso se selecciona el botón *Intersect* que se encuentra en la parte inferior central, al ser presionado el siguiente paso es ejecutar la acción mediante el botón de revelación (*disclosure*) que se encuentra a lado del nombre del marcador seleccionado como posición inicial y que servirá para determina sobre qué calle se realizará la consulta, como se ve en la Figura 5.14.

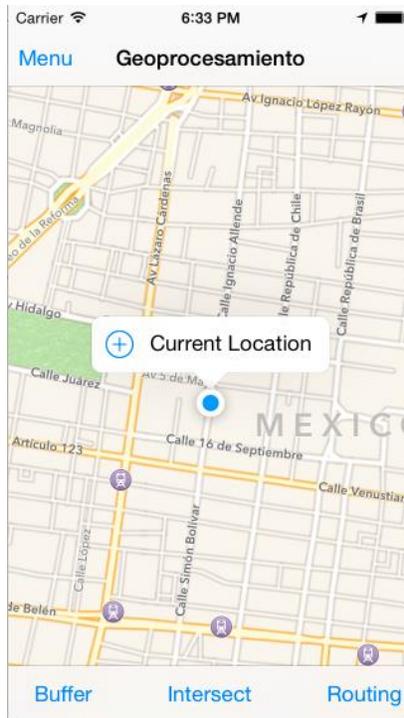


Figura 5.14 *Intersect* sobre calle del usuario.

Usando la posición de la calle donde se encuentra el usuario se determina que es muy cerca de un cruce por lo cual los sitios devueltos serán los pertenecientes a las calles que forman el cruce; Para esto se realiza una petición al servicio Web en la cual se envía la posición de referencia, procediendo a realizar una consulta en GeoSPARQL la cual tomando esa posición pregunta cuál es el nombre del segmento de calle más cercano, esto para después preguntar sobre todos los segmentos de calle que tienen el mismo nombre.

Una vez recuperados los segmentos que conforman la calle, se procede a realizar un buffer sobre estos y preguntar por los sitios que existan sobre estos segmentos, al final nos devuelve un arreglo con los nombres de los sitios y posición que se encuentran sobre toda una calle, en nuestro ejemplo nos devuelve de dos calles al ser un cruce como se ve en la Figura 5.15.

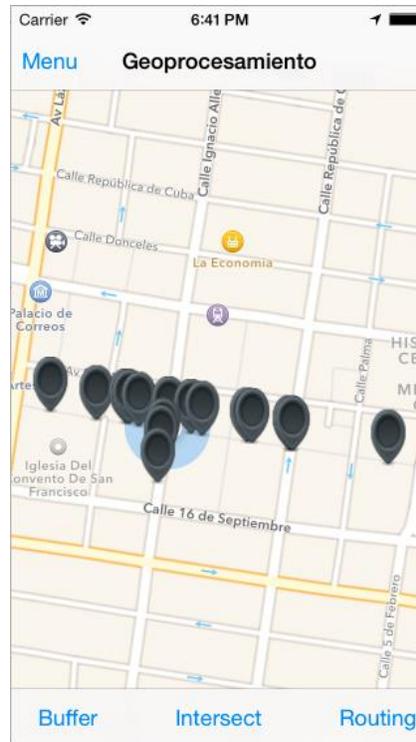


Figura 5.15 Resultado *Intersect*.

5.6.2.3 Ruta entre dos sitios

Para esta operación se requiere de la posición de inicio y de destino para realizar la ruta, al igual que en los experimentos anteriores el usuario puede seleccionar las posiciones, en este caso usaremos la posición del usuario y un marcador creado por nosotros, para esto el usuario presionara el botón *Routing* donde se desplegarán dos celdas donde se almacenarán el inicio y el destino como se ve en la Figura 5.16.

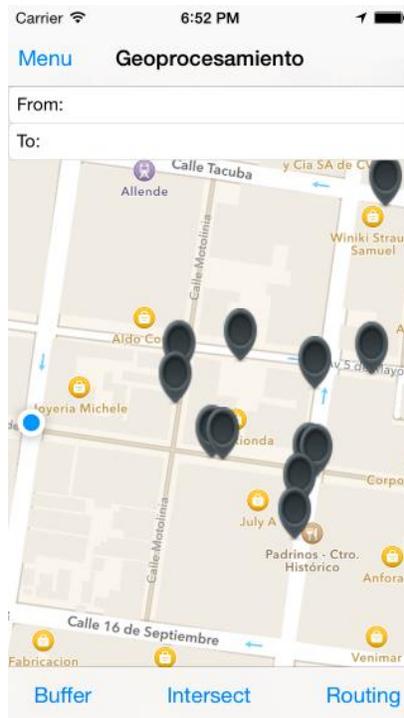


Figura 5.16 Visualización del Routing

Para determinar que sitio será origen el usuario selecciona su marcador y presiona el botón (+), a continuación se desplegará un menú emergente donde seleccionaremos si es inicio, para el destino se crea un nuevo marcador y se realiza el mismo procedimientos, solo que en vez de seleccionar inicio seleccionamos destino, cuando las dos celdas posean información se desplegará un botón en la parte superior derecha que dice mostrar ruta como se ve en la Figura 5.17.



Figura 5.17 Preparando información para *Routing*

Al pulsar el botón *Mostrar Ruta*, dado las posiciones de los marcadores seleccionados se procede a realizar una petición al servicio de ruteo de *Apple*, éste se encarga de encontrar la ruta más corta entre los sitios proporcionados y nos devuelve la línea que será trazada sobre la cartografía, la cual representa el camino que debe seguir el usuario entre dos lugares la cual como se ve en la Figura 5.18.

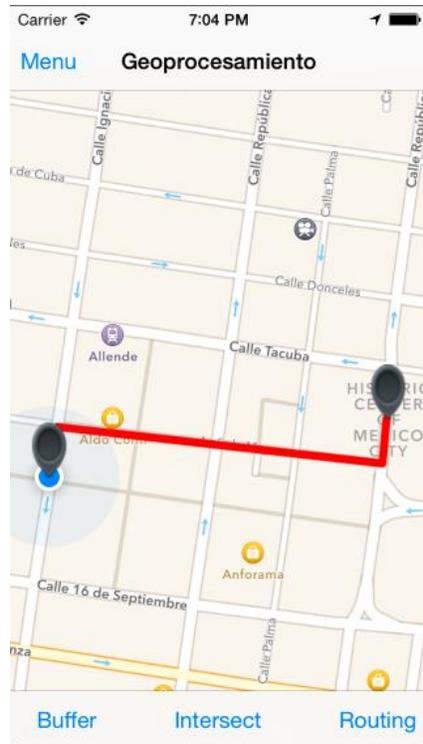


Figura 5.18 Resultado *Routing*

5.6.3 Generación de itinerario

Este experimento es el que combina varios de los experimentos por separado, para realizarlo se debe recolectar información de tres, donde se recuperan los sitios, actividades y preferencias de los usuarios para generar el itinerario con los sitios a recomendar, para acceder se selecciona en el menú principal la opción Itinerario, y nos mostrará la ventana principal del itinerario como se ve en la Figura 5.19.



Figura 5.19 Ventana principal Itinerario

En la Figura 5.19 se muestran las opciones que tiene el usuario para determinar las cosas que desea realizar, Actividades para agregar actividades que desea realizar conforme a una sugerencia; Sitios para buscar y seleccionar un sitio en específico que desee conocer y Preferencias donde se capturan el interés del usuario sobre los criterios de búsqueda. A continuación se describirá cada una de las subsecciones.

5.6.3.1 Actividades

En esta sección el usuario podrá seleccionar una actividad a la vez que desea realizar, cabe resaltar que las actividades como las características de las mismas son llenadas mediante consultas de *SPARQL* al *Endpoint*. Para el experimento el usuario querrá ir a un lugar a comer comida española y que el lugar cuente con servicio de bar, como se ve en la Figura 5.20.

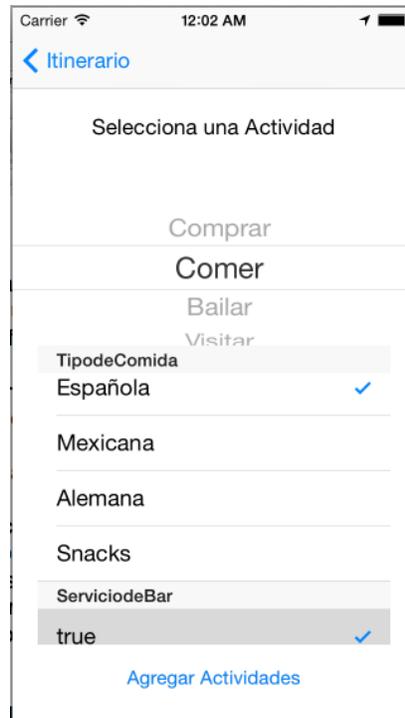


Figura 5.20 Llenado de Actividades

Para agregar la actividad se pulsa sobre el botón Agregar Actividades habiendo seleccionado la actividad y el valor de sus propiedades que se desee, para que este regrese al menú principal donde será desplegada la actividad general(En el caso del ejemplo sería la actividad de Comer) que será añadida al arreglo que controla la tabla como se ve en la Figura 5.21, además de guardar las características en un arreglo de n-dimensiones(una por cada actividad general que se vaya agregando) que será enviado junto con el arreglo de actividades generales al servicio Web.



Figura 5.21 Menú principal con Actividad

5.6.3.2 Sitios

En esta sección el usuario tiene la posibilidad de seleccionar los sitios que desea visitar, mediante una barra de búsqueda como la implementada en el experimento de búsqueda de sitio, para el experimento el usuario desea ir de compras a la tienda departamental “El Palacio de Hierro”, para después ir en búsqueda de un libro a la librería “Gandhi” como se ve en la Figura 5.22, después de seleccionar los sitios el usuario debe pulsar en el botón Agregar Sitios para guardarlos en al tabla del menú principal.



Figura 5.22 Ventana de para agregar sitios

5.6.3.3 Preferencias

En esta sección se le pregunta al usuario que tanto interés tiene acerca de la cercanía de los lugares entre ellos y las evaluaciones que usuarios de las redes sociales hacen acerca de ellos, para esto se implementaron dos barras con una escala de gusto de cinco escalones, al momento de agregar las preferencias se regresan los valores numéricos para ser enviados después, cabe mencionar que las preferencias se usan únicamente para realizar las búsquedas de los sitios donde se llevan a cabo las actividades . Para el experimento el usuario le importa la cercanía entre los lugares y le interesa mucho la opiniones de otros usuarios acerca de los lugares como se ve en la Figura 5.23.

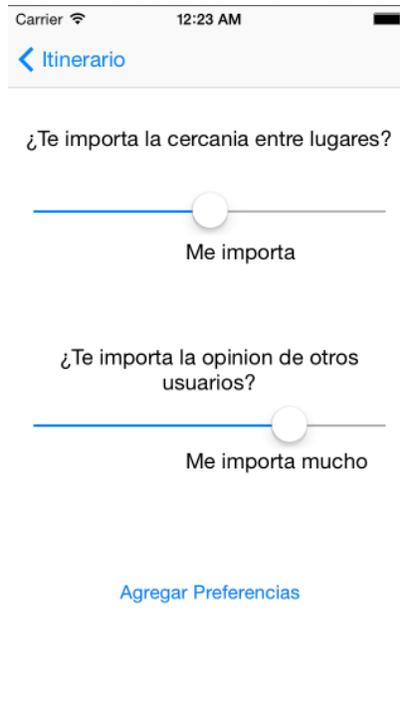


Figura 5.23 Ventana para preferencias.

El itinerario puede ser generado a partir de llenar las secciones de Preferencias, Actividades y/o Sitios, y estos pueden ser modificados en el menú principal, ya sea de orden u borrándolos según sea el caso, una vez definidos los lugares y actividades a realizar, al igual que sus preferencias como se ve en la Figura 5.24, se puede proceder a generar el itinerario personalizado del usuario mediante una petición al servicio Web.

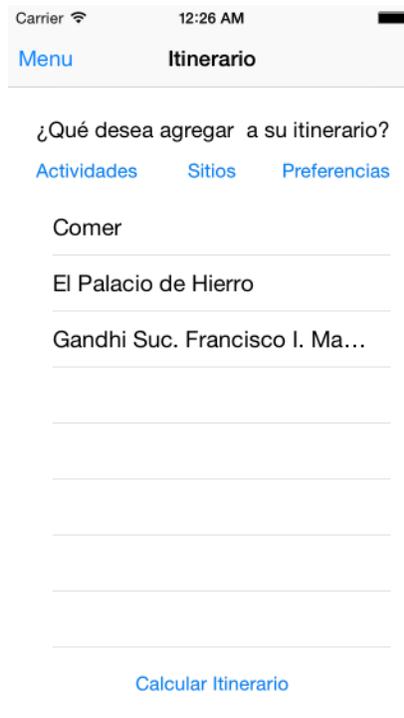


Figura 5.24 Perfil de usuario completo.

Cuando el usuario pulsa el botón Calcular Itinerario, la lista de actividades y sitios, así como sus preferencias y posición actual, son mandados al servicio Web para ser procesados, este dividirá el vector para extraer las variables y poder realizar las consultas correspondientes, existen dos casos para armar el itinerario, el primero son los sitios, donde se pregunta su nombre así como sus coordenadas y se almacenan, el segundo son las actividades a realizar, en este caso existen tres escenarios posibles para realizar esta consulta.

El primer escenario es cuando el usuario prefiere la cercanía entre lugares que la opinión de usuarios, para esto se implementa una distancia y rating dinámicos, esto con el fin de realizar una consulta flexible de tal manera que siempre nos regrese un resultado. Para este escenario se varía los ratings de tal forma que al llegar a la peor calificación registrada, en caso de no regresar ningún

resultado se procede a aumentar la distancia y a restablecer al valor establecido por el usuario. Esto hasta que se encuentra un solo resultado.

El segundo escenario es cuando el usuario prefiere las opiniones de usuarios sobre la cercanía entre los lugares, al igual que en el primer escenario se implementa distancias y ratings dinámicos. Para este escenario se varía la distancia hasta 900 metros, mientras que el rating se mantiene fijo, en caso de que la consulta no regrese algún resultado, se disminuye el valor del rating y se restablece el valor asignado por el usuario a la distancia.

El tercer escenario se divide en dos casos, cuando el usuario desea que se le recomiende lugares con muy buena calificación y cercanos y el segundo cuando al usuario no le interesa ni la calificación ni la cercanía en absoluto. Cuando el primer caso sucede, se toma la decisión de darle un mayor peso a la cercanía que a las opiniones de los usuarios, esto debido a que no existe recomendación alguna que cumpla ambos criterios a plenitud, dado que cuando se consulta por los lugares más cercanos, no siempre son los que tienen la mejor calificación y viceversa cuando se consulta por los lugares con mejor rating. Para el segundo caso se procede a recomendar los primeros lugares que cumplan con las características que el usuario requiere, sin importar su rating o su cercanía.

Estos escenarios nos arrojan como resultado siempre un sitio, en dado caso que el usuario pida un lugar donde la combinación de las propiedades nos retornen un conjunto vacío al realizar la consulta, se procede a descartar las propiedades con valores booleanos, dejando las características que nos ofrezcan una mejor descripción del lugar, por ejemplo si un usuario pide un sitio para ir a comer que sirva comida alemana y que tenga servicio de cafetería incluido, en caso de no existir una instancia con estas propiedades, se procede a solo tomar en cuenta el tipo de comida que ofrezca el lugar para la recomendación.

Una vez completado el proceso, el servicio Web devuelve como resultado un arreglo con los nombres y posiciones de los sitios que se recomendarán junto con lo que da el usuario. La respuesta es procesada por el sistema y genera los marcadores de los sitios como guía base, después realizar la petición al servicio de ruteo de Apple para determinar los segmentos de ruta por los cuales debe de transitar el usuario como se ve en la Figura 5.25.

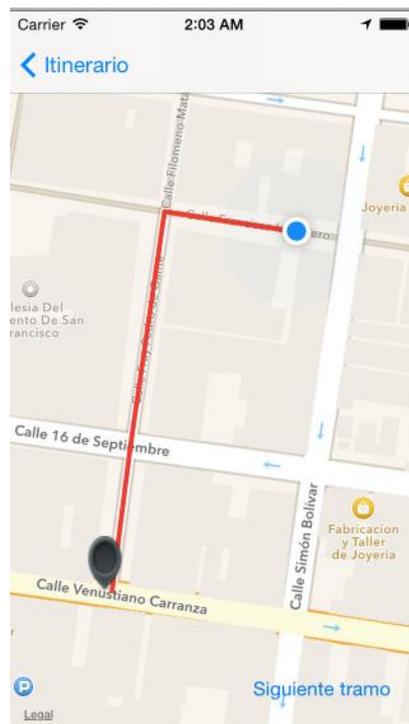


Figura 5.25 Resultado del itinerario turístico personalizado

Como se pudo ver a través de los experimentos realizados a la metodología propuesta se obtuvieron resultados satisfactorios con respecto a la aplicabilidad de la misma, haciendo uso de *Linked Data* para el enriquecimiento de los datos, así como la recomendación de sitios por medio de un perfil de usuario. A parte de poder implementar el estándar de *GeoSPARQL* para realizar las operaciones de geoprocesamiento necesario para proporcionarle al usuario los sitios más apegados a sus preferencias.

Capítulo 6 Conclusiones y Trabajo a futuro

En este trabajo se propuso crear una metodología que sea capaz de generar itinerarios turísticos personalizados, mediante el uso de un perfil de usuario y ponderaciones proporcionadas por el mismo, para realizar recomendación más precisa sobre sitios de interés que pueda visitar. La metodología es capaz de recuperar información de la Web mediante consultas a los *SPARQL Endpoint* de fuentes de datos externas, esto por la implementación de los principios de *Linked Data* en la metodología.

La metodología es capaz de eliminar algunas limitantes de los sistemas de recomendación como el *cold start* y la *dispersión*, el primero se resuelve gracias a la implementación de un perfil de usuario que recopila las actividades y/o sitios que desea realizar en su estadía, así como factores de ponderación como son la distancia entre sitios de interés o evaluaciones de los sitios a recomendar conforme a otros usuarios, mientras el segundo se resuelve con el uso de redes sociales de donde se recuperan las valoraciones hechas por todos sus usuarios, por tal motivo ya existen evaluaciones con las cuales poder realizar una buena sugerencia.

Otro de los alcances realizados es el enriquecimiento del dato original, que nos da la capacidad de poder recuperar información adicional de nuestros datos, esto gracias a *Linked Data* podemos recuperar información de diferentes fuentes de información y extraer datos para complementar ya sea una descripción o imágenes del lugar, o nuevos sitios de interés que pueden ser incorporados a nuestro *SPARQL Endpoint*.

También se logró transformar los datos originales (*shapefiles*) para implementar la integración de la especificación del *Open Geospatial Consortium GeoSPARQL* sobre un sistema de recomendación desarrollado para dispositivos móviles, usando una red de ontologías que poseen una componente geográfica. Esto se hizo posible gracias a su almacenamiento en un *triple store*, eliminando el uso de una Base de Datos Espacial (BDE), y generando un *SPARQL Endpoint* de consultas, sobre el cual se realizan las búsquedas sobre la red de ontologías además de realizar las operaciones espaciales que requiera el servidor.

En el marco de este trabajo se aplicaron los principios de *Linked Data* para generar y publicar información turística y geoespacial conforme a dichos principios sobre las instancias pertenecientes a la red de ontologías, con el fin de proporcionarle al usuario una mayor cantidad de información relacionada al sitio turístico de su interés.

En cuanto a las operaciones espaciales que permite realizar el estándar de *GeoSPARQL*, se tuvo que realizar previamente un tratamiento y transformación de los datos originales de las vialidades del área de estudio y los sitios de interés que se encuentran ambos en archivos con formato *shapefiles* y que se transforman a un archivo de tripletas con formato *RDF*. Para poder realizar dicha transformación se realizó un programa desarrollado en Java, el cual toma la información del objeto espacial junto a su componente geométrica para generar los archivos *RDF* y almacenarlos en el *SPARQL Endpoint* para su posterior explotación utilizando el estándar *GeoSPARQL*.

6.1 Trabajo a futuro

Como líneas de trabajo futuro para dar continuidad a este trabajo se propone abordar los siguientes tópicos:

- Análisis de sentimientos provenientes de información disponible en redes sociales, con el fin de mejorar las evaluaciones de los sitios de interés.
- Implementación de una medida de similitud semántica para realizar una mejor recomendación de sitios de interés, con el fin de aumentar la precisión en el caso de que no exista un sitio con las características exactas que pide un usuario.
- Incorporación de información de contexto en el sistema de recomendación con el objetivo de incrementar y mejorar la información facilitada al usuario.
- Desarrollo de algoritmos para el aumento de la cantidad de conjuntos de datos conectados geoespacialmente en la nube de *Linked Data*.

Bibliografía

Adomavicius, G.; Tuzhilin, A., (2005) *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions*, IEEE Transactions on Knowledge and Data Engineering, 17(6) (pp. 734–749).

Adomavicius, G.; YoungOk Kwon, (May-June 2007) *New Recommendation Techniques for Multicriteria Rating Systems*, Intelligent Systems, IEEE , 22(3), (pp. 48-55).

Ahmad, A. Kardan; Mahnaz, E., (January 2013) *A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups*, Information Sciences, 219(10), (pp. 93-110).

Allocca, C.; D'Aquin, M.; Motta, E., (2009) *DOOR - Towards a Formalization of Ontology Relations*. In Jan L. G. Dietz, editor, KEOD, (pp. 13–20).

Ashrafal A.; Latifur K.; Bhavani M., (2011) *Thuraisingham: Geospatial Resource Description Framework (GRDF) and security constructs*. Computer Standards & Interfaces, 33(1), (pp. 35-41).

Auer, S., Lehmann, J., Ngomo, A. C. N., & Zaveri, A. (2013). Introduction to linked data and its lifecycle on the web. In *Reasoning Web. Semantic Technologies for Intelligent Data Access* (pp. 1-90). Springer Berlin Heidelberg.

Batet, M.; Moreno, A.; Sánchez, D.; (15 June 2012) Isern, D.; Valls, A.; *Turist@: Agent-based personalised recommendation of tourist activities*, *Expert Systems with Applications*, 39(8), (pp. 7319-7329).

Becker, C.; Bizer, C., (2008). *DBpedia Mobile: A Location-Enabled Linked Data Browser*. *LDOW*, 369.

Berners, Tim.; Fielding. R.; Masinter L., (August 1998) *RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax*.

Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A., (July 2013) *Recommender systems survey*, *Knowledge-Based Systems*, 46(), (pp. 109-132).

Buján, D.; Martín, D.; Torices, O.; López, D.; Lamsfus, C.; Abaitua, J.; & Alzua, A., (2013) *Context Management Platform for Tourism Applications*. *Sensors*, 13(7), (pp. 8060-8078.)

Buriano, L.; Marchetti, M.; Carmagnola, F.; Cena, F.; Gena, C.; Torre, I., (10-12 May 2006) *The Role of Ontologies in Context-Aware Recommender Systems*, *Mobile Data Management, 2006. MDM 2006. 7th International Conference on* (pp.80).

Burke, R., (2002) *Hybrid recommender systems: survey and experiments*, *User Modeling and User-Adapted Interaction*, 12 (4), (pp. 331–370).

Candillier, L.; Meyer, F.; Boullé, M.; (2007) *Comparing state-of-the-art collaborative filtering systems*, *Lecture Notes in Computer Science*, 4571, (pp. 548–562).

Corcho, Ó.; Fernández, M.; Gómez, A.; López, A., (2005) *Construcción de ontologías legales con la metodología METHONTOLOGY y la herramienta WebODE*.

Fernández, M.; Gómez, A.; Juristo, N., (1997) *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*, AAAI Symposium on Ontological Engineering, Stanford.

Fonseca, F.; Egenhofer, M.; Agouris, P.; Câmara, C., (2002) *Using ontologies for integrated geographic information systems*. Transactions in GIS 6(3), (pp. 231-257).

Fonseca, F., Câmara, G., Monteiro, A. M. (2006) *A Framework for Measuring the Interoperability of Geo-Ontologies*, Spatial Cognition and Computation, vol. 6, no. 4, (pp. 307-329).

García, A.; Chamizo, J.; Rivera, I.; Mencke, M.; Colomo, R.; Gómez, J.M., (August 2009) *SPETA: Social pervasive e-Tourism advisor*, Telematics and Informatics, 26(3), (pp. 306-315).

García, Á.; López, J.L.; Colomo, R.; González, I.; Ruiz, B.; (15 September 2011) *Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain*, Expert Systems with Applications, 38(10), (pp. 13310-13319).

Golemati, M.; Katifori, A.; Vassilakis, C.; Lepouras, G.; Halatsis, C., (2007) *Creating an Ontology for the User Profile: Method and Applications*, First IEEE International Conference on Research Challenges in Information Science (RCIS), Morocco.

Gruber, T., (1995) *Towards principles for the design of ontologies used for knowledge sharing*, International Journal of Human-Computer Studies, 43(5/6), (pp. 907-928).

Guarino, N., (1995) *Formal ontology, conceptual analysis and knowledge representation*, International Journal of Human-Computer Studies, 43(5- 6), (pp. 625-640).

Gutiérrez Losada, I. (2010). *Ontologías Turísticas Geográficas: creación de una ontología sobre rutas turísticas (a Pie o en Bicicleta) por espacios naturales*.

Guzmán, J. A.; López, M.; Torres, I. D., (Mayo 2012) *Metodologías y métodos para la construcción de ontologías*. Scientia et Technica, [S.l.], 2(50), (pp. 133-140).

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), (pp. 1-136).

Hi-Touch Working Group (2003) *Semantic web methodologies and tools for intra-European sustainable tourism*.

Höpken, W.; Clissmann, C., (2006) *Harmo-TEN - Tourism Harmonisation Trans-European Network - Final Ontology Report*.

Hyung, J.A., (2 January 2008) *A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem*, Information Sciences, 178(1), (pp. 37-51).

Klien, E.; Probst, F., (2005) *Requirements for geospatial ontology engineering*. Conference on Geographic Information Science (AGILE 2005) Estoril (Portugal).

Lacasta, J.; Muro-Medrano, R.; Nogueras, J.; Zarazaga, F., (2005) *Web Ontology Service, a key component of a Spatial Data Infrastructure*. Proceedings of the 11th EC GI&GIS Workshop: ESDI: ESDI Setting the Framework.

Linden, G.; Smith, B.; York, J., (Jan/Feb 2003) *Amazon.com recommendations: item-to-item collaborative filtering*, Internet Computing, IEEE , 7(1), (pp.76,80).

Lutz, M.; Sprado, J.; Klien, E.; Schubert, C.; Christ, I., (2009) *Overcoming semantic heterogeneity in spatial data infrastructures*. Computers & Geosciences, 35(4), (pp. 739-752).

Mao Y.; Peifeng Y.; Wang-Chien L., (2010) *Location recommendation for location-based social networks*, In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '10), (pp. 458-461), New York, NY, USA.

Middleton, S. E.; Roure, D. D.; Shadbolt, N. R., (2009) *Ontology-based recommender systems*. In S. Staab and R. Studer, editors, Handbook on Ontologies, International Handbooks Information System, (pp.779–796), Springer Berlin Heidelberg,

Moreno, A.; Valls, A.; Isern, D.; Marin, L.; Borràs, J., (January 2013) *SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities*, Engineering Applications of Artificial Intelligence, 26(1), (pp. 633-651).

Neches, R., Fikes, R.E., Finin, T., Gruber, T.R., Patil, R., Senator, T. & Swartouy, W.R., (1991) *Enabling technology for knowledge sharing*, AI Magazine, 12(3), (pp. 16-36).

Parra, G.; Klerkx, J.; Duval, E., (2010). *More!: Mobile Interaction with Linked Data*.

Pazzani, M., (Dec. 1999) *A Framework for Collaborative, Content-Based, and Demographic Filtering*, Artificial Intelligence Rev., (pp. 393-408).

Ressler, J.; Dean, M., (2007) *Geospatial Ontology Trade Study*. Proceedings of the Second International Ontology for the Intelligence Community Conference OIC-2007.

Rich, E., (1983) *Users are individuals: individualizing user models*, International Journal of Man-machine Studies, 18(3), (pp. 199-214).

Roche, C., (2003) *Ontology : a survey*, in: 8th Symposium on Automated Systems Based on Human Skill and Knowledge, (pp. 28–41), Gteborg, Sweden.

Ruotsalo, T.; Haav, K.; Stoyanov, A.; Roche, S.; Fani, E.; Deliai, R.; Mäkelä, E.; Kauppinen, T.; Hyvönen, E., (May 2013) *SMARTMUSEUM: A mobile recommender system for the Web of Data*, Web Semantics: Science, Services and Agents on the World Wide Web, 20, (pp. 50-67).

Russomanno, D.; Kothari, C.; Thomas, O., (2005) *Building a Sensor Ontology: A Practical Approach Leveraging ISO and OGC Models*, The 2005 International Conference on Artificial Intelligence, Las Vegas, NV, (pp. 637-643).

Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J., (2000) *Application of Dimensionality Reduction in Recommender Systems—A Case Study*, Proc. ACM WebKDD Workshop.

Sarwar, B. M.; Karypis, G.; Konstan, J.; Riedl, J., (2002, December). *Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering*. In Proceedings of the fifth international conference on computer and information technology, 1().

Sarwar, B.; Karypis, G.; Konstan, J.; Reidl, J., (2001) *Item-Based Collaborative Filtering Recommendation Algorithms*. In Press, A. (Ed.).

Schafer, J.B.; Frankowski, D.; Herlocker, J.; Sen, S., (2007) *Collaborative filtering recommender systems*, The Adaptive Web, (9), (pp. 291–324).

Seok, K. L.; Yoon, H. C.; Soung, H. K., (1 June 2010) *Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations*, Information Sciences, 180(11), (pp. 2142-2155).

Smith, B.; Mark, M. (1998) *Ontology and Geographic Kinds*. In T. K. Poiker and N. Chrisman (eds.), Proc. 8th International Symposium on Spatial Data Handling (SDH'98), Vancouver: International Geographical Union, (pp. 308-320).

Steiniger, S.; Hunter, A.J.S., (May 2013) *The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption*, Computers, Environment and Urban Systems, 39(), (pp. 136-150).

Suárez, M. C.; Gómez, A.; Fernández, M., (2012). *The NeOn Methodology for Ontology Engineering*. In *Ontology Engineering in a Networked World*, (pp. 9-34), Springer Berlin Heidelberg.

Sure, Y.; Staab, S.; Studer, R., (2003) *On-to-knowledge methodology*, Handbook on Ontologies, Series on Handbooks in Information Systems, 6, (pp. 117-132).

Sure, Y.; Bloehdorn, S.; Haase, P.; Hartmann, J.; Oberle, D., (2005) *The SWRC Ontology - Semantic Web for Research Communities*. In Proc. 12th Portuguese Conf. on Artificial Intelligence. LNAI, 3808, (pp. 218—231). Springer Berlin, Heidelberg.

Uschold, M.; Gruninger, M., (1996) *Ontologies: principles, methods and applications*. The Knowledge Engineering Review, 11, (pp. 93-136).

Vilches Blázquez, L.M.; Cañete Pérez, J.A.; Corcho, O.; Bernabé Poveda, M.A., (2008) *Interrelaciones entre las tecnologías de la información geográfica y la ingeniería ontológica para la mejora de la gestión de los recursos geoespaciales*. Actas del XIII Congreso Nacional de Tecnologías de la Información Geográfica. Hernández Calvento, L.; Parreño Castellano, J.M. (eds.). Las Palmas de Gran Canaria. ISBN: 978-84-96971-53-0

Vilches-Blázquez, L. M. (2011). *Metodología para la integración basada en ontologías de información de bases de datos heterogéneas en el dominio hidrográfico* (Doctoral dissertation, Topografía).

Vilches-Blázquez, L. M., Villazón-Terrazas, B., Corcho, O., & Gómez-Pérez, A. (2013). Integrating geographical information in the Linked Digital Earth. *International Journal of Digital Earth*, (ahead-of-print), (pp. 1-22).

Virrantaus, K.; Markkula, J.; Garmash, A.; Terziyan, V.; Veijalainen, J.; Katanosov, A.; Tirri, H., (3-6 Dec 2001) *Developing GIS-supported location-based services*, Web Information Systems Engineering, 2001. Proceedings of the Second International Conference on , 2(), (pp. 66-75)

Wan-Shiou, Y.; San-Yih, H., (January 2013) *iTravel: A recommender system in mobile peer-to-peer environment*, Journal of Systems and Software, 86(1), (pp. 12-20).

Wöber, K. (2003) *Information supply in tourism management by marketing decision support systems*. Tourism Management. 24(3), (pp. 241-255).

Zadeh, L., (1965) *Fuzzy sets*. *Information & Control*, 8(3), (pp. 338–353).

Zablith, F.; Fernandez, M.; Rowe, M., (2011) *The OU linked open data: production and consumption*. In Proceedings of the 1st International Workshop on eLearning Approaches for the Linked Data Age (Linked Learning) at ESWC, Heraklion, Greece.