

**Instituto Tecnológico de Orizaba
División de Estudios de Posgrado e
Investigación
Maestría en Sistemas Computacionales**

**Desarrollo de un Sistema de recomendación geoespacial de
puntos de venta**

PRESENTADO POR:

Edith Verdejo Palacios M09011386

DIRECTOR

Dr. Giner Alor Hernández.

2016, Orizaba, Veracruz, México

Contenido

Índice de figuras.....	iv
Índice de tablas.....	v
Abstract	1
Resumen.....	2
Introducción	3
Capítulo 1. Antecedentes.....	4
1.1. Marco Teórico	4
1.1.1. Sistema de información geográfica (SIG).....	4
1.1.1.1 Clasificación de los SIG.	5
1.1.1.2 Características de los SIG.....	6
1.1.1.3 Software para SIG.....	6
1.1.2. Sistemas de Recomendación	7
1.1.3. Sistemas de Geo-recomendación.....	16
1.1.3.1.1. Herramientas de Geo-recomendación	17
1.1.3.1.2. Herramientas para el desarrollo de sistemas de recomendación.....	17
1.1.3.1.2.1. Apache Mahout	18
Características de Apache Mahout.....	18
Algoritmos que soporta.....	18
1.1.3.1.3. APIs de Geo-localización	20
1.1.3.1.3.1. Google™ Maps	20
1.1.3.1.4. INEGI.....	21
1.1.3.1.4.1. INEGI DENUÉ	21
1.2. Planteamiento del problema.....	21
1.2. Objetivo general y específicos	23
1.2.1. Objetivo general	23
1.2.2. Objetivos específicos	23
1.3. Justificación	23
Capítulo 2. Estado de la práctica.....	25
2.1. Estado de la práctica	25
2.1.1. Sistemas de recomendación aplicados en diferentes dominios.....	25

2.1.2.	Sistemas de información geográfica aplicados en estudios ambientales y urbanos....	28
2.2.	Propuesta de solución.....	36
2.2.1.	Alternativas de solución.....	36
2.2.2.	Solución propuesta.....	37
2.2.3.	Justificación de la solución propuesta.....	37
Capítulo 3.	Aplicación de la metodología.....	38
3.1.	Aplicación del proceso de fases de UWE para el diseño y desarrollo de un sistema de geo-recomendación.....	38
3.2.	Diseño de un sistema de geo-recomendación.....	38
3.3.	Arquitectura del sistema de geo-recomendación.....	44
3.3.1.	Descripción de las capas de la arquitectura.....	46
3.3.1.1.	Descripción de componentes.....	48
3.3.2.	Funcionamiento de la arquitectura.....	48
3.4.	Esquema relacional de la base de datos.....	51
3.5.	Identificación de fuentes de información para el proceso de geo-recomendación.....	53
3.6.	Métricas de similitud para el proceso de geo-recomendación.....	54
3.6.1.	Coseno Ajustado.....	54
3.6.2.	Coficiente de Correlación de Pearson.....	55
3.6.3.	Coseno.....	55
3.7.	Proceso de Geo-recomendación.....	56
3.8.	Diseño del sistema de Geo-recomendación.....	59
Capítulo 4.	Resultados.....	59
4.1.	Caso de estudio en la recomendación de puntos de venta.....	59
4.2.	Pruebas del sistema.....	65
Capítulo 5.	Conclusiones y recomendaciones.....	74
5.1	Conclusiones.....	74
5.2	Recomendaciones.....	75
Productos Académicos	76
Bibliografía	77

Índice de figuras

Figura 1. 1 Definición de SIG ¹	4
Figura 1. 2 Modelos ¹	5
Figura 1. 3 Representación de un sistema de recomendación	7
Figura 1. 4 Clasificación de los sistemas de recomendación	8
Figura 1. 5 Basado en contenido	8
Figura 1. 6 Filtrado Colaborativo ⁷	9
Figura 1. 7 Filtrado Demográfico.....	9
Figura 1. 8 Basados en conocimiento ⁷	10
Figura 1. 9 Recomendación Híbrida ⁷	10
Figura 1. 10 Características de los sistemas de recomendación	12
Figura 1. 11 Arquitectura Apache Mahout tomada del sitio de ibm	19
Figura 3. 1. Diagrama de Casos de Uso	39
Figura 3. 2 Diagrama de Clases (modelo).....	40
Figura 3. 4 Diagrama de Clases (Vista)	41
Figura 3. 6 Modelo de Presentacion (Geo-recomendación).....	43
Figura 3. 7 Arquitectura del sistema de geo-recomendación.	45
Figura 3. 8 Diagrama de actividades del sistema	49
Figura 3. 9 Esquema de Base de Datos.	52
Figura 4. 2 Formulario de Búsqueda.	61
Figura 4. 3 Resultados.....	62
Figura 4. 4 Resultados Zonas Óptimas.....	63
Figura 4. 5 Detalle de la zona.....	64
Figura 4. 7 Retroalimentación por usuario.....	71
Figura 4. 8 Modelo.....	72

Índice de tablas

Tabla 1. 1 Definición de características	13
Tabla 2. 1 Tabla comparativa	¡Error! Marcador no definido.
Tabla 2. 2 Alternativas de solución	37
Tabla 4. 1 Características de los sistemas de recomendación	67
Tabla 4. 2 Valores normalizados	68
Tabla 4. 4 Características	69

Abstract

Nowadays, it is more common to offer recommendations to users on services or products. Many companies have implemented recommendation mechanisms and they study the preferences and interests of the users in such a way that the services of the companies are adapted to the users' needs. From another perspective, companies have the constant worry about the total of income and expenditures, which is related to the location of the business.

In this work, the principal aim is to develop a recommender system for the ideal location of commercial establishments through Apache Mahout with the integration of geographical models.

Through of the development of the recommender system, we have geographically analyzed the situation of the company taking into account parameters such as location, competition, clients, among others. The result of this development allows to know better the clients, to locate to the potential and sure market, to optimize investment and contributes in the field of the studies of geographical integrations with recommender systems.

Resumen

Actualmente en la sociedad cada vez es más conveniente ofrecer recomendaciones a los usuarios sobre servicios o productos. Dicho motivo permite que muchas compañías implementen mecanismos de recomendación a sus sistemas de manera que se estudien los gustos e intereses de los usuarios de tal forma que los servicios de las compañías sean adecuados a las necesidades de los usuarios. Visto desde otro punto las empresas tienen la constante preocupación del total de ingresos y egresos, que en medida se ve relacionado con la ubicación del negocio.

En este trabajo el objetivo principal es el desarrollo de un sistema de recomendación para la ubicación óptima de establecimientos comerciales a través de Apache Mahout con la integración de modelos geográficos.

Mediante el desarrollo del sistema de recomendación analizó geográficamente la situación de la empresa tomando en cuenta ciertos parámetros como ubicación, competencia, clientes, entre otros. El resultado de este desarrollo permite conocer mejor a los clientes, localizar al mercado potencial y seguro, optimizar inversión y contribuye en el campo de los estudios de integraciones geográficas con sistemas de recomendación.

Introducción

La gran cantidad de negocios existentes ha hecho que éstos sean más competitivos en la forma de diseñar estrategias para captar un mayor número de clientes. La utilización de modelos geográficos con integración del entorno socioeconómico ha conformado un nuevo punto de vista para el estudio del mercado que no ha explotado lo suficiente por las empresas.

El presente trabajo plantea el problema de selección óptima que enfrentan las empresas para la ubicación de sus instalaciones, así como la incertidumbre que les ocasiona saber si la ubicación influye en el éxito de sus negocios.

Por lo tanto este trabajo propone abordar el problema desde el punto de vista de los sistemas de recomendación con la integración de modelos geográficos y determinados factores como la proximidad de los posibles clientes, distancia del local, niveles socioeconómicos de la zona, grado de afluencia, nivel de competitividad, ubicación de servicios complementarios, entre otros.

El documento está organizado de la siguiente manera:

En el primer capítulo se presentan los conceptos utilizados a lo largo del documento, el planteamiento del problema, la justificación y los objetivos a perseguir.

En el segundo capítulo se presenta un estudio y análisis de los trabajos existentes y relacionados con el trabajo presente.

En el tercer capítulo se describe la arquitectura del sistema, presentando cada una de las capas y módulos que la conforman.

Dentro del cuarto capítulo se ejemplifican el caso de estudio que describe la utilidad del sistema. Finalmente en el capítulo cinco se presentan las conclusiones, productos académicos y recomendaciones.

Capítulo 1. Antecedentes

En este capítulo, se dan a conocer los conceptos que se utilizan a lo largo de la tesis, como los Sistemas de Información Geográfica (SIG), Sistemas de Recomendación (SR) y Sistemas de Geo-recomendación así como las características y aplicación de los conceptos dentro del problema a resolver, los objetivos específicos y generales; y finalmente la justificación del mismo.

1.1. Marco Teórico

1.1.1. Sistema de información geográfica (SIG)

Son sistemas diseñados para almacenar, recuperar, manipular, analizar y correlacionar datos geográficos. El elemento central de un SIG es el uso de un sistema de localización de referencias para que los datos sobre una ubicación específica se analicen en su relación con otros lugares [1]; esta habilidad los distingue de otros sistemas de información ya que permiten realizar planes y estrategias geográficas. Estos sistemas son capaces de trabajar con información organizada en bases de datos y georreferenciar cada unidad de análisis espacial a través de sus coordenadas geográficas en mapas. La figura 1.1 representa la definición de los SIG.

1

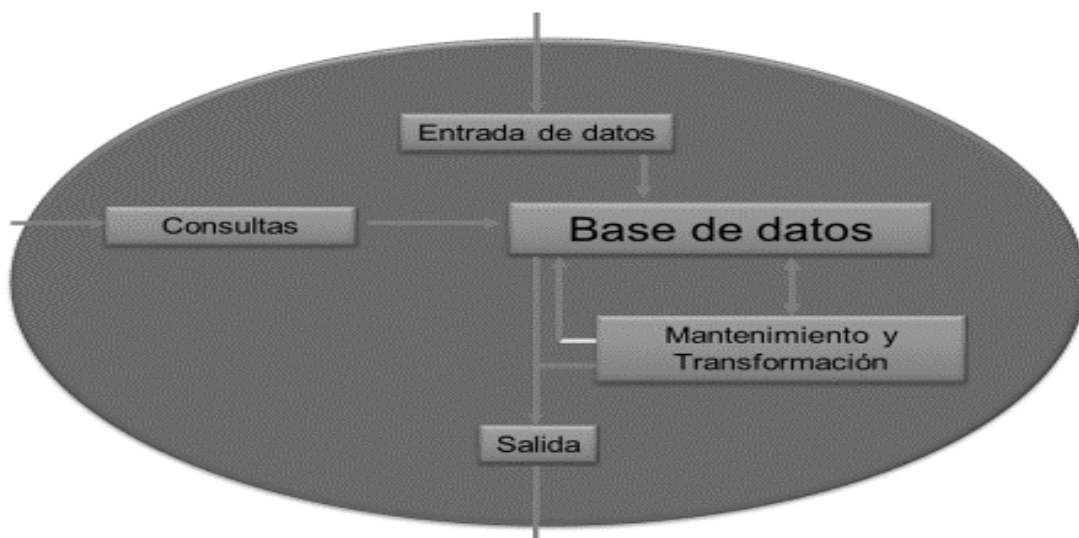


Figura 1. 1 Definición de SIG ¹

¹ goo.gl/55KOHA

Sin embargo, es capaz de ofrecer mucho más poder que un sistema automatizado de cartografía, por sus funciones analíticas y la importancia de las relaciones espaciales entre objetos para asignar atributos [2].

1.1.1.1 Clasificación de los SIG.

Existen dos formatos básicos para el almacenamiento y procesamiento de datos de coordenadas geográficas: vector y ráster. El método vector almacena las ubicaciones de elementos geográficos utilizando el sistema de coordenadas X, Y; en cambio con el método ráster los datos se almacenan en una matriz de píxeles, los puntos se representan como un píxel, las líneas como una serie de píxeles contiguos y las áreas como grupos de píxeles [3]. La figura 1.2 muestra cómo se visualizan ambos modelos.

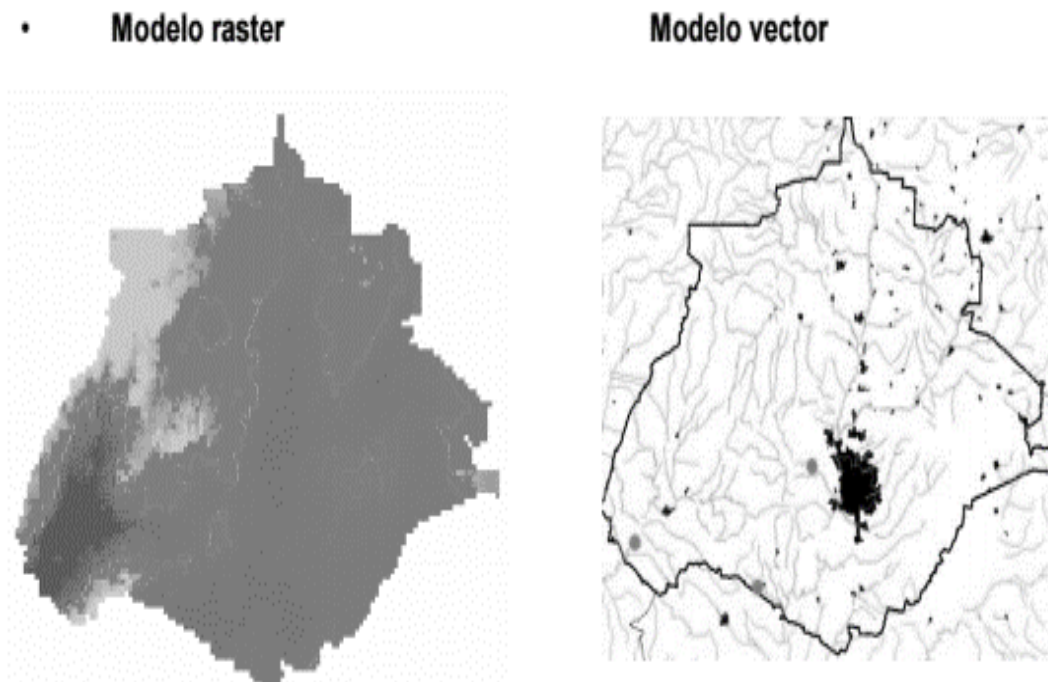


Figura 1. 2 Modelos¹

1.1.1.2 Características de los SIG

Los sistemas de información geográfica generalmente cuentan con las siguientes características:

1. Permiten organizar la información geográfica en capas, ya que modelan el espacio geográfico en conjuntos de datos denominados capas de información.
2. Posibilitan el análisis estadístico con respecto a la clase de objeto y la información de sus atributos [4].
3. La superposición implica la unión de datos vectoriales o de mapa de bits a partir de dos conjuntos de datos.
4. Permiten el análisis multi-variable, el cual es un análisis de información de operaciones geográficas que facilita permutaciones de combinaciones de atributos y las relaciones entre los atributos y las posiciones geográficas. Donde las variables se combinan a menudo en las ecuaciones lineales que simplemente proporcionan la combinación aditiva de atributos y relaciones [5].
5. Integran muchas fuentes de información, tales como mapas digitales, fotografías aéreas, levantamientos topográficos e imágenes satelitales, entre otros.
6. Permiten almacenar información descriptiva de los elementos geográficos a través de administrar e integrar la información geográfica y alfanumérica.

1.1.1.3 Software para SIG

En la actualidad existen programas SIG, que ayudan a realizar un análisis exhaustivo del territorio para la toma de decisiones en las que la información espacial tiene una especial relevancia. Algunos ejemplos de este tipo de software son ArcGIS^{®2}, GRASS GIS^{®3} (Geographic Resources Analysis Support System), PostGIS^{®4} e IDRISI^{®5}, entre los más destacados. Es importante profundizar sobre ArcGIS[®] ya que es la plataforma líder mundial para crear y utilizar SIG. ArcGIS[®] es un sistema que permite recopilar, organizar, administrar, analizar, compartir y distribuir información geográfica. Además de conectar mapas, aplicaciones y datos, de manera que la toma de decisiones sea inteligente y rápida [6]. Una

² <https://www.arcgis.com>

³ <https://grass.osgeo.org/>

⁴ <http://postgis.net/>

⁵ <https://clarklabs.org/>

alternativa viable para ArcGIS® son GRASS GIS® y PostGIS® ya que tienen características similares al primero.

1.1.2. Sistemas de Recomendación

Los sistemas de recomendación son herramientas que tienen como objetivo apoyar a los usuarios en la toma de decisiones mientras interactúan con grandes espacios de información ofreciendo sugerencias personalizadas sobre determinado servicio o producto [7]. Los sistemas de recomendación estudian las características de cada usuario y mediante un procesamiento de los datos, encuentran un subconjunto de productos o servicios que satisfacen al usuario. La figura 1.3 ilustra un sistema de recomendación en términos generales.



Figura 1. 3 Representación de un sistema de recomendación⁶

1.1.2.1. Clasificación de los sistemas de recomendación

Los sistemas de recomendación se clasifican en híbridos, basados en contenido, filtrado colaborativo, filtrado demográfico y basados en conocimiento. La figura 1.4 ejemplifica esta clasificación.

⁶ goo.gl/IFUHz8



Figura 1. 4 Clasificación de los sistemas de recomendación

1. **Basados en contenido:** las recomendaciones provienen de evaluaciones subjetivas de algunos elementos, realizadas anteriormente; y de evaluaciones futuras de otros elementos similares [8]. En la figura 1.5 se representa el funcionamiento de las recomendaciones basadas en contenido.

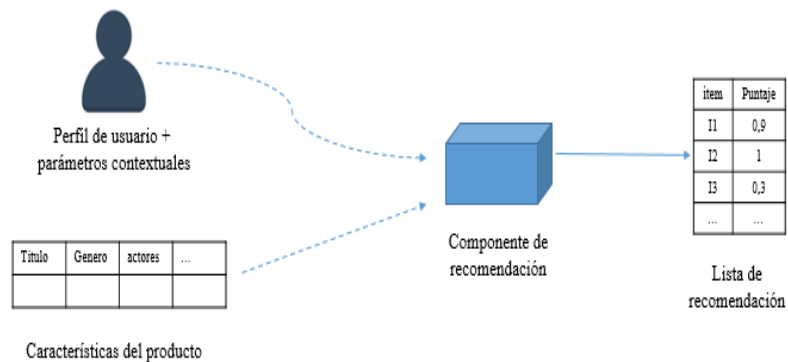


Figura 1. 5 Basado en contenido⁷

⁷ goo.gl/xF3OJr

2. **Filtrado colaborativo:** consiste en realizar las predicciones de los usuarios activos en artículos basados en las preferencias de usuarios y artículos similares [8]. La figura 1.6 ejemplifica la técnica de filtrado colaborativo.

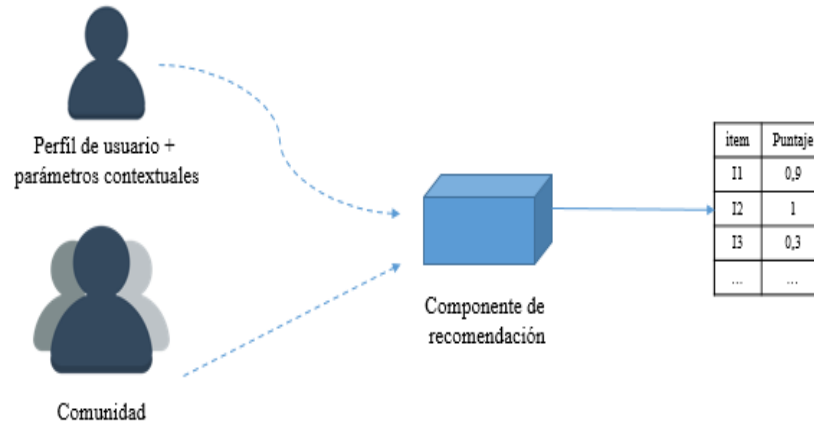


Figura 1. 6 Filtrado Colaborativo⁷

3. **Filtrado demográfico:** realiza predicciones justificadas en el principio de que individuos con ciertos atributos personales comunes (sexo, edad, país, entre otros) también tienen preferencias comunes [9], tal como se representa en la figura 1.7.

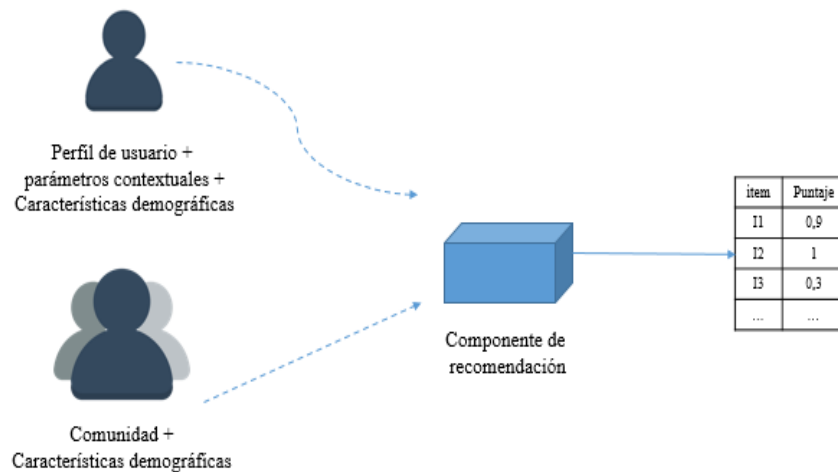


Figura 1. 7 Filtrado Demográfico

4. **Basados en conocimiento:** modelan el perfil de usuario a través de algoritmos de inferencia e identifican la correlación entre sus preferencias y los productos o servicios existentes [10] tal como se muestra en la figura 1.8.

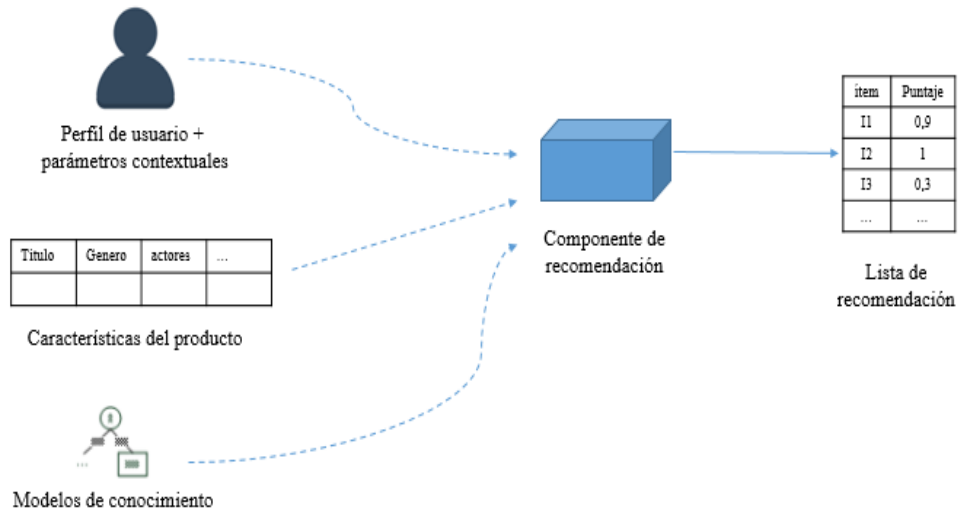


Figura 1. 8 Basados en conocimiento⁷

5. **Híbridos:** es la combinación de cualquiera de los métodos descritos anteriormente [10], esta técnica es mostrada en la figura 1.9.

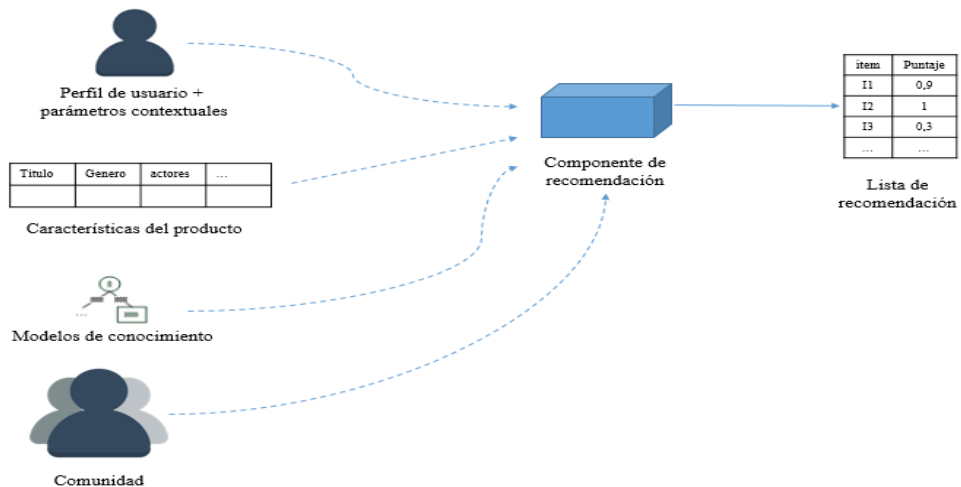


Figura 1. 9 Recomendación Híbrida⁷

1.1.2.2. Características generales de los sistemas de recomendación

Existen cinco puntos importantes dentro de los sistemas de recomendación, los cuales son independientes de la clasificación de los sistemas de recomendación. Estos puntos son adquisición de conocimiento (implícito o explícito), información compartida, representación del perfil, fuente de conocimiento y técnica de recomendación, los cuales se encuentran expresados en la figura 1.4. La adquisición de conocimiento es el proceso responsable de reunir y ampliar la información perteneciente a un perfil de usuario, este conocimiento tiene dos opciones, ser explícito o implícito [11]. Cuando el proceso es explícito se dice que es procesado, almacenado e incluso compartido por medio de computadoras y cuando el proceso es implícito es debido a la adquisición de factores intangibles, perspectivas y experiencias [12]. En cambio la información compartida es evidente; para ello es necesario que los sistemas de recomendación mejoren el alcance del conocimiento general por medio de la retroalimentación del usuario y dominio del conocimiento [13]. Por otro lado, la representación de perfiles es la característica más importante dentro de los sistemas de recomendación; ya que es necesario procesar los intereses de los usuarios por medio de técnicas de aprendizaje y adaptación de perfiles [14]. Las fuentes de conocimiento son las formas en que el conocimiento es extraído ya sea por medio del mismo usuario, por otros usuarios, datos sobre la recomendación y el dominio de la recomendación [15]. Finalmente las técnicas de recomendación están implícitamente relacionadas con la fuente de conocimiento, es decir con el tipo de información que se va a utilizar para ofrecer la recomendación; en este caso siempre es bueno visualizar que tipo de elementos se recomendarán [16]. Existen ciertas formas de obtener las características de los sistemas de recomendación, las cuales se muestran en la figura 1.10 y se describen en la tabla 1.1.

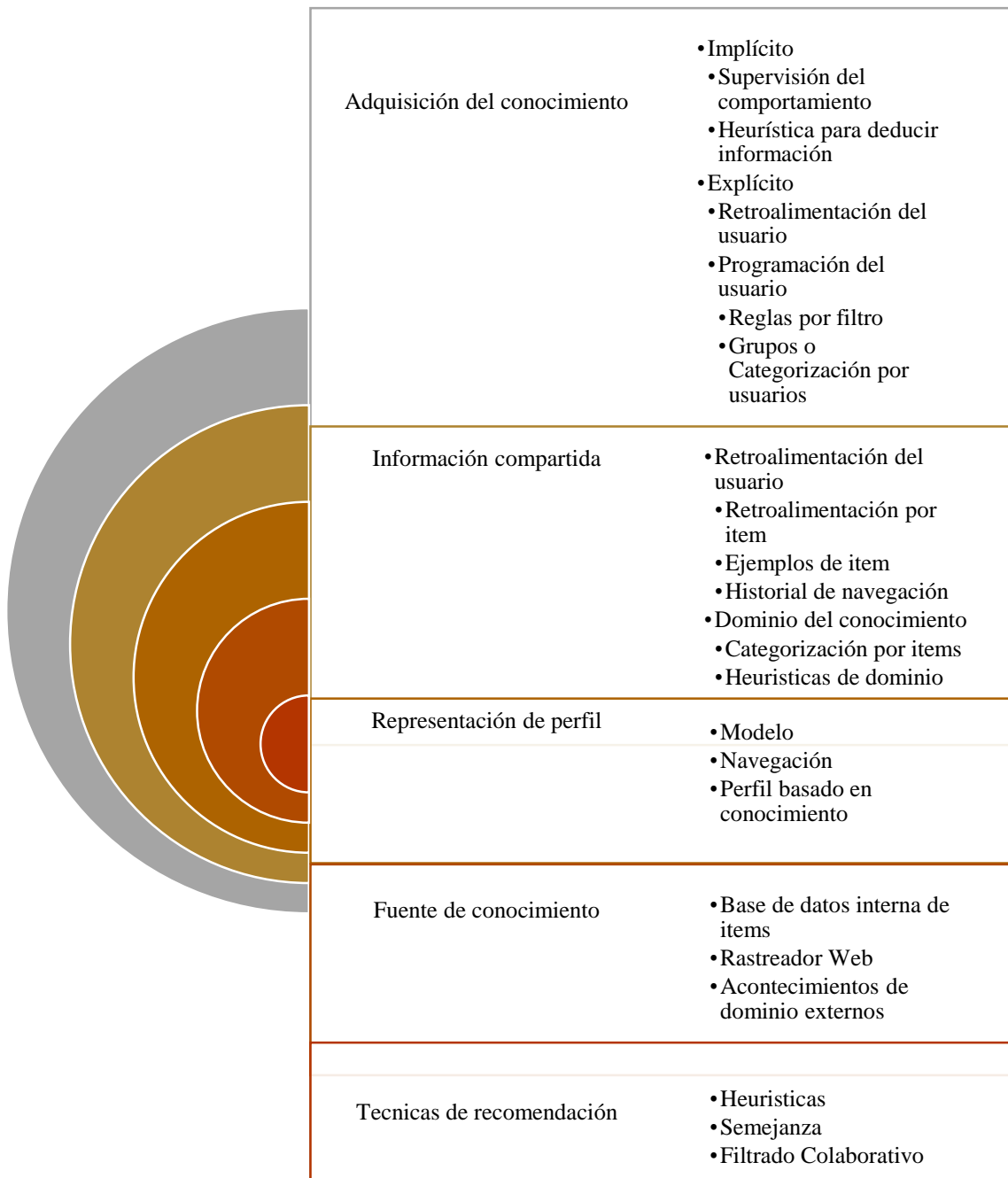


Figura 1. 10 Características de los sistemas de recomendación

Tabla 1. 1 Definición de características

Característica	Descripción
Supervisión del comportamiento	El sistema observa el comportamiento de los usuarios y lo registra.
Heurística para deducir información	Reglas usadas para la deducción de la información de los usuarios.
Retroalimentación del usuario	Es la calificación que dan los usuarios para ayudar a otros usuarios en la selección de artículos.
Reglas por filtro	Los usuarios proporcionan reglas de filtro.
Categorización o Grupos de Usuario	El sistema crea grupos o categorías de usuarios.
Retroalimentación por ítem	Es la retroalimentación de artículos que el sistema da para ayudar a los usuarios a decidir
Ejemplos de ítem	El sistema reúne artículos para crear un conjunto de entrenamiento
Historial de navegación	El sistema registra el historial de navegación para ayudar a los usuarios.
Categorización por ítems	El sistema crea grupos o categorías de artículos.
Heurísticas de dominio	Reglas para filtrar el dominio.
Modelo	El sistema usa estructuras vectoriales (modelos) para almacenar el perfil de los usuarios.
Navegación	El sistema guarda la historia de navegación.
Perfil basado en conocimiento	Los sistemas crean perfiles basados en conocimiento.
Base de datos interna de ítems	Las recomendaciones se hacen desde una base de datos propia.
Rastreador Web	El sistema rastrea la web para recomendar artículos.
Acontecimientos de dominio externo	Acontecimientos que ocurren al ejecutar los triggers de recomendación.

Característica	Descripción
Heurísticas	Reglas para encontrar el mejor artículo.
Semejanza	Función basada en contenido para encontrar artículos que pertenecen a un perfil.
Filtrado Colaborativo	Son funciones estadísticas para encontrar gente con perfiles similares.

1.1.2.3. Características específicas de los sistemas de recomendación

Los sistemas de recomendación tienen sus características propias de acuerdo al tipo que pertenezcan.

En el caso de los basados en contenido las características son las siguientes [17]:

- Lista de características del perfil de usuario y de los elementos. Aprendizaje de las preferencias del usuario por medio del perfil del usuario.
- Evaluaciones subjetivas para localizar los elementos que son similares a las preferencias del usuario.

Para los de filtrado colaborativo, contienen las siguientes particularidades:

- Uso del enfoque sabiduría comunitaria, es decir, se guía de las puntuaciones más altas que los usuarios han otorgado a un elemento.
- Supone que los usuarios con gustos similares en el pasado tendrán los mismos gustos en el futuro.
- Contiene una matriz usuarios-elementos.
- Realiza predicciones numéricas sobre el grado de interés de cada elemento.

Los sistemas de recomendación de filtrado demográfico se caracterizan con los siguientes puntos [18]:

- Uso de los estereotipos demográficos (raza, sexo, edad, ubicación).
- Asocia la información demográfica de los usuarios con las características demográficas de usuarios similares.

- Uso de vectores demográficos para realizar el cálculo de similitud.

Los basados en conocimiento se diferencian de los demás, ya que son capaces de [17]:

- Realizar recomendaciones deterministas, ya que relacionan las características de los usuarios y elementos con la construcción de modelos para ofrecer una sugerencia.
- Encontrar una serie de elementos a sugerir por medio del conocimiento; por lo tanto, no existe el problema de arranque en frío.

Finalmente en los híbridos es común [17]:

- Combinar dos o más de las clasificaciones anteriores de los sistemas de recomendación.
- Realizar mejores sugerencias, ya que al combinar algunas clasificaciones los resultados son más completos.

1.1.2.4. Ejemplos de caso de éxito de sistemas de recomendación

En la actualidad, existen una gran cantidad de sitios especializados en Internet que ofrecen servicios o productos utilizando mecanismos de recomendación. Cada día son más los servicios que incorporan funcionalidades de recomendación a sus usuarios, su veloz auge radica en el hecho de proporcionar asesoramiento a sus clientes de manera de que se conozcan y estudien los intereses y necesidades de los clientes. Este descubrimiento contribuyó al estudio de los sistemas de recomendación que van desde ámbitos de recreación a ventas como los casos de Amazon[®], Netflix[®], Reddit[®], Pandora[®], Last.fm[®], iDreamBook[®], StumbleUpon[®], entre otros, por mencionar sólo algunos.

Amazon, Inc. es una empresa estadounidense de comercio electrónico, dedicada a la venta de productos por Internet, en ella se usa la técnica de filtrado colaborativo para calcular las sugerencias de cada usuario en función de los usuarios similares que han comprado o mostrado interés en un producto.

Netflix, Inc[®]. es una empresa estadounidense de entretenimiento, que mediante una tarifa mensual ofrece transmisión de películas y series. Netflix[®] es un buen ejemplo de la técnica híbrida, ya que hace recomendaciones por medio de comparar las observaciones de las preferencias de los usuarios similares con las valoraciones que el usuario realizó anteriormente.

Reddit, Inc® es una plataforma web que consiste en compartir información por medio de técnicas de filtrado colaborativo, dicho mecanismo se basa en sugerir información interesante o relevante; que es valorada por la comunidad activa de Reddit®.

Pandora® es un servicio de radio personalizada ya que el usuario crea sus propias emisoras. Pandora utiliza un mecanismo de recomendación basado en contenido, en donde se presenta al usuario temas musicales y este decide si es adecuado a sus gustos.

Last.fm® es un servicio de recomendaciones músicas basadas en filtrado colaborativo que construye perfiles y estadísticas sobre gustos musicales, basándose en los datos enviados por los usuarios registrados.

iDreamBook, Inc ® es un sistema de recomendación de libros, en el que se usa la técnica de filtrado colaborativo al recomendar obras literarias con base a las opiniones de los críticos y las valoraciones de los usuarios.

StumbleUpon® es un sitio web comercial que integra una red social que utiliza la técnica basada en conocimiento de los sistemas de recomendación para que los usuarios intercambien páginas de interés.

1.1.3. Sistemas de Geo-recomendación

Los sistemas de geo-recomendación son una nueva categoría de los sistemas de recomendación. Representan el nuevo cambio dentro de los sistemas de recomendación, ya que la inclusión de la ubicación geográfica permite un mayor alcance en el procesamiento de los datos. Los sistemas de geo-recomendación son capaces de ofrecer recomendaciones tomando en cuenta las ubicaciones geográficas del usuario y de los negocios [19]. La principal atracción que ofrecen los sistemas de geo-recomendación reside en combinar las características de los sistemas de recomendación con los sistemas de información geográfica, donde el ámbito principal es el de actividades de ocio [19] [20]. La palabra clave dentro de estos sistemas es la “movilidad” y esto es por el uso de teléfonos inteligentes con sistema de posicionamiento global (GPS) dentro de los sistemas de recomendación; ya que con la información demográfica de los usuarios y junto con los intereses de estos es posible construir recomendaciones más acorde a los usuarios.

1.1.3.1.1. Herramientas de Geo-recomendación

Herramientas exclusivas para la construcción de sistemas de geo-recomendación, no existen como tal para ello es necesario realizar una combinación de algunas herramientas para la construcción de sistemas de recomendación, APIs de Mapeo y el uso de alguna API del gobierno para la extracción de datos geográficos del territorio.

1.1.3.1.2. Herramientas para el desarrollo de sistemas de recomendación.

Las herramientas para la construcción de sistemas de recomendación proveen un conjunto de funciones, documentación e instrumentos para el fácil desarrollo de motores de recomendación. Por medio de estas herramientas el análisis del comportamiento de los usuarios y la relación entre ellos y la selección de ítems. La selección de la herramienta de recomendación depende mucho del tipo de sistema a construir ya que integran diferentes estrategias de recomendación, el acceso de datos y la construcción del modelo de recomendación. Así como también el tipo de licencia (de paga o públicamente disponible). Generalmente las herramientas de paga son las pertenecientes al Software como Servicio (SaaS) ya que se proporcionan por un tercero siendo de fácil integración y la seguridad en la mejora de predicciones. Algunos de ejemplos de estas herramientas son Amazon Machine Learning la cual es una plataforma de aprendizaje automático para la construcción de modelos de datos y predicciones; por otro lado Azure ML es una herramienta de computo en la nube desarrollada por Microsoft para el análisis predictivo de un gran conjunto de datos y finalmente Google™ Cloud Prediction Api es servicio REST de aprendizaje automático para la predicción basado en el análisis de sentimientos [21]. Por otro lado, se tienen las herramientas de licencia pública que a pesar de no ser de fácil integración comparadas con las anteriores proveen un gran apoyo por parte de la comunidad de desarrolladores. Algunos ejemplos de este tipo de herramientas son MyMediaLite desarrollada en C+, la cual trabaja bajo la predicción basada en filtrado colaborativo [22]; Crab el cual es un motor de recomendación desarrollado en Python; el cual es útil para los sistemas de recomendación de filtrado colaborativo o híbridos; jCOLIBRI desarrollado en Java, dirigido para sistemas basados en contenido [23]. Finalmente Apache Mahout desarrollado en Java™, el cual está dirigido a los sistemas de recomendación basados en filtrado colaborativo con algoritmos algebraicos basados en minería de datos [24]

1.1.3.1.2.1. Apache Mahout

Es una API de programación que permite resolver problemas de agrupamiento, clasificación y filtrado colaborativo construida sobre el paradigma MapReduce. Apache Mahout tiene como objetivos trabajar con grandes volúmenes de datos, sistemas distribuidos y contar con el apoyo comunitario para ofrecer mejores funciones y soporte en errores que van surgiendo [24]. La siguiente figura ilustra la arquitectura de Apache Mahout.

Características de Apache Mahout

Apache Mahout cuenta con las siguientes características:

- Prueba de filtrado colaborativo que proporciona algoritmos de aprendizaje automático para ampliar las aplicaciones por medio de las recomendaciones personales.
- Implementaciones MapReduce en clustering y Naive Bayes [25].

Algoritmos que soporta

Actualmente Apache Mahout da soporte a los siguientes algoritmos:

- Regresión lineal resuelta por medio del gradiente estocástico descendente. Este algoritmo describe un tipo de modelo de clasificación simple y secuencial, que es capaz trabajar con grandes volúmenes de datos.
- Modelos ocultos de Markov; que son implementaciones secuenciales y paralelas de algoritmos de clasificación en etiquetado textual.
- Descomposición de valor singular es una técnica matemática generalmente usada para reducir el ruido de matrices.

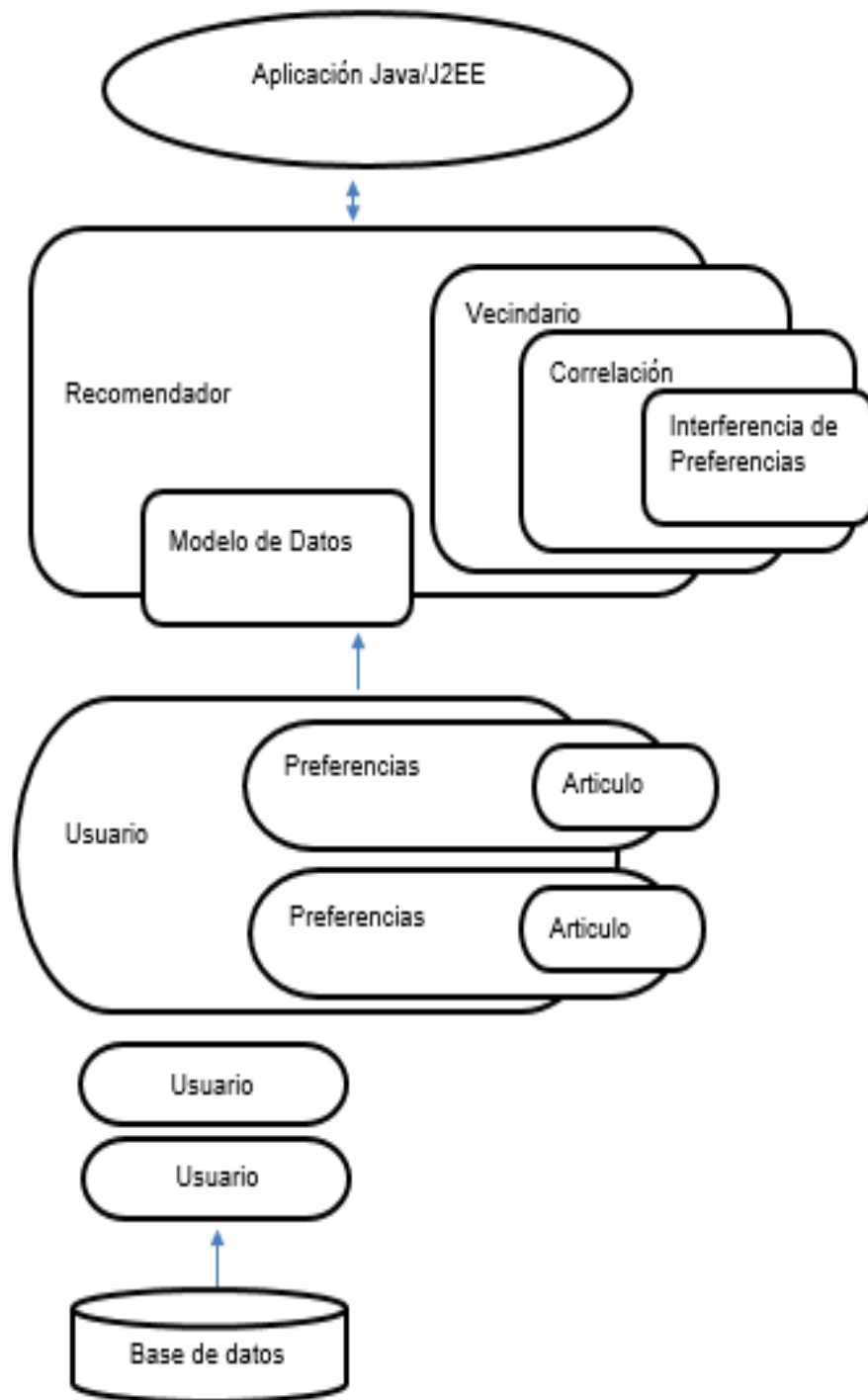


Figura 1. 11 Arquitectura Apache Mahout tomada del sitio de ibm⁸

⁸ goo.gl/F49Yj3

- Almacenamiento en clúster Dirichlet, es útil para determinar la propiedad en datos con superposición.
- Almacenamiento en clúster espectral; en donde usa un enfoque gráfico para determinar la partición del *clúster*.
- Almacenamiento en clúster Minhash a través de una estrategia hash para agrupar elementos similares en clústeres [26].

1.1.3.1.3. APIs de Geo-localización

Son las entidades que proveen servicios de Geo-localización, las cuales ofrecen el uso de mapas, la entrega de coordenadas de una determinada posición o ubicación, así como la exactitud de un radio basado en una dirección IP detectada generalmente por dispositivos móviles. Este tipo de servicios acceden a muchas bases de datos para obtener la información necesaria de un punto, así como la integración de características de Sistemas de Información Geográfica (SIG). Algunos ejemplos de proveedores de servicios son OpenStreetMap™ API, Bing™ Maps Geocode, and Google™ Maps. Los cuales son servicios RESTful que devuelven los datos en formato XML o JSON.

1.1.3.1.3.1. Google™ Maps

Es una API de Geo-localización la cual se integra por una colección de clases en JavaScript, dicha herramienta permite la creación de *mashups* para dibujar o compartir mapas por medio de datos geográficos [27]. Google™ Maps combina el uso de imágenes satelitales con mapas convencionales [28]. El servicio de Google™ Maps usa la Geo-localización para especificar la posición geográfica sobre un mapa. Los datos usados por Google™ Maps usan la posición geográfica definida por una latitud [-90 (S), 90 grados (N)] y la longitud [-180 (O), 180 grados (E)] el sistema de coordenada [29]. Finalmente, el API también proporciona una unidad eficiente para construir usos que usan datos espaciales geográficos y proporciona rasgos de GIS.

1.1.3.1.4. INEGI

El Instituto Nacional de Estadística y Geografía (INEGI) es un organismo público con autonomía técnica y de gestión, personalidad jurídica y patrimonio propios, responsable de normar y coordinar el Sistema Nacional de Información Estadística y Geográfica, así como de realizar:

- Realizar los censos nacionales.
- Integrar el sistema de cuentas nacionales.
- Elaborar los índices nacionales de Precios al Consumidor, e Índice Nacional de Precios Productor.

El INEGI representa la fuente confiable para la extracción de la geografía del territorio mexicano. Una herramienta importante que ofrece esta institución es el Servicio INEGI DENUÉ [30].

1.1.3.1.4.1. INEGI DENUÉ

Es una herramienta proporcionada por el INEGI, la cual es fuente de información del Directorio Nacional de Unidades Económicas de la república mexicana. El Diario Oficial de la Federación del 16 de abril del 2008 publica la Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG). En su artículo 23 establece que “contará con un Directorio Nacional de Unidades Económicas” (DENUÉ) y en los artículos 94 y 95 determina “la responsabilidad del INEGI para establecer, normar y operar el Directorio”. En su página de internet enuncia: “El Directorio Estadístico Nacional de Unidades Económicas (DENUÉ, 03/2011) ofrece información sobre la identificación y ubicación de todos los establecimientos activos en el territorio nacional, de acuerdo a los datos recabados durante la Actualización de Unidades Económicas, 2010; por ello, constituye una herramienta fundamental para la toma de decisiones en los ámbitos público y privado [31]”.

1.2. Planteamiento del problema

Los sistemas de recomendación ayudan a filtrar los elementos de información recuperados, usando distintas técnicas, para identificar aquellos elementos que mejor satisfacen las preferencias o necesidades de los usuarios. Las recomendaciones se generan a partir de las

opiniones que proporcionan otros usuarios sobre esos elementos en búsquedas previas, o bien a partir del perfil del usuario.

En la sociedad actual es cada vez más importante la recomendación y los usuarios utilizan las recomendaciones para contrastar si los productos, servicios o contenidos que, quieren consumir son de su agrado. Es por eso que en los últimos años muchas compañías y sitios Web implementan sistemas para estudiar los gustos de los usuarios y adaptar sus productos, información y servicios más adecuadamente a dichos intereses con la finalidad de mejorar sus resultados.

Por otra parte, las empresas comerciales tienen la necesidad de preocuparse por el comportamiento de sus ingresos y de sus pérdidas, los cuales en muchas ocasiones son directamente afectados por la localización de sus puntos de venta. Estos elementos por lo general dependen de la cercanía de las instalaciones entre los clientes potenciales; una mala ubicación conlleva egresos en vez de ganancias.

Es por eso que en los últimos años algunas empresas realizan estudios de mercado para evaluar a la competencia y su posicionamiento; intentando captar el negocio de sus contrarios investigando al consumidor y al entorno que le rodea. Dichos estudios suelen apoyarse de técnicas de minería de datos, sistemas geográficos de información y sistemas de recomendación con la finalidad de mejorar la productividad en las ventas.

Expresado lo anterior, se propone desarrollar un sistema de geo-recomendación de localización de los puntos de venta con estrategias de mercado implementado técnicas de logística y minería de datos, así como algoritmos de recomendación. Para ubicar los establecimientos comerciales en función de su potencial comercial que satisfaga al público perteneciente a la zona seleccionada del punto de distribución.

1.2. Objetivo general y específicos

1.2.1. Objetivo general

Desarrollar un sistema de recomendación para la ubicación de establecimientos que permita identificar las áreas de la zona de localización óptima.

1.2.2. Objetivos específicos

- Analizar el estado del arte sobre sistemas de recomendación basados en filtrado colaborativo, sistemas de información geográfica y dinámica de sistemas.
- Identificar y explorar herramientas que soporten el desarrollo de sistemas de recomendación que permitan la integración con modelos geográficos.
- Identificar los diversos factores (ubicación de los clientes, nivel socio-económico de la zona, grado de afluencia, ubicación de servicios complementarios y nivel de competitividad) que se relacionan y correlacionan con la colocación de los puntos de venta.
- Definir una arquitectura de integración para el desarrollo del sistema de recomendación que integrará los modelos geográficos.
- Definir los niveles de prestación de servicios, módulos y componentes de la arquitectura de integración.
- Desarrollar los módulos y componentes para la ubicación de establecimientos de puntos de venta usando modelos geográficos.
- Validar la arquitectura de integración mediante un caso de estudio real en una empresa de la región como prueba de concepto.

1.3. Justificación

Las empresas comerciales se enfrentan constantemente al reto de proporcionar a sus usuarios el acceso adecuado a sus servicios de una forma ágil y cómoda a través de la ubicación del negocio, es decir la localización es un factor determinante para el nivel de ingresos del negocio. Visto desde otra perspectiva los usuarios se enfrentan a una vida atareada y variable, de manera que necesitan que los servicios se adapten a ellos. Por lo tanto, para tomar la decisión correcta con

respecto a la ubicación de la empresa es necesario conocer los comportamientos de competencia, demanda, afluencia y las tendencias.

Los factores identificados muestran que un simple estudio basado en entrevistas o experiencias muestra incertidumbre, por lo que es necesario hacer uso de información objetiva por medio de la construcción de modelos geográficos utilizando herramientas de Geo-localización, así como el uso de técnicas de sistemas de recomendación para tomar en cuenta la necesidad de la futura comunidad de clientes potenciales que pertenecen a la zona de ubicación del negocio.

Dichos motivos proponen el desarrollo de un Sistema de recomendación para la ubicación óptima de establecimientos comerciales en la zona, que ofrece ventajas frente al comportamiento de los clientes y la competencia; ya que estos estudios no son comunes.

Analizado lo anterior, se identifica que el presente estudio tenga impacto económico para los comercios, dé confianza a los dueños de los negocios al seleccionar la ubicación estratégica y en estudios de integraciones geográficas y de sistemas de recomendación de futuros trabajos.

Capítulo 2. Estado de la práctica.

En este capítulo se da a conocer el estado de la práctica del proyecto de innovación, es decir, algunos trabajos relacionados con el proyecto de innovación.

2.1. Estado de la práctica

A continuación se presenta una recopilación de trabajos relacionados con esta propuesta; que describen el área de oportunidad que se desea abordar en el proyecto de innovación.

2.1.1. Sistemas de recomendación aplicados en diferentes dominios

En [32] desarrollaron un sistema de recomendación para dispositivos móviles basado en agentes; donde mostraron recomendaciones sobre actividades cercanas e interesantes para el usuario. En [20], se discutió que los cambios en el turismo electrónico requieren que los servicios proporcionen a los usuarios información relevante de acuerdo a sus contextos físicos actuales, teniendo en cuenta los gustos y preferencias; sin problemas en las redes por el ancho de banda. Con base en lo anterior, se propuso la implementación móvil sensible al contexto 3D de REJA que era un sistema Web de recomendación de restaurantes de la provincia de Jaén, España. Para ello se desarrolló un módulo de pre-filtrado sensible a la ubicación del usuario y se proporcionó un protocolo de petición-respuesta ligero construido directamente sobre TCP/IP con el fin de comunicar de manera eficiente los clientes móviles con el servidor del SIG. Finalmente se obtuvo una aplicación intuitiva (capaz de informarles el lugar en donde se ubican, los restaurantes que se encuentran cerca y cómo llegar a ellos) basada en mapas 3D que supera a la mayoría de aplicaciones para dispositivos móviles. En [33], se propuso un sistema de recomendación basado en los cupones de descuento con el fin de promover los productos pertenecientes a las plataformas en línea, ya que recientemente ha surgido un nuevo tipo de comercio grupal. Por medio de un mecanismo de recomendación de cupones para implementar el comercio social basado en la localización e influencia por la semejanza de los usuarios sobre la compra; consistiendo en los criterios de similitud entre las preferencias del usuario y las características del producto, la conveniencia geográfica de la tienda, y la influencia de las evaluaciones de los amigos sobre los productos. El caso de estudio consistió en la

recomendación de 20 productos, donde la mayoría de ellos fueron elegidos desde el sitio web en línea Groupon. Para dicho estudio se utilizaron técnicas de minería de datos (Construcción de árboles para categorizar los productos) y un tratamiento de los datos para construir la red de usuarios y recoger datos de comportamiento de éstos. Con el fin de evaluar el desempeño de la recomendación grupal de los cupones, los experimentos utilizados se evaluaron con la aplicación de cuestionarios en línea para recopilar información sobre la satisfacción. El trabajo permitió identificar con precisión a los clientes con alta disposición a la compra de los productos promocionados y la observación del mecanismo de formación de grupos cohesivos que aumentan significativamente el valor derivado de la compra colectiva. En [19], se enfatizó que el auge en el mercado de teléfonos inteligentes y los avances en las redes inalámbricas han traído el desarrollo de sistemas de recomendación sensible al contexto. La mayoría de las ramas a la que pertenecen estos sistemas son turismo y entretenimiento. En la rama del entretenimiento se observó que los sistemas de recomendación no tomaban en cuenta a los factores de hora, ubicación y multitud así como la información perteneciente a las salas en donde se transmite determinada película. De manera que se propuso el desarrollo de un sistema de recomendación móvil híbrido sensible al contexto, en donde se integra a la ontología Movie Ontology (MO) las clases y propiedades correspondientes a la duración de una película dentro de una sala de cine así, como el número de salas en donde se presenta determinada película. La aplicación descarta las películas que no se presentan en las salas de cine y desecha los horarios de las películas a las que el usuario activo es probable que no asista basado en la distancia entre la posición del usuario activo actual (origen) y la ubicación del cine (destino) a partir de la identificación de las preferencias del usuario por medio de una matriz de similitudes, el análisis del perfil de usuario y un pre y post filtrado de la información proveniente de la multitud de personas que asisten a determinada función y la ubicación del usuario a través del uso de Foursquare mediante los sensores en los dispositivos móviles. El resultado final es una herramienta capaz de ofrecer recomendaciones a los usuarios con 81 -83% de precisión. En [34] propusieron un sistema de inferencia basado en servicios de ubicación y conocimiento, el cual construye el conocimiento a través una aplicación móvil; mediante esta información el sistema ofrece recomendaciones. En [35] manifestaron la creación de un método de recomendaciones turísticas basadas en imágenes obtenidas de redes sociales basadas en servicios de localización; donde a través de la

construcción de un modelo con los atributos de las imágenes y junto con la información geográfica concluyen que las mejores recomendaciones turísticas se obtienen al usar la indexación semántica latente, clasificación y la ubicación de imágenes. En [36] propusieron un sistema de recomendación de rutas turísticas a partir de información proveniente del geo-etiquetado de una red social basada en servicio de localización donde es tomado en cuenta el lugar visitado, la época en la que fue visitado y el tiempo empleado y por medio del cálculo de semejanza entre imágenes ofrecen un plan de viaje estructurado. En [37] manifestaron la necesidad de un sistema planificador de viaje por medio el análisis de los atributos (popularidad, tiempo de visita, visitas por día y apariencia a través del tiempo) de fotos provenientes de TripAdvisor y Yahoo!Travel. La recomendación fue construida por medio de un algoritmo capaz de insertar los lugares y tiempos destinados a visitar tomando en cuenta los intereses de los usuarios y la relación entre los destinos. En [38] expusieron que la atracción a los puntos de interés se encuentra relacionada a las características de un grupo de personas que lo visita. De manera que fue realizado un análisis de las características del grupo (cercanía y tamaño), las cuales son obtenidas a través de la relación de datos geo-sociales de una red. Finalmente fue demostrado que las mejores recomendaciones de puntos de interés son aquellas que toman en cuenta el comportamiento del grupo. Por otra parte en [39] declararon que es necesario contar con herramientas que ayuden en la toma de decisión de rutas a tomar por los peatones en la ciudad de México por lo que propusieron un prototipo de sistema de recomendación móvil basado en el filtrado colaborativo. De tal forma en [40] desarrollaron algoritmos híbridos de recomendaciones musicales basados en la integración geo-espacial de las semejanzas entre los usuarios y el contenido de información musical y realizaron experimentos que indican que las estrategias híbridas de recomendación son capaces de superar a las recomendaciones que usan sólo un tipo de datos. En [41] expusieron el problema de movilidad humana al ofrecer recomendaciones de puntos de interés, por lo que propusieron un modelo basado en gráficos y la integración de la distribución geográfica así como el comportamiento del usuario para recomendar puntos de interés apropiados a los objetivos de los usuarios. Concluyen que combinar la información geográfica y social de los usuarios, ofrecen mejores resultados.

2.1.2. Sistemas de información geográfica aplicados en estudios ambientales y urbanos

En [42], se discutió la falta de herramientas tecnológicas en Colombia, basadas en dinámicas de propagación de frentes de epidemias que permitan abordar, visualizar y contrarrestar su propagación. Se decidió tomar el modelo genérico de epidemiología matemática propuesto por Navarro Juárez para entender y analizar por medio de la dinámica de sistemas las fluctuaciones en las poblaciones de sanos, recuperados e infectados por polígono en una región de influencia para representar al modelo susceptible-infectado-removido (SIR), relacionado con las enfermedades que confieren inmunidad permanente. Con el fin de modelar factores críticos en la propagación de epidemias, se tomaron los datos de una epidemia de asma en el distrito de Manhattan de Nueva York, pertenecientes a ArcGIS®. El modelo SIR se representó por medios de diagramas de Forrester utilizando herramientas de dinámica de sistemas, se implementó y se ejecutó dicho modelo usando herramientas SIG con el propósito de visualizar e interpretar las fluctuaciones de sanos, infectados y recuperados. La adecuación del modelo matemático SIR en el SIG permitió hacer análisis cuantitativos sobre la distribución espacial y la variación poblacional. Además de que se utilizó el método del Krigin para calcular el mapa de predicción. El modelo computacional propuesto usando SIG permitió observar la fluctuación espacial de las poblaciones por polígonos en una región de influencia; correspondiente a la primera fase de desarrollo del proyecto de investigación modelación y simulación computacional. Por otra parte en [43] se manifestó que los efectos de la deposición de los residuos de las granjas de peces ubicadas en la costa oeste de Escocia, contienen bacterias (*etry bathym*) que afectan al fondo del mar, así como también la salud de las personas. Debido a ello se planteó un estudio, en donde utilizó una combinación de hojas de cálculo y un SIG por medio de un módulo de dispersión desarrollado en DELPHI3 e integración con el software IDRISI32 usando IDRISI Application Programming Interface para estimar la distribución de los residuos de carbono en partículas derivadas de la actividad marina; describiendo un modelo de distribución de residuos que no afecten el ecosistema marino. El lugar utilizado para la recolección de datos de campo consistió en 70 metros de circunferencia de la costa, 26 metros de profundidad, 3 metros por debajo de la superficie en la marea más baja prevista y 3 metros por encima del fondo del mar. Los datos se guardaron en un documento en excel y estos se importaron al modelo de dispersión previamente construido. Posteriormente produjo un conjunto de información gráfica y estadística; que

reportó una reducción en 32% la materia fecal de los peces al encontrar el sitio óptimo de distribución de residuos. En [44] se planteó la falta de una herramienta capaz de predecir los riesgos por la intrusión de agua proveniente de alcantarillas, drenajes y zanjas a los sistemas de distribución de agua. Debido a ello, fue necesario realizar una investigación y construir la herramienta IRA-WDS para evaluar el riesgo de los contaminantes en el sistema de distribución de agua; mediante la elaboración de los modelos de entrada de contaminantes, el cual estima las zonas fuentes de éstos; el modelo de condiciones en la tubería en donde se estima la condición de las tuberías de la red de distribución de agua y evalúa la potencial vía de intrusión y finalmente el modelo de evaluación de riesgos que genera el índice de riesgo para cada tubo mediante el uso que genera las ponderaciones para cada uno de los factores de riesgo y normaliza un factor de riesgo en particular. Los módulos fueron construidos en C y éstos se integraron a un SIG. El área estudiada fue una población de Guntur, India; dicha área tiene una población de alrededor de 60, 000 y un área de 4 km². Además de que su sistema de suministro de agua es propenso a la contaminación, por lo tanto, las autoridades encargadas de la gestión de este sistema se dispusieron a entender los riesgos en el suministro de agua. El resultado del uso de esta herramienta permitió a los ingenieros priorizar el mantenimiento y se obtuvo que el sistema de tuberías presenta un nivel medio-bajo de riesgo por lo que se recomendó rehabilitar tuberías que se consideren en mal estado, llevar a cabo una detección de fugas y el programa de reparación en las zonas con muchas conexiones. En [45] se discutió que la producción de ostión japonés en Funka Bay está aumentando, pero el desarrollo de ésta se encuentra obstaculizado por los limitados sitios adecuados e inquietudes con respecto a los impactos sobre el medio ambiente, los cuales provocan la sobre explotación en el uso de los recursos naturales. Se presentó una evaluación multicriterio basada en un SIG que utiliza datos de detección satelital y datos de verificación de campo para identificar los sitios más adecuados para el desarrollo de la producción de vieira japonesa en Funka Bay. Por medio de una recolección de 287 imágenes con una buena resolución pertenecientes al periodo comprendido entre junio de 2002 a agosto de 2004, se combinaron para construir un mapa compuesto, que se utilizó para generar valores promedio de la temperatura del mar; a través de las herramientas SeaDAS® para el procesamiento de las imágenes y ArcGIS® para la construcción del modelo. Posteriormente se escaneó un mapa hidrográfico, que fue importado al SIG, utilizando la técnica de red triangular

irregular, para reclasificar las puntuaciones ideales para la producción de vieiras. El resultado obtenido demostró que sitios con profundidad de agua entre 15 y 30 metros son ideales para esta actividad. A medida que se disponga de más datos (imágenes de satélite o mediciones de campo), la utilidad de esta herramienta aumentará y proporcionará una gama de funciones incorporadas en diversos componentes que se adaptan para la selección del sitio óptimo. En [46] se enfatizó que la construcción y el mejoramiento de las áreas residenciales urbanas es un tema crítico en el proceso de todo el desarrollo urbano, el cual está conectado con una serie de desafíos y problemas. En virtud de lo señalado se propuso un estudio con enfoque integrado para la evaluación de la sostenibilidad del desarrollo urbano residencial, teniendo en cuenta los indicadores de sostenibilidad, de equilibrio de vivienda y construcción. Se presentó un caso de estudio para la Región de Stuttgart del estado de Baden-Württemberg, Alemania; para analizar el crecimiento poblacional; por medio de la construcción de un sistema GISSD que consistió en el análisis del SIG y de los sistemas dinámicos (modelo y la visualización 3D) y la visualización espacial en 2D a través de ArcGIS® y en 3D por medio de CityEngine, de las zonas residenciales para evaluar el desarrollo urbano. El resultado de este estudio contribuyó con la aplicación de un GISSD para la construcción de modelos de sostenibilidad urbano residencial de desarrollo y con la integración del SIG para generar simulación espacial y vistas en 3D. A través del modelo se conoció el potencial económico urbano, la migración de la población y el medio ambiente. Por otra parte, en [47] se mencionó que la competencia y el desempeño de las franquicias son afectados en parte por los factores de selección en la ubicación y la calidad de las instalaciones, así como la distancia máxima que los clientes están dispuestos a recorrer en busca de un servicio o producto. Se propuso una investigación en donde se emplearon modelos de localización competitiva para representar las interrelaciones entre los tomadores de decisiones, y herramientas de SIG para proporcionar un mapa que representa la cuota de mercado donde la empresa busca maximizar sus ganancias o ingresos por medio de modelos de localización competitiva espaciales que incorporan varios aspectos relacionados con los consumidores y los proveedores de bienes o servicios. Para este estudio se desarrollaron dos herramientas utilizando el ArcGIS® Model Builder; dichas herramientas permiten estimar las zonas óptimas para establecer los puntos de venta por medio de la ponderación geográfica y los factores de costo y competitividad. El desarrollo de estas herramientas facilitó resolver el problema de la elección

de la ubicación para una nueva franquicia, teniendo en cuenta los factores de geografía, demanda y competencia. De manera similar, en [48] se discutió que la apertura de un nuevo establecimiento es un factor crítico para las empresas en el sector minorista, ya que la decisión lleva consigo una serie de riesgos muy graves de imagen corporativa y financiera; donde la geografía y los criterios asociados a ésta juegan un papel clave en el éxito de un negocio. Se desarrolló de una metodología para el proceso de selección de puntos de venta combinando a los SIG para visualizar los datos espaciales que influyen en la toma de decisión y a el proceso de jerarquía analítica (AHP), el cual consiste en definir un modelo a través de los criterios asociados a la localización y las alternativas de ubicación mediante un análisis en la geodemanda y geocompetencia. Para la investigación se analizó el sector de la distribución comercial de Murcia España. El proceso consistió en determinar y ponderar los factores principales que afectan el éxito de los negocios, consecutivamente se obtuvo los rangos posibles de ubicación con base a las ponderaciones anteriores a través de la integración del análisis en geodemanda y geocompetencia por medio del software ArcGIS®, además de la realización en los cálculos poblacionales con respecto a los datos de las unidades catastrales y el promedio de estos. Finalmente, a la información obtenida le fue aplicado AHP donde se evaluó los criterios mejor ponderados con respecto a la densidad de concentración de clientes potenciales usando el método Kernel. La investigación reveló que el éxito de un supermercado es en un 17.4% al tipo de comercio con el que cuentan, 14.62% a la publicidad web, 14.49% a la distancia entre los supermercados pertenecientes a la competencia, 9.75% al mercado potencial del supermercado, 9.71% al acceso por medio de carretera y 9.17% al acceso a pie. El estudio actuó como variable para el desarrollo de un modelo que combina a los SIG y los criterios múltiples para determinar nuevos puntos de venta. Finalmente, en [49] se expuso que tanto el proceso de urbanización como el uso de suelo cambian de manera sustancial la cobertura de la tierra al aumentar las áreas construidas con materiales impermeables; además de la disminución de zonas con vegetación, por lo que se altera el balance de energía nativa de la población, aumentando la temperatura del aire en las ciudades con respecto a sus alrededores; este fenómeno se conoce como Isla Urbana de Calor (IUC). En razón de lo expuesto se propuso realizar un estudio de la ciudad de Mexicali, Baja California, México; al ser una zona industrial con un clima muy extremo en el año. De manera que fue necesario usar las temperaturas registradas en período de 2000 a 2005,

pertenecientes a la zona urbana de Mexicali y a la zona rural de Calexico, California en Estados Unidos. Por medio de un algoritmo genético implementado con el software Stella se interpolaron las temperaturas y se compararon con las calculadas por técnicas de interpolación elaboradas con un sistema de información geográfica; posteriormente fue necesario distribuir en dos subsistemas los días representativos de invierno, primavera y otoño. En el primer subsistema el proceso principal fue interpolar temperaturas urbanas con datos proporcionados de estaciones meteorológicas ubicadas dentro de la ciudad. El segundo subsistema interactúa con el primero tomando como entrada la temperatura máxima, su objetivo principal es generar como salida datos más detallados del fenómeno de IUC. La investigación demostró que la transformación de las áreas nativas en áreas urbanas, es capaz de incrementar el calentamiento del sistema climático. Este modelo resulta ser una herramienta confiable para detectar la intensidad, hora y ubicación geográfica de la IUC y amigable para desarrolladores urbanos y personal dedicado al estudio ambiental.

La Tabla 1 presenta un análisis comparativo de la literatura, tomando en cuenta los SIG, Sistemas de Recomendación.

Tabla 2. 1. Análisis Comparativo de la literatura.

NU=No Utilizado	NE= Especificado	No H=Híbrido	FC=Filtrado Colaborativo
Artículo	Objetivo	SR	SIG
Castro et al. [42]	Plantear, desarrollar y simular un modelo epidemiológico usando SIG y dinámica de sistemas.	NU	ArcGIS®
Corner et al. [43]	Analizar, diseñar y construir el módulo de dispersión de desechos marinos basado en técnicas de SIG	NU	TerrSet (IDRISI)

Artículo	Objetivo	SR	SIG
Vairavamoorthy et al. [44]	Desarrollar un software predictor de riesgos asociados a los sistemas de distribución de agua fundamentados en SIG.	NU	ArcGIS®
Radiarta et al. [45]	Construir un modelo de evaluación multicriterio basado en técnicas SIG.	NU	ArcGIS®
Yu et al. [34]	Desarrollo de un sistema de recomendación a partir de la construcción de conocimiento colectivo.	H NE	Un
Xu y Volker [46]	Desarrollar un GISSD en 3D para la evaluación de la sostenibilidad del desarrollo urbano residencial.	NU	ArcGIS®
Suárez et al. [47]	Diseñar y construir modelos y herramientas de localización óptima para franquicias comerciales.	NU	ArcGIS®
Batet et al. [32]	Desarrollo un sistema de recomendación de películas híbrido para dispositivos móviles basado en agentes.	H NE	Un
Noguera et al. [20]	Desarrollar e integrar un motor de recomendación móvil	H NE	NE

Artículo	Objetivo	SR	SIG
	híbrido sensible a la ubicación con una arquitectura 3D-SIG.		
Roig et al.[48]	Desarrollar un método de selección de puntos de ventas fundamentado en SIG y el proceso de jerarquía analítica.	NU	ArcGIS®
Casillas et al [49]	Aplicar y validar la técnica de modelado dinámico en la estimación de intensidad y hora en que ocurre la isla urbana de calor.	NU	TerrSet (IDRISI)
Li et al. [33]	Desarrollar un mecanismo de recomendación para compras grupales con cupones de descuentos a través de análisis de preferencia y ubicación geográfica.	H NE	NU
Colombo et al [19]	Crear un sistema móvil de recomendación de funciones de películas; sensible a la ubicación, el tiempo y a los espectadores.	BC NE	NU
Pliakos et al. [35]	Crear un método de recomendaciones turísticas basadas en la indexación	H NU	NU

NU=No Utilizado	NE= Especificado	No	H=Híbrido	FC=Filtrado Colaborativo
Artículo	Objetivo		SR	SIG
	semántica latente de las imágenes.			
Yin et al. [36]	Desarrollo de un sistema de recomendación turística basada en el geo-etiquetado de la información de usuarios que previamente han visitado ese lugar.		H NE	NU
Fu et al. [37]	Desarrollo de un sistema planificador de viaje por medio de la inserción de actividades y tiempos destinados en lugares a visitar.		H NE	NU
Wei et al. [38]	Analizar las características de comportamiento que tienen los grupos de personas que visitan ciertos puntos de interés.		un	NU
López-Ornelas et al. [39]	Diseño de un prototipo de sistema recomendación peatonal.		FC NE	NU
Schedl et al. [40]	Diseño de algoritmos híbridos de recomendaciones musicales basados en la integración geo-espacial y el contenido musical.		H NE	NU
Huang et al [41]	Diseño de modelo de geo-recomendación basado en		H NE	NU

NU=No Utilizado	NE= Especificado	No	H=Híbrido	FC=Filtrado Colaborativo
Artículo	Objetivo			SR SIG
	gráficos y en la integración de la distribución geográfica y el comportamiento humano.			

Con base al análisis de los artículos relacionados se exhiben los distintos ámbitos en los que fueron abordados los sistemas de información geográfica y sistemas de recomendación; donde el más predominante es el comercial. Se observa que actualmente la búsqueda de puntos de interés está dominada por los sistemas de información geográfica, el cual carece de las características y necesidades de los usuarios; por lo que hasta el momento no existe una herramienta que proporcione de manera completa e integrada un análisis geográfico para establecer puntos de venta por medio de mecanismos de recomendación que permita ofrecer recomendaciones de puntos de venta.

2.2. Propuesta de solución

En esta sección se presenta la propuesta de solución desde el enfoque de las tecnologías de información planteadas en el capítulo primero; la cual permite alcanzar el objetivo general de la tesis.

2.2.1. Alternativas de solución.

En la tabla 2.2 se muestran las diferentes alternativas de solución al problema abordado en este proyecto.

Tabla 2. 1 Alternativas de solución

Propuesta	Metodología	API para desarrollo del sistema de recomendación	Herramienta de Geo-localización	Framework para aplicaciones web
1.	UWE	Apache Mahout	Google™ Maps	JSF
2.	WAE	Apache Mahout	OpenStreetMap™	JSF
3.	HFPM	jCOLIBRI	Google™ Maps	JSF
4.	OOHMD	MyMediaLite	OpenStreetMap™	Django
5.	UWE	Crab	OpenStreetMap™	Django

2.2.2. Solución propuesta.

Tomando en cuenta la disponibilidad de los recursos de material de apoyo que acompañan a algunas tecnologías, así como el grado de explicación de éstos, se ha determinado que la mejor alternativa es la propuesta número uno como solución propuesta para el presente proyecto de innovación.

2.2.3. Justificación de la solución propuesta.

La alternativa seleccionada como remedio al problema planteado está justificada bajo los siguientes argumentos. En lo que respecta a la selección de metodología fue más conveniente UWE, ya que se cuentan con muchos recursos educativos para ponerla en marcha como metodología para el desarrollo del proyecto, así como el hecho de ser de las pocas metodologías que cumplen completamente el ciclo de vida de las aplicaciones web; en el caso del API de desarrollo se decidió usar Apache Mahout; en la selección de la herramienta de Geo-localización se optó por Google™ Maps. Finalmente se decidió escoger JSF como marco de trabajo, ya que es compaginable con la API de Apache Mahout.

Capítulo 3. Aplicación de la metodología

En este capítulo se realiza la descripción de cada una de las capas, así como de los componentes arquitectura del sistema desarrollado en este trabajo de tesis. Complementario a la descripción de la arquitectura del sistema, se menciona el funcionamiento de la arquitectura y pormenorización del esquema de base de datos, fuentes de información y métricas de similitud.

Con la intención de proporcionar al lector la información necesaria para entender el proceso de geo-recomendación, es posible el comprender mejor el flujo de trabajo del sistema desarrollado.

3.1. Aplicación del proceso de fases de UWE para el diseño y desarrollo de un sistema de geo-recomendación.

Para dar solución al problema identificado en este trabajo de tesis, se describe la aplicación del proceso de fases de la metodología UWE en las actividades relacionadas con el diseño y desarrollo de un sistema de geo-recomendación.

3.2. Diseño de un sistema de geo-recomendación

En la fase de diseño del sistema, como primera actividad a realizar es la identificación de requisitos, y en UWE se especifican mediante el modelado de casos de uso con UML. La aplicación web permite conocer la información geográfica de una ciudad y la Geo-recomendación. El proceso de información de la ciudad incluye una serie de consultas de establecimientos por determinados criterios como la selección de ciudades, asentamientos y calles. Por otra parte el proceso de geo-recomendación está compuesto en primera parte por las calificaciones con respecto al nivel de tráfico que los usuarios dan a las vialidades (calles o asentamientos) seleccionadas consecutivamente es mostrado al usuario un mensaje con las tres vialidades más similares a la vialidad calificada. Como segunda parte el usuario tiene la capacidad de seleccionar las vialidades similares y verificar la ubicación de estas de acuerdo a las reglas de distancias radiales para la colocación de establecimientos. Las funcionalidades mencionadas anteriormente se expresan a través de un diagrama de casos de uso en la figura 3.1.

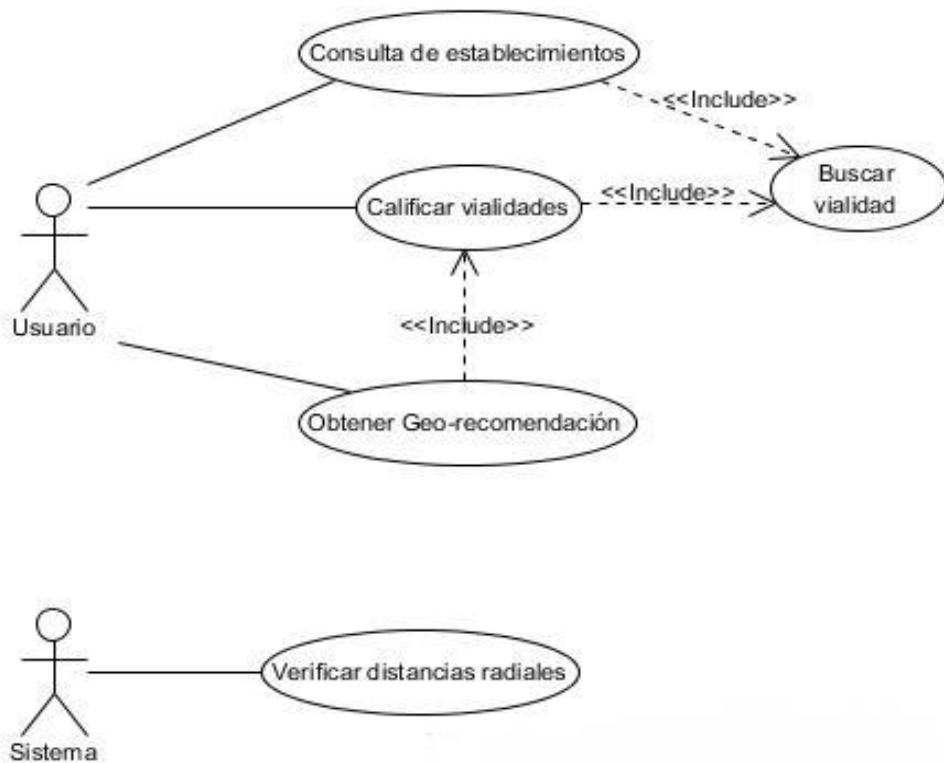


Figura 3. 1. Diagrama de Casos de Uso

Una vez identificados los requisitos es necesario proporcionar una especificación de la implementación de los requisitos por medio de diagramas de clases del modelo, del controlador y la vista.

El primer diagrama de clases es representado en la figura 3.2 en dicha figura se visualizan las clases pertenecientes al modelo de la aplicación. Este diagrama muestra las clases necesarias para representar el modelo utilizado dentro de la aplicación. Las relaciones expresan lo siguiente un establecimiento pertenece a una ciudad, un asentamiento y una calle pero en el caso contrario a una ciudad, asentamiento o calle le pertenecen muchos asentamientos. Por otro lado un usuario recurre a calificar a muchas calles o asentamientos (de manera que sea posible guardar la calificación de las vialidades y generar la recomendación) Finalmente teatro, orfanato, asilo, cine, escuela, estadio, guardería, hospital, iglesia, mercado y templo heredan las características de los establecimientos, es decir son un tipo de establecimientos.

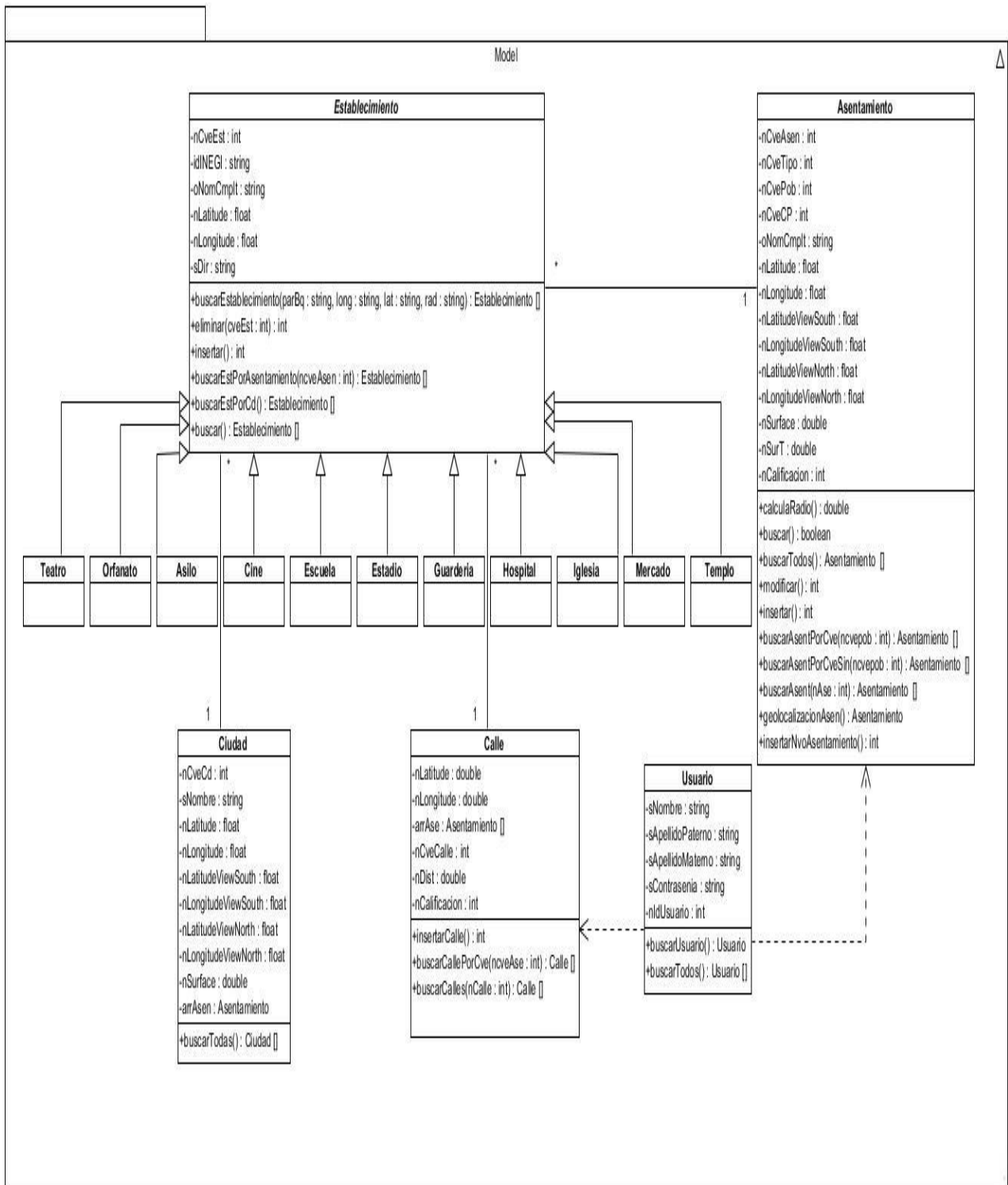


Figura 3. 2 Diagrama de Clases (modelo)

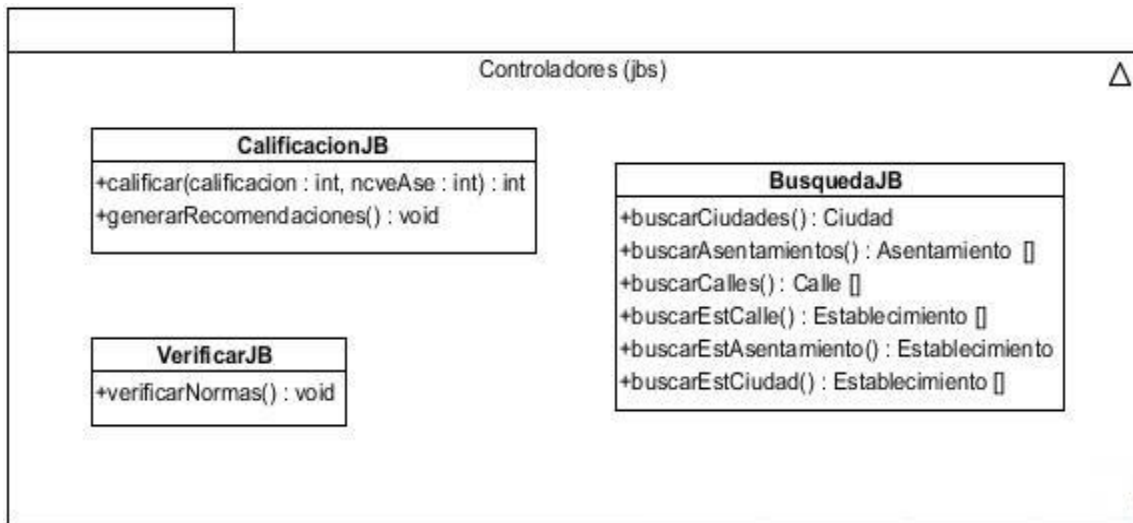


Figura 3. 3 Diagrama de Clases (Controlador)

La figura 3.3 muestra las clases correspondientes a los controladores, los cuales hacen uso de las clases pertenecientes al paquete del modelo. Estos controladores se encargan básicamente de realizar las búsquedas de establecimientos, calificar las vialidades, verificar las vialidades con base a las normas de distancia radial y generar la recomendación.

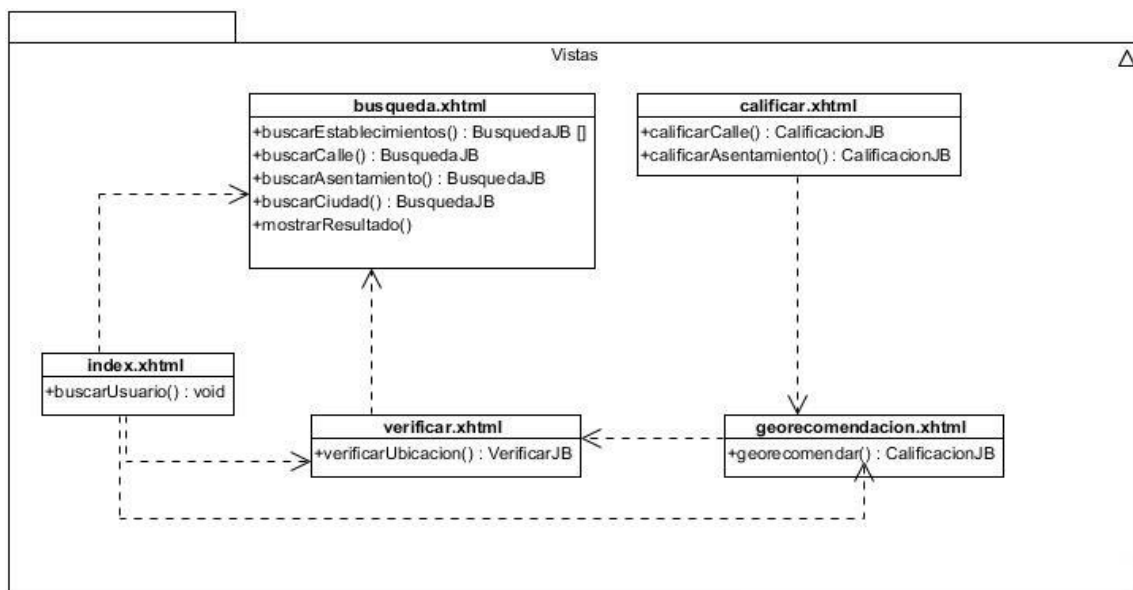


Figura 3. 4 Diagrama de Clases (Vista)

Finalmente la figura 3.4 se muestran las vistas utilizadas en el sistema de georecomendación las cuales hacen uso de los métodos pertenecientes a las clases de los controladores.

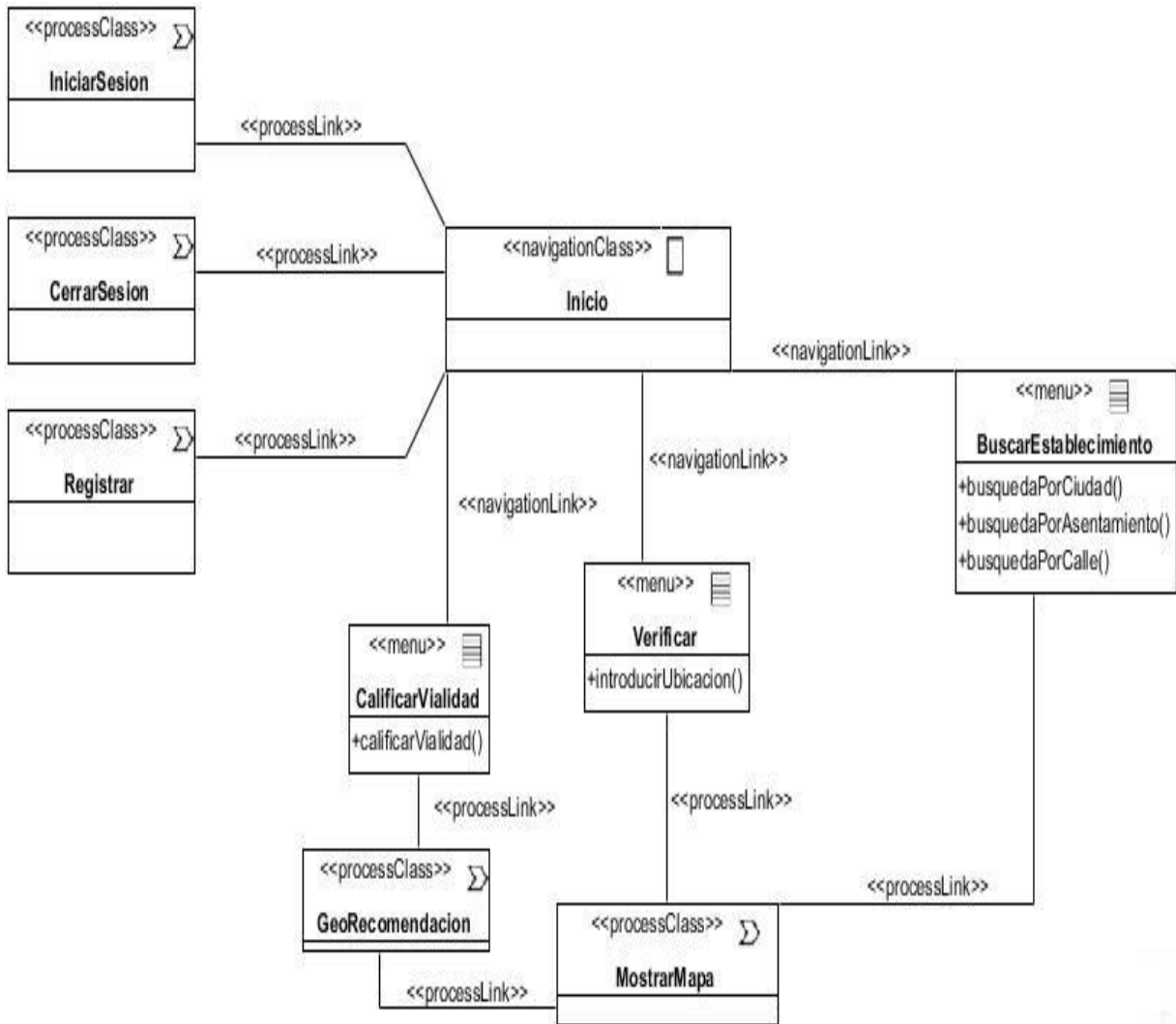


Figura 3. 5 Modelo de navegación

Por otro lado, la figura 3.5 muestra el diagrama de navegación que representa la integración existente entre los componentes de la aplicación para proveer funcionalidad al usuario, con el objetivo de realizar la geo-recomendación.

La aplicación parte de inicio que es la vista principal, aquí se alojan todas las vistas disponibles; donde el usuario después de ser autenticado selecciona la opción de consultar la información de los tipos de establecimientos pertenecientes a una ciudad, asentamiento o calle. A su vez el usuario introduce una ubicación en formato longitud/latitud y verifica si cumple las normas de distancias radiales. Por otro lado, calificar las vialidades (calles y asentamientos) en cuanto al nivel de tráfico y obtener las tres vialidades más similares de acuerdo a las previamente calificadas y verificar si cumplen con las normas de distancias radiales.

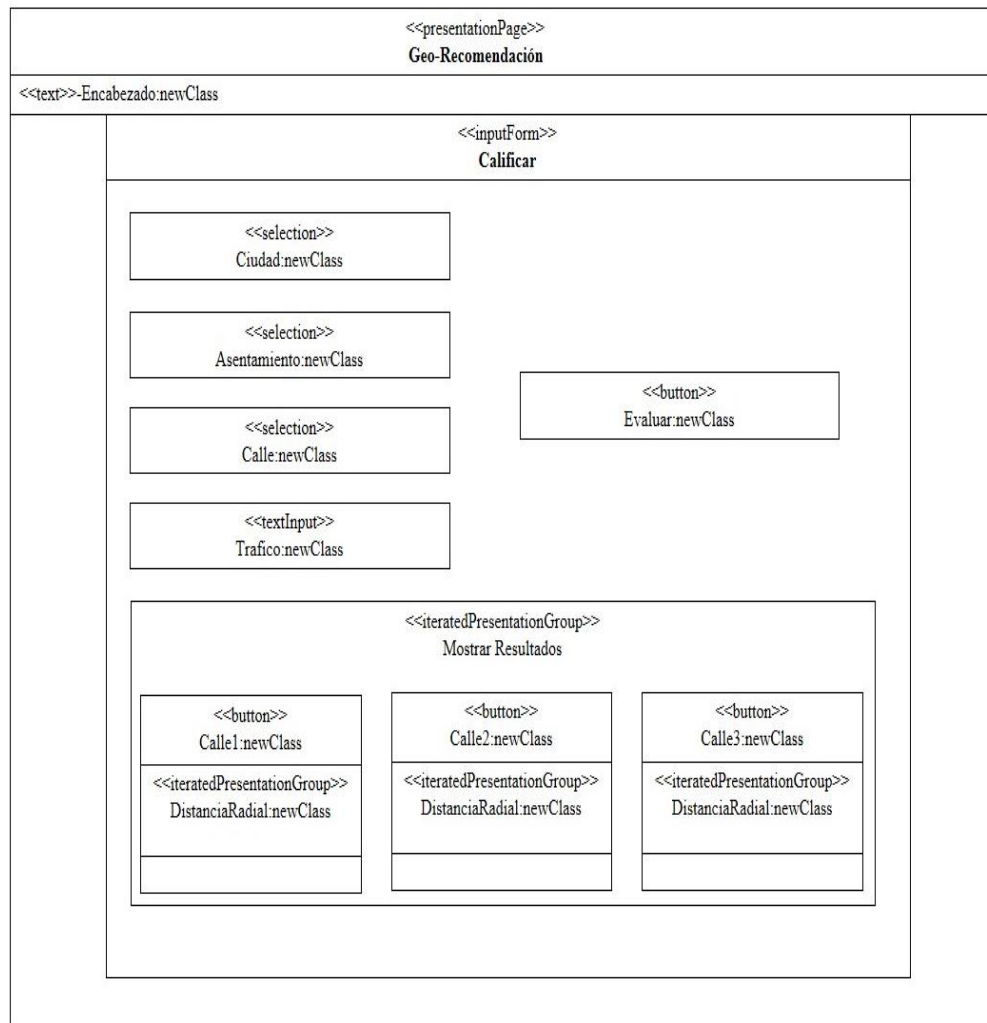


Figura 3. 6 Modelo de Presentacion (Geo-recomendación)

3.3. Finalmente se presenta el modelo de presentación de la geo-recomendación. En la figura 3.6 se ilustra el ejemplo de la calificación de acuerdo al nivel de tráfico de una calle. Para esto es necesario seleccionar una ciudad, conforme a la selección de esta se muestran los asentamientos pertenecientes a esa ciudad, posteriormente se selecciona un asentamiento y a su vez se actualizan las calles relacionadas a ese asentamiento. En este momento el usuario emite una calificación de acuerdo a su percepción del nivel de tráfico y el sistema revela las calles con similitud a la evaluada. Finalmente es posible evaluarlas con base a las normas de distancias radiales.**Arquitectura del sistema de geo-recomendación.**

El sistema desarrollado propone una arquitectura basada en un diseño de capas, para el establecimiento de puntos de venta. Donde las funcionalidades de sus componentes e interrelaciones están distribuidas para un mejor mantenimiento y escalabilidad; el cual permite que sus tareas y responsabilidades se encuentran distribuidas.

La arquitectura presentada posibilita la identificación de zonas adecuadas para establecer puntos de venta por medio de los sistemas de recomendación, sistemas de información geográfica y dinámica de sistemas; a partir de los intereses de los usuarios. Actualmente se considera que un factor clave para obtener el éxito comercial es la ubicación de los negocios.

El diseño de la arquitectura del sistema concede entender la solución propuesta por medio de la descripción de los componentes que la estructuran, la relación entre estos y su flujo de su información.

El objetivo del sistema es auxiliar al usuario mediante la selección de áreas óptimas para el establecimiento de puntos de venta. La arquitectura general se muestra en la Figura 3.7, donde se exhiben los elementos internos que integran la arquitectura del sistema desarrollado en este trabajo de tesis. Cada componente tiene una función que se explica cómo sigue.

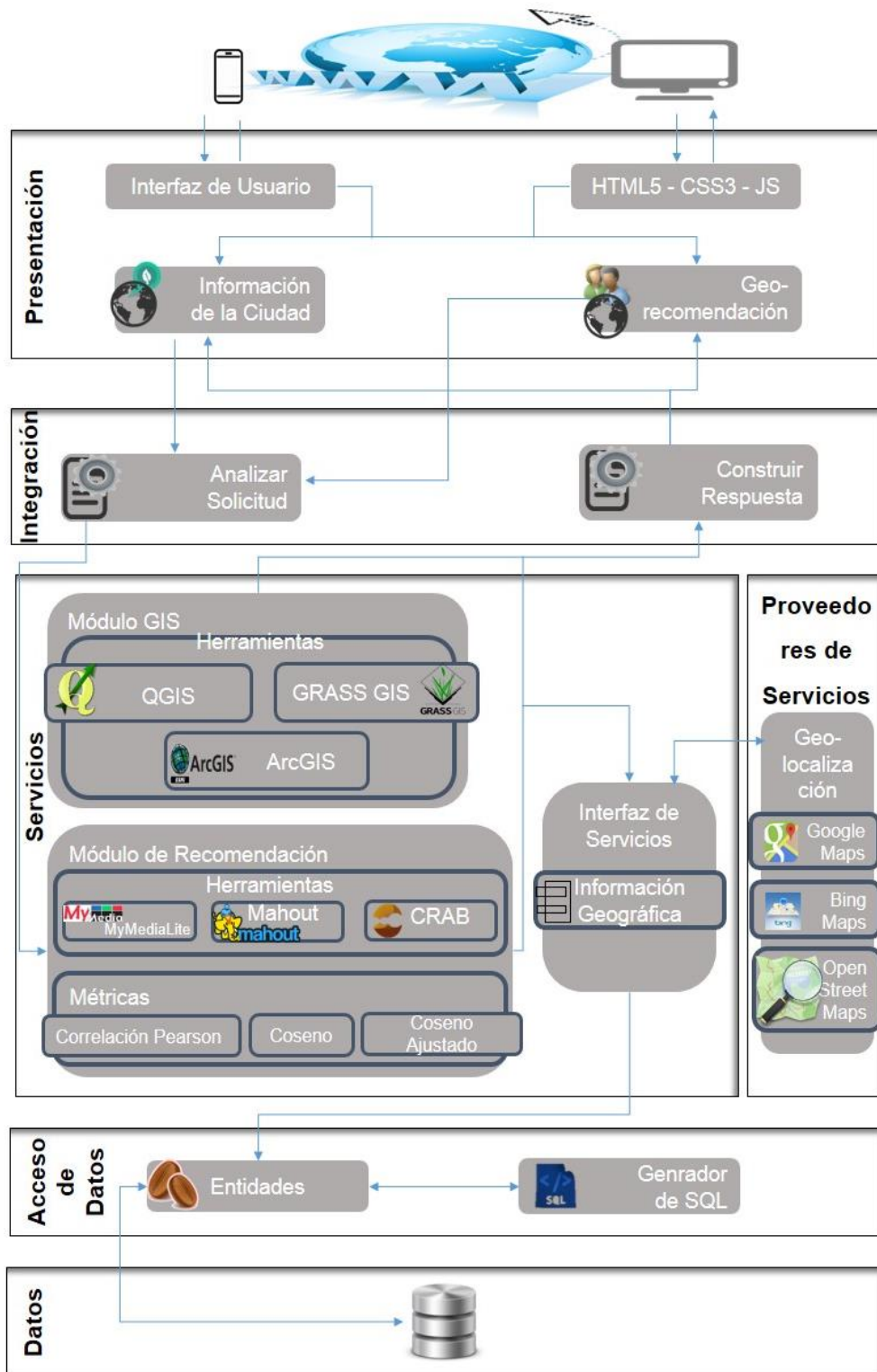


Figura 3. 7 Arquitectura del sistema de geo-recomendación.

3.3.1. Descripción de las capas de la arquitectura.

La arquitectura propuesta se compone de 6 capas. Las capas superiores son más cercanas al usuario, lo que requiere un mayor nivel de abstracción para su implementación, mientras que las capas inferiores son menos cercanas al usuario, lo que supone un menor nivel de abstracción en su implementación. A continuación se describe de manera precisa la función de cada una de las capas en la arquitectura.

Capa de Presentación: La capa de presentación actúa como medio de comunicación entre los resultados obtenidos por las demás capas para los usuarios. Dentro de ella el usuario conoce la información de la ciudad, enviar la dirección del predio en forma de latitud y longitud para establecer un punto de venta; además de mostrar la recomendación final en formato de mapa. Actúa como la interfaz, en donde el usuario envía la dirección del predio y recibir la recomendación en formato de mapa.

Capa de Integración: Esta capa permite re direccionar las solicitudes a los servicios que fueron solicitados en la capa de presentación. Así como la construcción de las respuestas.

Capa de Servicios: En esta capa se encuentra gran parte de operaciones con las cuales funciona el sistema. En ella se ubican los módulos de SIG, de recomendación e interfaz de servicios. Es importante mencionar que esta capa realiza el trabajo de ofrecer la recomendación así como también permite tener la información necesaria de las restricciones relacionadas con los puntos de venta.

Proveedores de Servicios: Dentro de esta capa se encuentran las entidades que prestan el servicio de Geo-localización. Estas ofrecen un servicio Web de aplicaciones de mapas que permite obtener la ubicación geográfica a partir de la devolución de un radio preciso de localización. Algunos proveedores de servicio de localización son OpenStreetMap™ API, Bing™ Maps Geocode y Google™ Maps. Estos son servicios Web REST que regresan los datos como archivos JSON o archivos XML.

Capa de Acceso de Datos: La capa de acceso de datos busca y guarda la información en la Base de Datos que le solicita la capa de servicios. Es posible la encapsulación de tareas a través de las distintas entidades y la ejecución de las operaciones de inserción, eliminación, consulta y actualización por medio del generador de instrucciones basadas en sentencias SQL.

Capa de Datos: Esta capa almacena información acerca de las poblaciones, así como sus asentamientos y establecimientos (cines, escuelas, orfanatos, asilos, hospitales, templos, guarderías, mercados, auditorios, estadios y teatros)

En esta arquitectura, cada módulo tiene una función bien definida la cual se describe a continuación:

Módulo SIG: Este módulo es responsable de construir el modelo geográfico con base en las características geográficas obtenidas a partir de la consulta de las características geográficas de la zona y de establecimientos que son solicitadas a la interfaz de servicios. Algunas de las herramientas a utilizar en este módulo son QGIS, GRASS GIS y ArcGIS™. Los dos primeros son software de distribución pública y mientras que el tercero es de licencia comercial. Dichas herramientas permiten la creación de modelos geográficos a partir de la unificación y el análisis de datos geoespaciales.

Módulo de Recomendación (RS): Es el responsable de ofrecer las sugerencias al correlacionar el perfil del usuario con respecto a otros perfiles. Así como también es responsable de evaluar que la ubicación de un punto de venta es la adecuada. Dentro de las distintas clasificaciones de los sistemas de recomendación se encontró que la más adecuada para construir la recomendación es la perteneciente al filtrado colaborativo; ya que mide la similitud entre los distintos puntos de ubicación a partir de métricas que calculen la similitud. Los algoritmos que calculan la similitud son correlación de Pearson, coseno y coseno ajustado. Lo aconsejable es la creación de modelos de recomendación con diferentes algoritmos para el cálculo de similitud y se realicen evaluaciones de las recomendaciones a partir de las métricas recall y precisión para una mejor decisión del modelo de recomendación a utilizar. Algunas APIs especializadas para el desarrollo de sistemas de recomendación basados en filtrado colaborativo son Apache Mahout, MyMediaLite y CRAB. Estas APIs son de distribución pública y los lenguajes compatibles con ellas son Java y Python; Java para Apache Mahout y Python para las últimas dos.

Interfaz de servicios Web: Permite conectar con los servicios Web de geo-localización y de información de establecimientos; posibilita la creación de contenidos completos por medio de combinar datos provenientes de distintos servicios Web.

3.3.1.1. Descripción de componentes.

En esta arquitectura, cada módulo tiene una función bien definida la cual se describe a continuación:

Módulo SIG: Este módulo es responsable de construir el modelo geográfico con base en las características geográficas obtenidas a partir de la consulta de las características geográficas de la zona y de establecimientos que son solicitadas a la interfaz de servicios.

Módulo de Recomendación (RS): Es el responsable de ofrecer las sugerencias al correlacionar el perfil del usuario con respecto a otros perfiles. Así como también es responsable de evaluar que la ubicación de un punto de venta es la adecuada. Pide el histórico de las demás recomendaciones a la capa de datos y por medio de las métricas de correlación de Pearson, coseno y coseno ajustado es capaz de ofrecer la geo-recomendación.

Interfaz de servicios Web: Permite conectar con los servicios Web de geo-localización y de información de establecimientos; posibilita la creación de contenidos completos por medio de combinar datos provenientes de distintos servicios Web.

3.3.2. Funcionamiento de la arquitectura.

La figura 3.8 muestra el flujo de trabajo de un generador de sistemas de recomendación.

Este flujo de trabajo describe la funcionalidad de la arquitectura propuesta como se describe a continuación:

- 1) El usuario hace una solicitud HTTP a través del sistema de Geo-recomendación; dicha solicitud permite conocer la información geográfica de una ciudad y obtener la Geo-recomendación.

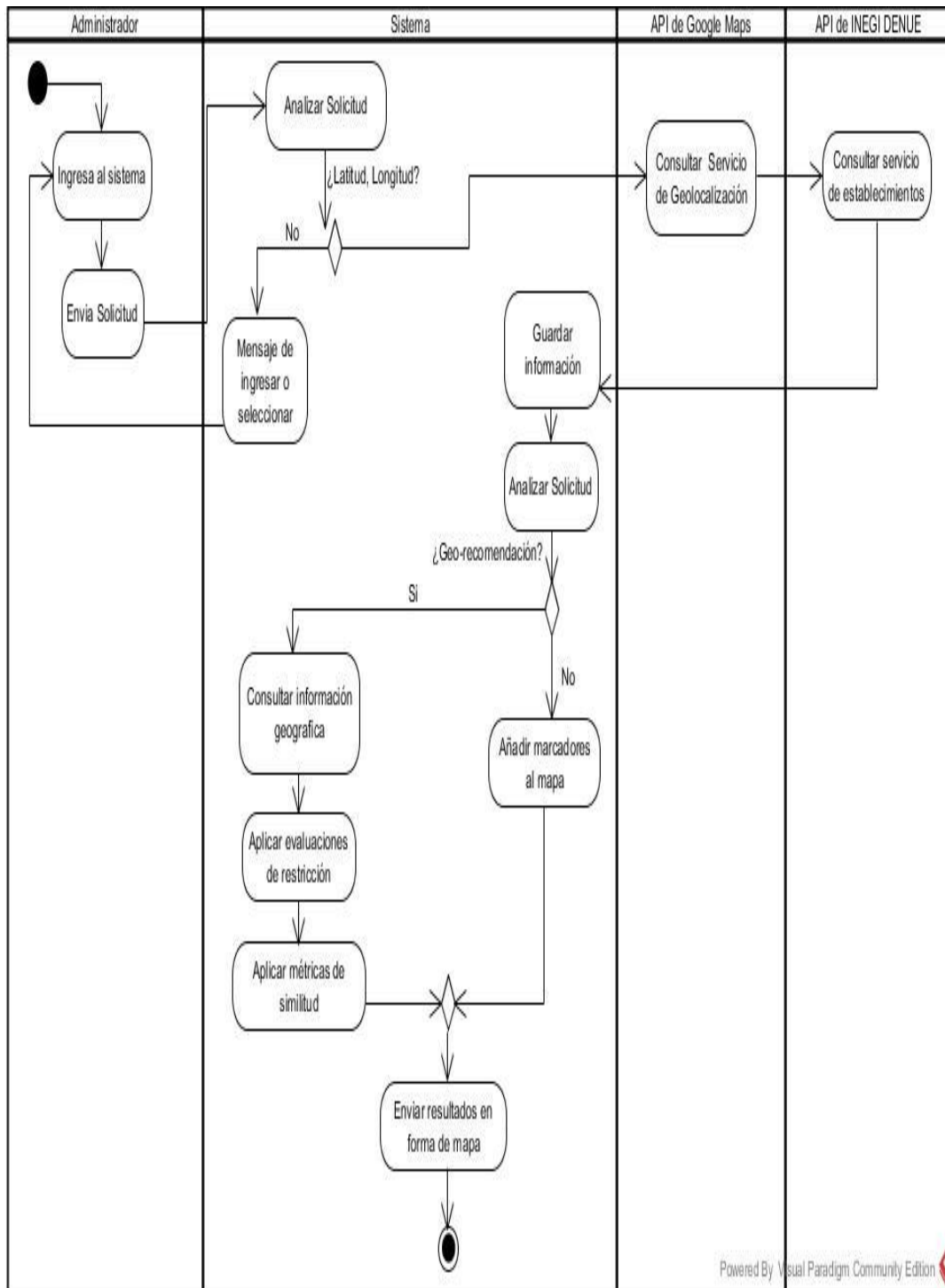


Figura 3. 8 Diagrama de actividades del sistema

- 2) Sin importar la opción seleccionada, todas las solicitudes pasan por el analizador de solicitud, el cual es el responsable de redireccionar la solicitud al módulo correspondiente.
- 3) Este paso tiene distintos flujos de acción de acuerdo a la opción seleccionada.
 - a) En este punto el módulo de Sistema de Información Geográfica (SIG) es el responsable de solicitar cierta información a la interfaz de servicios. El proceso de consulta de información es descrito a partir del punto 4, ya que el módulo de sistema de recomendación necesita la información.
 - b) El módulo de sistema de recomendación solicita la información del módulo de sistema de información geográfica para la evaluación de colocación de puntos de venta. Este módulo es el responsable de ofrecer las recomendaciones a partir de los algoritmos correlación de Pearson, coseno y coseno ajustado.
- 4) El módulo SIG solicitada a la interfaz de servicios conocer la información geográfica de un determinado punto.
- 5) Se realiza la consulta a la API de Geo-localización de Google™ Maps.
- 6) Posteriormente se solicita a la interfaz de servicios conocer cuáles son los establecimientos que se encuentran a los alrededores de punto en cuestión.
- 7) A través de la consulta a la API de INEGI DENUÉ se obtiene la información de establecimientos alrededor de ese punto.
- 8) Las solicitudes anteriores son responsables de enviar su información a las entidades de la capa de acceso de datos.
- 9) Es necesario construir la solicitud en formato SQL para que las entidades sean capaces de llevar a cabo la solicitud de almacenamiento.
- 10) Se ejecuta esa solicitud a través de la conexión entre la capa de acceso de datos a la capa de datos.
- 11) Al ser ejecutadas las solicitudes, estas se envían al generador de respuestas.
- 12) De acuerdo a lo seleccionado desde un principio, el usuario obtendrá su respuesta.

3.4. Esquema relacional de la base de datos.

Antes mostrar al esquema relaciona de base de datos. Es necesario introducir a nuestro amigo lector en ciertos términos que le ayuden a comprender mejor la decisión del diseño presentado por medio de una serie de premisas que son relatadas a continuación:

- a) Un municipio está conformado por códigos postales.
- b) Los códigos postales pertenecen a más de un asentamiento.
- c) Los asentamientos pertenecen solo a un municipio y código postal.
- d) Los asentamientos tienen una clasificación (colonia, fraccionamiento, unidad habitacional, barrio, ranchería, ampliación, condominio, congregación, parque industrial y residencial)
- e) Un asentamiento está compuesto por muchas calles pero en el caso de calles, ellas solo le pertenecen a un asentamiento.
- f) Tanto para asentamiento y calle existe un nivel de tráfico, el cual es otorgado en función de la percepción de los usuarios.
- g) Para el establecimiento de puntos de venta de tipo gasolinera es necesario almacenar la información con respecto a determinados establecimientos, los cuales son asilos, auditorios, cines, escuelas, estadios, guarderías, hospitales, iglesias, mercados, orfanatos, teatros y templos para cada uno de los asentamientos que integra un municipio.

El presente esquema relacional de base de datos contempla 20 tablas, de las cuales cuatro corresponden a información relacionada con los municipios pertenecientes a un estado de la república mexicana. Dichas tablas son “poblado”, “codigopostal”, “asentamiento” y “tipoasentamiento”. En la tabla “poblado” se guarda las longitudes, latitudes, radios y nombres de los municipios. Por otro lado en “codigopostal” se encuentran las claves de los códigos postales que integran a los poblados. En la tabla “tipoasentamiento” se concentran los tipos de clasificaciones de los asentamientos que integran a un municipio, dichas clasificaciones son: colonia, fraccionamiento, unidad habitacional, barrio, ranchería, ampliación, condominio, congregación, parque industrial y residencial. Además en “asentamiento”, se encuentra

almacenada la información de los asentamientos que constituyen a un municipio, en dicha tabla se contemplan los nombres, longitudes, latitudes, radios y tipo de asentamiento. Por medio de ellas es posible mantener actualizado al sistema con información referente a los establecimientos.

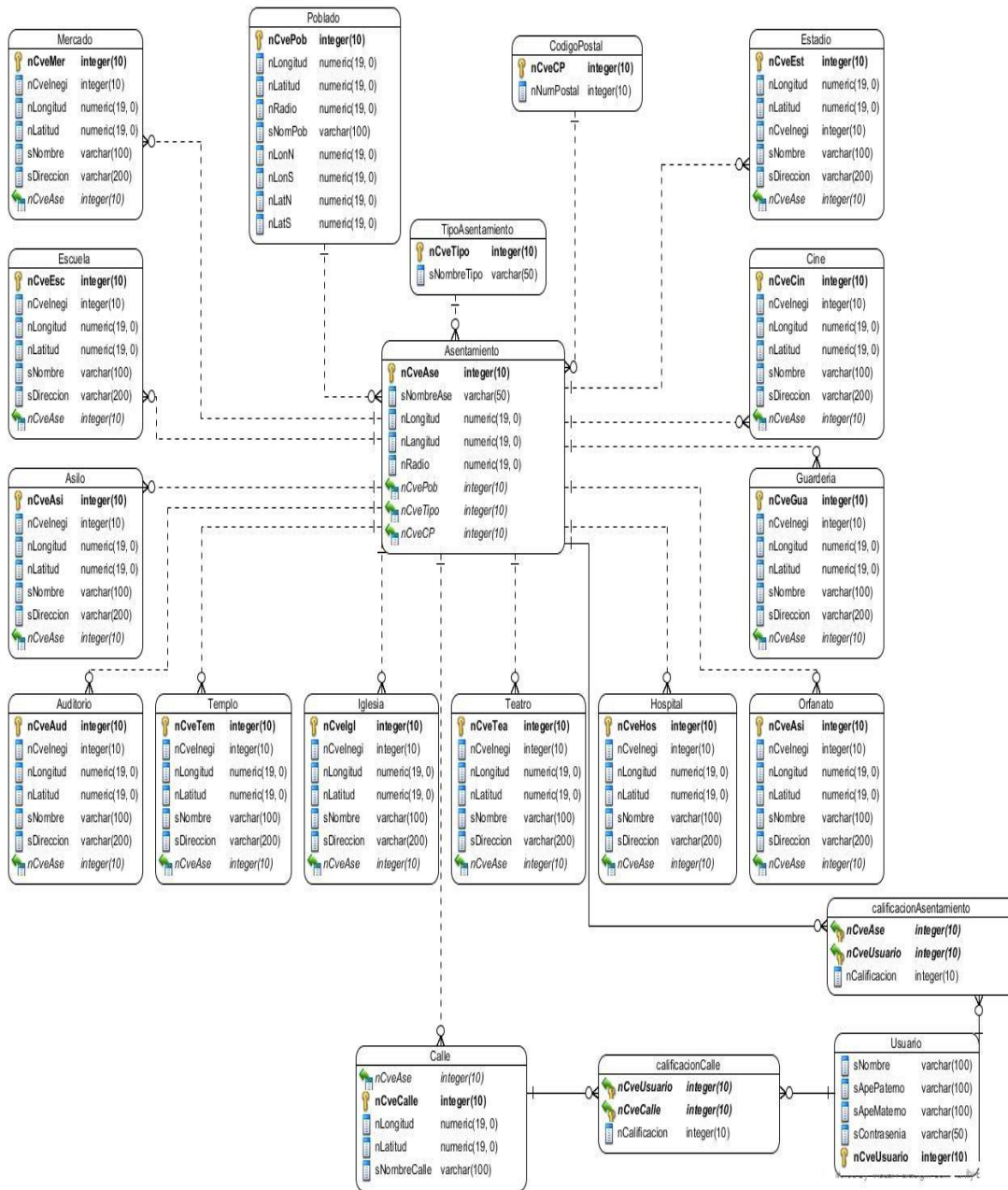


Figura 3. 9 Esquema de Base de Datos.

Por otro lado la tabla “calle” guarda el nombre, longitud, latitud de la calle así como el asentamiento al que pertenece. En el caso de las tablas “calificaciónCalle” y “calificacionAsentamiento” guardan las calificaciones que los usuarios dan a las calles y asentamientos. Por otro lado la tabla de “usuario” guarda los datos del usuario (nombre, apellidos y contraseña. Finalmente las tablas “asilo”, “auditorio”, “cine”, “escuela”, “estadio”, “guarderia”, “hospital”, “iglesia”, “mercado”, “orfanato”, “teatro” y “templo” se han condensado la longitud, latitud, nombre del establecimiento en cuestión, dirección y clave del asentamiento al que pertenecen.

La figura 3.9 muestra los campos y las relaciones entre las tablas que conforman el repositorio de información.

3.5. Identificación de fuentes de información para el proceso de geo-recomendación.

Antes de contar con una estructura de almacenamiento en forma de tablas, fue necesario comprender que fuentes de información eran de vital importancia para el proceso de geo-recomendación. Dichas fuentes fueron las siguientes:

1. Reglamento para las acciones de construcción, instalación, conservación y operación de estaciones de servicio en gasolinera y carburación.
2. Google™ Maps Geolocation API.
3. API DENUÉ del INEGI.

Como primer paso fue consultar la existencia de normativas o restricciones para la colocación de puntos de ventas existen ciertas normativas y regulaciones a cumplir. Dichas se encuentran establecidas en el en el artículo 10 del Reglamento para las acciones de construcción, instalación, conservación y operación de estaciones de servicio en gasolinera y carburación. El cual indica que la ubicación del predio respeta una distancia mínima de resguardo de 300 metros radiales de centros de concentración masiva, tales como escuelas, hospitales, orfanatos, guarderías, asilos; así como a 150 metros radiales de mercados, cines, teatros, estadios, auditorios y templos [37]. Estas restricciones de ubicación indican que las estaciones de servicio respetan cierta distancia con respecto a museos, escuelas, hospitales, orfanatos, guarderías, teatros, cines, auditorios y templos.

Para obtener esta información fue necesario conocer los códigos y asentamientos que integran los municipios. Dentro del territorio mexicano la única fuente que cuenta con determinada información es el Servicio Postal Mexicano (SEPOMEX), esta institución no cuenta con una API de distribución pública que permita por medio de la conexión de servicios web consultar los códigos postales y asentamientos de un poblado. Por otra parte, es posible descargar esa información en un archivo Excel.

Por otro lado era necesario conocer la geográfica de los poblados, a través Google™ Maps Geolocation API. Este servicio devuelve una ubicación en función de información. De manera que para guardar la información se realizó una conexión a la API de Google™ Maps para obtener por cada asentamiento la su información geográfica (longitud y latitud).

No obstante la información recabada hasta ese momento no estaba completa, era necesario conocer los establecimientos correspondientes a cada asentamiento. Mediante la API DENUÉ del INEGI es posible consultar datos de identificación, ubicación, actividad económica y tamaño de un determinado asentamiento. Para acceder al servicio es necesario indicar la longitud, latitud y radio del asentamiento. A través de una conexión por medio de una llamada al servicio proporcionado por la API DENUÉ del INEGI y guardar la información de los establecimientos, así como también la construcción de un mecanismo para la actualización de la información.

3.6. Métricas de similitud para el proceso de geo-recomendación.

Para el cálculo de similitud en el proceso de geo-recomendación es necesario utilizar una métrica de similitud, existen diversas métricas. Para el presente trabajo se consideran tres métricas de similitud; coseno ajustado, correlación de Pearson y coseno.

3.6.1. Coseno Ajustado.

Se utiliza esta métrica cuando un usuario tiene opiniones muy distintas con respecto a los productos evaluados. De manera que para eliminar ese inconveniente se restan las calificaciones promedio para cada usuario de las calificaciones promedio de los productos [50]. La siguiente fórmula expresa el cálculo de coseno ajustado:

$$\text{sim}(i, j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{item}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{item}} (R_{jc} - \bar{R}_j)^2}}$$

Donde I_{ij} representa los ítems que el usuario i y j correlaciona y \bar{R}_i, \bar{R}_j es el promedio de usuario i y j . Es decir primero se calculan los promedios de los usuarios, se calcula el denominador y numerador tomando en cuenta el promedio obtenido.

3.6.2. Coeficiente de Correlación de Pearson.

El coeficiente de correlación de Pearson, el cual mide la desviación entre valoraciones comunes y medias que los usuarios hacen a un respectivo producto. El intervalo para medir la similitud va de -1 a 1 , donde el valor de -1 indica que la correlación es negativa, el valor de 0 indica la inexistencia de correlación y el valor de 1 indica la corrección de una variable con respecto a si misma [50].

$$\text{sim}(i, j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{jc} - \bar{R}_j)^2}}$$

Es decir, primero se realiza el cálculo de las medias aritméticas \bar{R}_i, \bar{R}_j . Posteriormente se calcula la covarianza $\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)$, posteriormente las desviaciones típicas $\sqrt{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{jc} - \bar{R}_j)^2}$ para i y j .

3.6.3. Coseno.

Toma el producto escalar de dos vectores. El rango de valores que nos da esta métrica va desde -1 y 1 , dichos valores nos indican la dirección y la magnitud de la asociación entre dos elementos o usuarios. Donde magnitud nos indica el nivel de asociación entre dos elementos y la dirección nos dice cómo varían las variables [51].

$$\text{sim}(i, j) = \frac{I \cdot J}{\|I\| \|J\|} = \frac{\sum_{c \in I_{item}} R_{ic} R_{jc}}{\sqrt{\sum_{c \in I_{item}} R_{ic}^2} \sqrt{\sum_{c \in I_{item}} R_{jc}^2}}$$

Donde I, J representan las calificaciones en vectores que los usuarios i y j hacen en n ítems. Item corresponde a todos los ítems. R_{ic} y R_{jc} son las evaluaciones del usuario i y j en el ítem c .

3.7. Proceso de Geo-recomendación.

Con el objetivo de ofrecer una recomendación de tipo geográfica, primero fue necesario establecer restricciones para la colocación de establecimientos las cuales se encuentran de la regulación anteriormente mencionada; estas restricciones actúan como primer filtro para considerar que la ubicación cumple con las estipulaciones establecidas. El segundo filtro es aplicar la biblioteca de Apache Mahout la cual es ideal para trabajar con grandes volúmenes de información.

Como primer paso en el proceso de geo-recomendación es referente a la construcción de matrices con la información de la sucursal (latitud, longitud, histórico de ventas) y de los usuarios responsables del historial de venta. Para ello es ejemplificado el proceso de recomendación mostrando los resultados obtenidos con cada una de las métricas anteriores. En la tabla 3.1 se tienen los siguientes usuarios los cuales calificaron determinadas calles con base al nivel de tráfico, en la tabla se maneja la calificación en una escala del 1 al 5, en donde una calificación de 5 indica un mayor nivel de tráfico.

Tabla 3. 1 Tabla de calificaciones

	Calle 1	Calle 2	Calle 3	Calle 4	Calle 5
Lisa Rose	2.5	3.5	3.0	3.5	2.5
Gene Seymour	3.0	3.5	1.5	5.0	3.5
Michael Phillips	2.5	3.0	0	3.5	0
Claudia Puig	0	3.5	3.0	4.0	2.5
Mick LaSalle	3.0	4.0	2.0	3.0	2.0

Después de haberse estructurado la información, se elaboran las recomendaciones basadas en usuarios y en ítems; las cuales utilizan el concepto similitud. Algunas de las implementaciones de este concepto en Apache Mahout son las métricas de coseno ajustado, coeficiente de

correlación. Se calcula bajo la métrica de coeficiente de relación de Pearson. Como primer paso se realizan las recomendaciones basadas en ítems, para ellos se toman como valoraciones el nivel de tráfico.

Como primer caso de la construcción de la recomendación se obtiene el promedio de las vialidades, el cual se muestra en la tabla 3.2.

Tabla 3. 2 Cálculo de Promedios

	Calle 1	Calle 2	Calle 3	Calle 4	Calle 5
Calle 1	2,50	3,50	3,00	3,50	2,50
Calle 2	3,00	3,50	1,50	5,00	3,50
Calle 3	2,50	3,00	0,00	3,50	0,00
Calle 4	0,00	3,50	3,00	4,00	2,50
Calle 5	3,00	4,00	2,00	3,00	2,00
PROMEDIO	2,20	3,50	1,90	3,80	2,10

Posteriormente se calcula el numerador (tabla 3.3) y denominador de la fórmula de Pearson (tablas 3.4 y 3.5):

Tabla 3. 3 Cálculo de numerador

Numerador Calle(1,2)	0,25
Numerador Calle(1,3)	-2,9
Numerador Calle(1,4)	-0,3
Numerador Calle(1,5)	-0,35

Numerador Calle(2,3)	1
Numerador Calle(2,4)	-0,25
Numerador Calle(2,5)	1
Numerador Calle(3,4)	-0,1
Numerador Calle(3,5)	4,3
Numerador Calle(4,5)	2,35

Para obtener el denominador es necesario primero calcular la suma cuadrática correlación para cada una de las calles. Los resultados se muestran en la tabla 3.4 y en la tabla 3.5 se presentan los resultados de los denominadores.

Tabla 3. 4 Cálculo de Suma Cuadrática Correlación

Suma Cuadrática Correlación calle 1	6,3
Suma Cuadrática Correlación calle 2	0,5
Suma Cuadrática Correlación calle 3	6,2
Suma Cuadrática Correlación calle 4	2,3
Suma Cuadrática Correlación calle 5	6,7

Tabla 3. 5 Cálculo denominador

Denominador calle(1,2)	1,77482393
Denominador calle(1,3)	6,2498

Denominador calle(1,4)	3,80657326
Denominador calle(1,5)	6,49692235
Denominador calle(2,3)	1,76068169
Denominador calle(2,4)	1,07238053
Denominador calle(2,5)	1,83030052
Denominador calle(3,4)	3,77624152
Denominador calle(3,5)	6,44515322
Denominador calle(4,5)	3,92555729

Más adelante se presentan los resultados de la matriz normalizada en la tabla 3.6 y se eligen los puntos de venta más adecuados para ser recomendados al usuario.

Tabla 3. 6 Matriz normalizada

	Calle 1	Calle 2	Calle 3	Calle 4	Calle 5
Calle 1	1,00	0,14	-0,46	-0,08	-0,05
Calle 2	0,14	1,00	0,57	-0,23	0,55
Calle 3	-0,46	0,57	1,00	-0,03	0,67
Calle 4	-0,08	-0,23	-0,03	1,00	0,60
Calle 5	-0,05	0,55	0,67	0,60	1,00

3.8. Diseño del sistema de Geo-recomendación.

Como parte del diseño del sistema de geo-recomendación se desarrollaron las siguientes funcionalidades, las cuales se describen a continuación. Como primera funcionalidad se tiene la búsqueda de establecimientos de un cierto tipo (escuelas, guarderías, teatros, orfanatos, asilos, hospitales, cines, entre otros.) pertenecientes a un asentamiento; a partir de la selección de ciudad, tipo de asentamiento y el nombre del asentamiento. Esta búsqueda es interesante ya que la selección de estos parámetros está fuertemente ligada con la visualización de los resultados; es decir que al seleccionar una ciudad automáticamente actualiza la lista de los asentamientos correspondientes a esa ciudad. Después de haberse seleccionado los parámetros, estos se representan dentro de un mapa a partir de marcadores.

Como segunda funcionalidad se presenta la búsqueda de puntos de venta viables a partir del ingreso de latitud y longitud. Donde como primer filtro se aplican las restricciones pertenecientes a la regulación vigente; las cuales determinan cuándo una ubicación es adecuada para la colocación de puntos de venta. Finalmente como segundo filtro se aplica el proceso de geo-recomendación el cual permite predecir que ubicaciones son las más idóneas. Una vez producidas las predicciones, estas se representan dentro de un área destinada para el mapa a partir de circunferencias de distintos tonos (rojo, amarillo y verde). Es decir es posible observar las regiones con mayor, media y menor probabilidad de establecer un punto de venta. Por otro lado estas circunferencias ofrecen ver el detalle de los resultados (probabilidad, latitud, longitud y establecimientos a sus alrededores).

Otra de las funcionalidades es la búsqueda radial de puntos de venta, la cual permite identificar a partir de un punto (longitud y latitud) y una distancia radial ubicaciones apropiadas para establecer un punto de venta; donde muchas veces las superficies de estas búsquedas implican más de un asentamiento derivando en evaluaciones de restricciones para cada uno de los puntos que integren dicha superficie y en algunas ocasiones encuentra más de un punto de venta óptimo.

Finalmente la búsqueda por código postal hace evaluaciones por todas las superficies de los asentamientos que pertenecen al código postal y al igual que la búsqueda radial en determinadas circunstancias devuelve más de un punto de venta adecuado. De manera que en casos con más

de un punto de venta optimo, la decisión depende del usuario. Hasta este momento el sistema no es capaz de solucionar el problema de arranque en frío, pero si es posible ofrecer geo-recomendaciones de puntos de venta a partir de los puntos de venta se colocaron.

Una de las restricciones del sistema, es el hecho de que el sistema no es capaz de solucionar el problema de arranque en frío, pero si es posible ofrecer geo-recomendaciones de puntos de venta a partir de los puntos de venta que se colocaron e información con respecto al historial de ventas.

Capítulo 4. Resultados

En este capítulo se muestra un caso de estudio y las pruebas de estudio. Como caso de estudio se presenta un Sistema de Geo-Recomendación GEOREMSYS, el cual es un sistema de recomendación que facilita la recomendación de puntos de venta, el objetivo de RESYGEN es asistir al usuario en el mediático proceso de selección de un punto de venta.

4.1. Caso de estudio en la recomendación de puntos de venta.

Para la validación de la arquitectura se plantearon los siguientes argumentos:

- a) Una empresa desea establecer puntos de ventas en la ciudad de Orizaba, Ver. para comercializar su producto.
- b) Es necesario que cada punto de venta cumpla con la regulación vigente perteneciente al Reglamento para las acciones de construcción, instalación, conservación y operación de estaciones de servicio en gasolinera y carburación.
- c) La empresa desea automatizar el proceso de búsqueda de establecimientos para la colocación de puntos de venta.

Con base en lo anterior; ¿Cómo podría la empresa satisfacer estas condiciones para establecer su punto de venta?

Como primer paso es necesario contar con la información dentro de una geo base de datos en la cual se guardan todas las ubicaciones de los negocios pertenecientes a la localidad. Estas ubicaciones se localizan dentro de un mapa a partir del servicio del API de Google™ Maps. El siguiente punto es la construcción de un módulo que permita la conexión con el API anterior y con el GPS del dispositivo móvil de manera que posibilite el reconocimiento de la ubicación actual. Posteriormente es necesario contar con un componente exclusivo para las empresas en donde ellas ingresen las promociones con las que cuentan de modo que al activarse el módulo sensible a la ubicación como paso continuo se accione el tercer módulo a desarrollar; el cual genera la recomendación basada en la ubicación actual y las preferencias de los usuarios. Contar con una aplicación de este tipo ayudaría en distintas formas a los acelerados estilos de vida de la población así como también al medio ambiente.

GEOREMSYS: Sistema de geo-recomendación para la localización de puntos de venta.

Las funcionalidades de este sistema derivan del primer caso de estudio presentado en este capítulo. El objetivo es auxiliar al usuario mediante la selección de áreas óptimas para el establecimiento de puntos de venta, el cual es una aplicación Web basada en el Marco de trabajo JSF con PrimeFaces. El sistema presenta las siguientes funcionalidades a la búsqueda de establecimientos de un cierto tipo (escuelas, guarderías, teatros, orfanatos, asilos, hospitales, cines, entre otros.) pertenecientes a un asentamiento; a partir de la selección de ciudad, tipo de asentamiento y el nombre del asentamiento. Esta búsqueda es interesante ya que la selección de estos parámetros está fuertemente ligada con la visualización de los resultados; es decir que al seleccionar una ciudad automáticamente actualiza la lista de los asentamientos correspondientes a esa ciudad. Después de seleccionarse, los parámetros, estos se representan dentro de un mapa a partir de marcadores. La figura 4.1 muestra la visualización de la información obtenida a partir de las consultas a los servicios Web de las APIs previamente mencionadas en la sección de los casos de estudio.

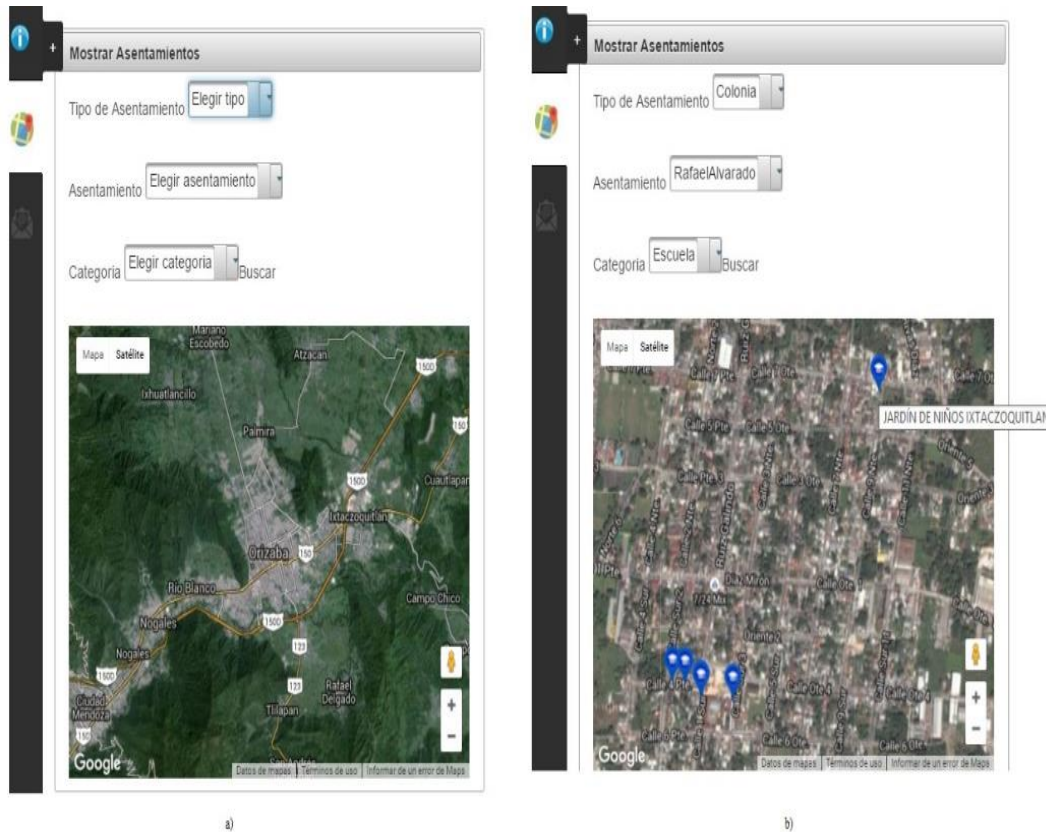


Figura 4. 1 Visualización de establecimientos.

Como segunda funcionalidad se presenta la búsqueda de puntos de venta viables a partir del ingreso de latitud y longitud. Donde como primer filtro se aplican las restricciones pertenecientes a la regulación vigente; las cuales determinan cuando una ubicación es adecuada para la colocación de puntos de venta.

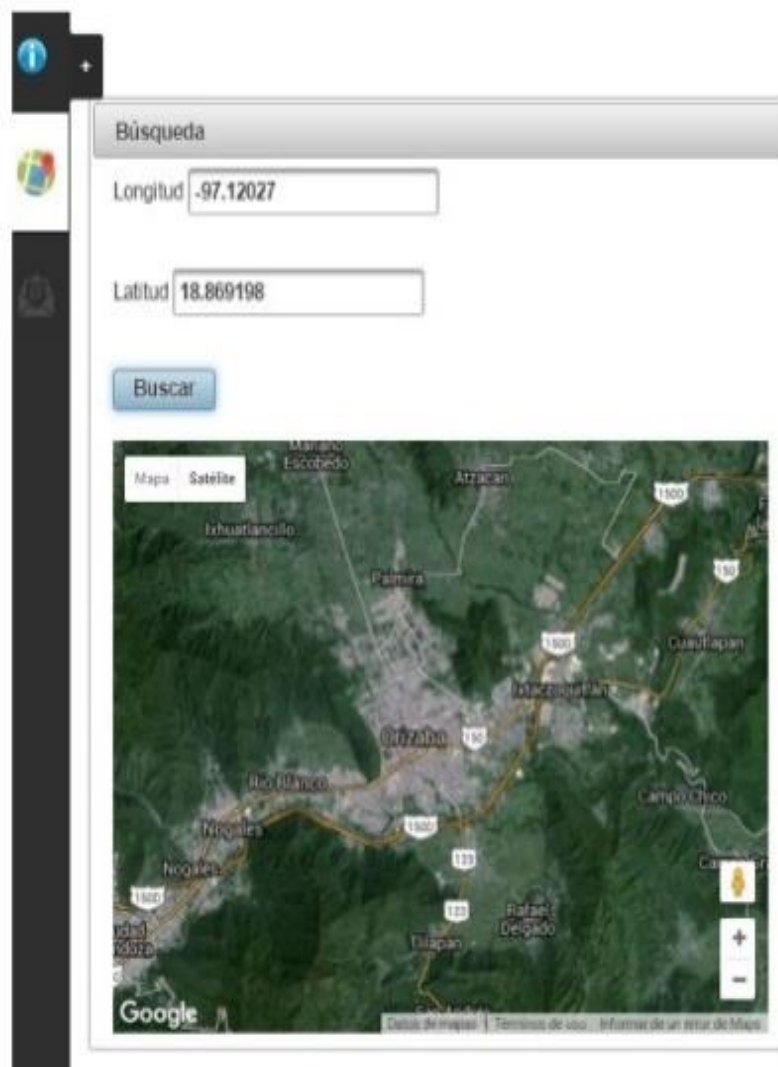


Figura 4. 2 Formulario de Búsqueda.

La Figura 4.2 muestra el formulario de búsqueda de puntos de venta, en él se introduce la dirección que se considera adecuada para establecer un punto de venta, en formato de longitud y latitud. En caso de no ser una ubicación adecuada de acuerdo con las restricciones del reglamento, especificará que la localización no es la adecuada

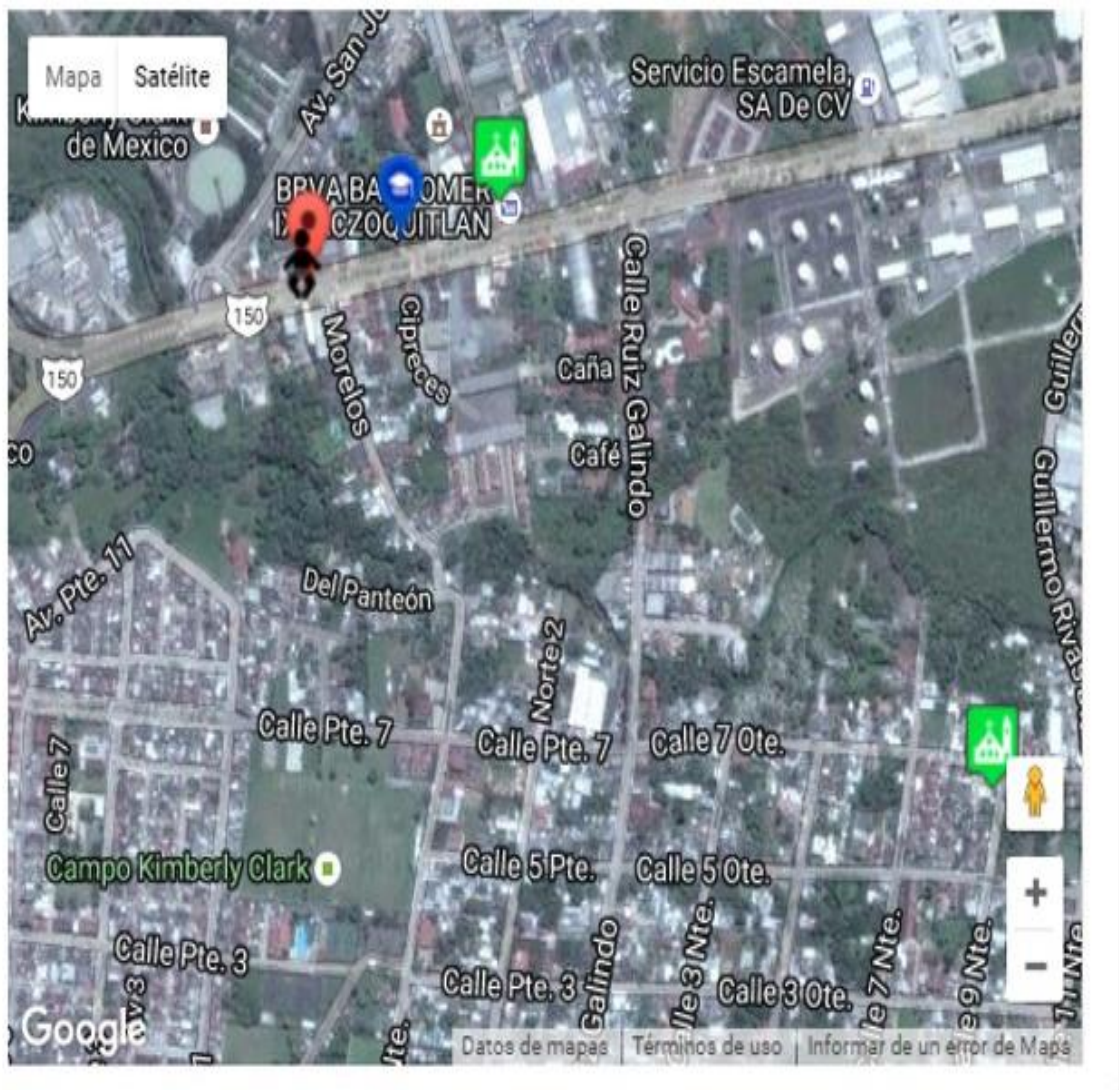


Figura 4. 3 Resultados

La figura 4.3 muestra el caso en donde la ubicación ingresada no satisface las normas radiales de la regulación vigente, en donde se presentan en diferentes iconos los establecimientos identificados.

Finalmente, como segundo filtro se aplica el proceso de geo-recomendación el cual permite predecir que ubicaciones son las más idóneas. Una vez producidas las predicciones, estas se representan dentro de un área destinada para el mapa a partir de circunferencias de distintos tonos (rojo, amarillo y verde). Es decir es posible observar las regiones con mayor, media y menor probabilidad de establecer un punto de venta.

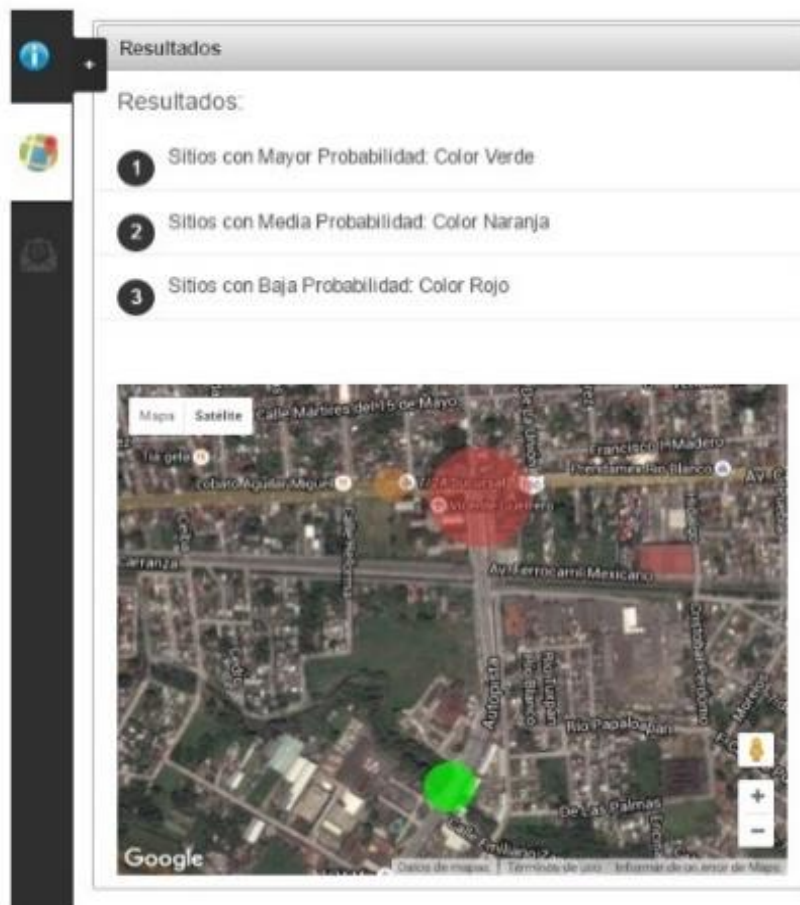


Figura 4. 4 Resultados Zonas Óptimas

La figura 4.4 presenta los resultados de sugerencia de posibles puntos de venta, en caso de que la ubicación si cumpla con las restricciones del reglamento; mostrando en forma circular de color verde los sitios más óptimos, en naranja los sitios con una probabilidad media y en rojo los sitios menos adecuados.



Figura 4. 5 Detalle de la zona.

Finalmente, la Figura 4.5 presenta el detalle de la forma circular seleccionada, donde un porcentaje nos indica que tan recomendable es la zona para establecer puntos de venta, así como también la identificación de establecimientos cercanos (hospitales, asilos, guarderías, orfanatos, escuelas, mercados, templos, auditorios, cines, estadios y teatros).

El sistema permite visualizar el detalle de cada una de las tres zonas (Alta, Media y Baja) encontradas mostrando los porcentajes y la puntualización de las reglas. Para este ejemplo se muestra el detalle de una zona con mayor posibilidad de ser adecuada para la colocación de un punto de venta; así como también la especificación de cumplir con las reglas de distancia de resguardo al no encontrarse con ningún establecimiento cercano. Otra de las funcionalidades es la búsqueda radial de puntos de venta, la cual identifica a partir de un punto (longitud y latitud) y una distancia radial ubicaciones apropiadas para establecer un punto de venta; donde muchas veces las superficies de estas búsquedas implican más de un asentamiento derivando en evaluaciones de restricciones para cada uno de los puntos que integren dicha superficie y en algunas ocasiones encuentra más de un punto de venta óptimo. Finalmente la búsqueda por código postal hace evaluaciones por todas las superficies de los asentamientos que pertenecen al código postal y al igual que la búsqueda radial en determinadas circunstancias devuelve más de un punto de venta adecuado. De manera que en casos con más de un punto de venta óptimo, la decisión depende del usuario. Hasta este momento el sistema no es capaz de solucionar el problema de arranque en frío, pero si es posible ofrecer geo-recomendaciones de puntos de venta a partir de los puntos de venta que se colocaron. Una de las restricciones del sistema, es el hecho de que el sistema no es capaz de solucionar el problema de arranque en frío, pero si es posible ofrecer geo recomendaciones de puntos de venta a partir de los puntos de venta que se colocaron e información con respecto al historial de ventas.

4.2. Pruebas del sistema

Los procesos de evaluación en los sistemas de recomendación ayudan a mejorar la diversidad y especificación de estos, pero sobre todo a enriquecer la experiencia de los usuarios [52]. Las evaluaciones de los sistemas de recomendación están clasificadas en las orientadas a los algoritmos de recomendación y en criterios/características propias de los sistemas de recomendación desde las perspectivas del usuario y del sistema. Las evaluaciones enfocadas en los algoritmos de recomendación representan la línea tradicional en el proceso de evaluación, ya que considera como punto central la exactitud de los algoritmos de predicción. Actualmente la exactitud no es suficiente para mejorar las necesidades de los usuarios [52,53]. Por otro lado las orientadas en las características de los sistemas de recomendación logran mejores resultados

en relacionar los ítems con los usuarios y la confianza del sistema en el proceso de predicción. Para el diseño de la evaluación, se decidió utilizar una orientada en las características de los sistemas de recomendación el fin de medir los aspectos principales del diseño GEOREMSYS. Por lo tanto, se presenta un diferente enfoque menos tradicional, con un fin de ofrecer un estilo de evaluación del sistema de recomendación.

Como primera fase es necesario identificar los criterios más relevantes de un sistema de geo-recomendación se emplea el proceso de análisis jerárquico (AHP), el cual se desarrolló por Thomas L. Saathy. Es un sistema flexible de metodología de análisis de decisión multi-criterio discreta [52]. Por lo tanto, este método posibilita que en el proceso de toma de decisiones se estructure un problema multi-criterio en forma visual, mediante la construcción de una jerarquía de atributos, la cual contiene como mínimo tres niveles:

- El propósito o el objetivo global del problema, situado en la parte superior.
- Los distintos atributos o criterios que definen las alternativas en el medio.
- Las alternativas que concurren en la parte inferior del diagrama.

En caso de que los atributos o los criterios no sean lo suficientemente explícitos o claros, es posible incluir sub-criterios más operativos en forma secuencial entre el nivel de criterios y el de las alternativas, lo que origina un modelo jerárquico multinivel. Una vez construido el modelo jerárquico, se realizan comparaciones por pares entre dichos elementos (criterios, sub-criterios y alternativas) y se atribuyen valores numéricos a las preferencias señaladas por las personas que intervienen en el proceso de decisión [53]. Este proceso es útil para evaluar las características de un producto y así, seleccionar las más relevantes. Para aplicar el proceso de análisis jerárquico, es comparada cada característica utilizando una escala de cinco puntos. La escala ordinal de comparación se mueve entre valores de 9 y 1. La escala adoptada por los expertos es:

- 1: igual importancia
- 3: algo más importante
- 5: bastante más importante
- 7: mucho más importante

- 9: absolutamente más importante

Los mismos valores se expresarán de forma fraccionaria si la comparación resulta inferior. Esta comparación por pares da lugar a matrices cuadradas en las cuales la correspondencia entre pares de valores es recíproca [54]. Para evaluar el sistema de geo-recomendación GEOREMSYS con otras herramientas, se utilizan algunos criterios encontrados en la literatura [11]. Las características representativas de los sistemas de recomendación se describen a continuación:

- **Supervisión del comportamiento (MH)**
- **Heurística para deducir información (HI)**
- **Retroalimentación por usuario (UF)**
- **Retroalimentación por ítem (IF)**
- **Historial de navegación (NH)**
- **Rastreador Web (CWP)**
- **Modelo (VM)**

Como primer paso, se crea una matriz comparación por pares con cada una de las características identificadas. Los valores sombreados de color naranja son las valoraciones que se le asignan con respecto a la característica activa (valor 1) en cada fila. El valor 1 es siempre asignado a la comparación de una alternativa con respecto a sí misma. Los resultados en color azul son los pertenecientes a la división de la característica activa entre la calificación de la característica. Finalmente se suman los resultados de cada una de las filas. La tabla 4.1 muestra el resultado:

Tabla 4. 1 Características de los sistemas de recomendación

	MH	HI	UF	IF	NH	CWP	VM	CF
MH	1,00	3,00	3,00	5,00	7,00	7,00	5,00	7,00
HI	0,33	1,00	5,00	5,00	7,00	9,00	9,00	9,00

UF	0,33	0,20	1,00	3,00	3,00	3,00	7,00	7,00
IF	0,20	0,20	0,33	1,00	5,00	7,00	3,00	5,00
NH	0,14	0,14	0,33	0,20	1,00	3,00	5,00	7,00
CWP	0,14	0,11	0,33	0,14	0,33	1,00	7,00	7,00
VM	0,20	0,11	0,14	0,33	0,20	0,14	1,00	9,00
CF	0,14	0,11	0,14	0,20	0,14	0,14	0,11	1,00
SUMA	2,50	4,88	10,29	14,88	23,68	30,29	37,11	52,00

Una vez que se comparó cada una de las características, los valores obtenidos se normalizaron. La Matriz normalizada se obtiene dividiendo cada número una columna de la matriz de comparación por pares por la suma total de la columna. La tabla 4.2 muestra el resultado:

Tabla 4. 2 Valores normalizados

	MH	HI	UF	IF	NH	CWP	VM	CF
MH	0,40	0,62	0,29	0,34	0,30	0,23	0,13	0,13
HI	0,13	0,21	0,49	0,34	0,30	0,30	0,24	0,17
UF	0,13	0,04	0,10	0,20	0,13	0,10	0,19	0,13
IF	0,08	0,04	0,03	0,07	0,21	0,23	0,08	0,10
NH	0,06	0,03	0,03	0,01	0,04	0,10	0,13	0,13
CWP	0,06	0,02	0,03	0,01	0,01	0,03	0,19	0,13
VM	0,08	0,02	0,01	0,02	0,01	0,00	0,03	0,17

CF	0,06	0,02	0,01	0,01	0,01	0,00	0,00	0,02
Total	1	1	1	1	1	1	1	1

A continuación se calcula el porcentaje de importancia relativa. El cual se obtiene calculado el promedio de cada fila de la matriz normalizada. Este promedio por fila representa el vector de prioridad de la alternativa con respecto a la seleccionada [55]. Los resultados se muestran en la tabla 4.3:

Tabla 4. 3 Características según grado de importancia

Característica	Porcentaje
MH	30%
CF	27%
UF	13%
NH	11%
VM	7%
HI	6%
CWP	4%
IF	2%

Utilizando como base los porcentajes de la tabla 4.3, se eligieron las características mejor calificadas, las cuales son presentadas en la tabla 4.4.

Tabla 4. 4 Características

Característica	Porcentaje
MH	30%

CF	27%
UF	13%
NH	11%
VM	7%

Con el fin de proporcionar una comparación de GEOREMSYS con otros sistemas se realizó una evaluación con dos sistemas, los cuales son: MyCityMX y tiendeo. Estos sistemas son basados en servicios de Geolocalización y cumplen con las características seleccionadas pertenecientes a la tabla 4.4. MyCityMX es una aplicación basada en geolocalización, diseñada para facilitar la vida del usuario concentrando puntos de interés como Turismo, Eventos, Hoteles, Bares y Restaurantes, así como todas las OFERTAS de las tiendas comerciales de tu ciudad. My CiTy también te dará información de tu ciudad como todas las oficinas de servicios públicos municipales, entre otros puntos de interés como Bancos, Hospitales, Transportes, entre otros [56]. Por otro lado TouristEye es una aplicación web y móvil (iPhone y Android™) que permite organizar por categorías lugares de atracción tales como restaurantes, hoteles, museos, centros de diversión, entre otros. Los Usuarios marcan los sitios turísticos como visitados, proporcionar un comentario para declarar un grado de satisfacción; así como también la descripción de visitas por medio de fotos, comentarios y valoraciones. Para visualizar los puntos de interés, el servicio de Google™ Maps es integrado dentro del sistema para mostrar sus rutas. El servicio de TouristEye integra un motor de recomendación tal que los nuevos puntos de interés automáticamente se muestran al usuario [57].

A continuación son descritas e ilustradas las características pertenecientes a la tabla 4.4.

Supervisión del comportamiento (MH)= El sistema observa el comportamiento de los usuarios y registros. La figura 4.6 muestra en el lado superior izquierdo como el sistema guarda el comportamiento de los usuarios con respecto a la calificación de las calles.

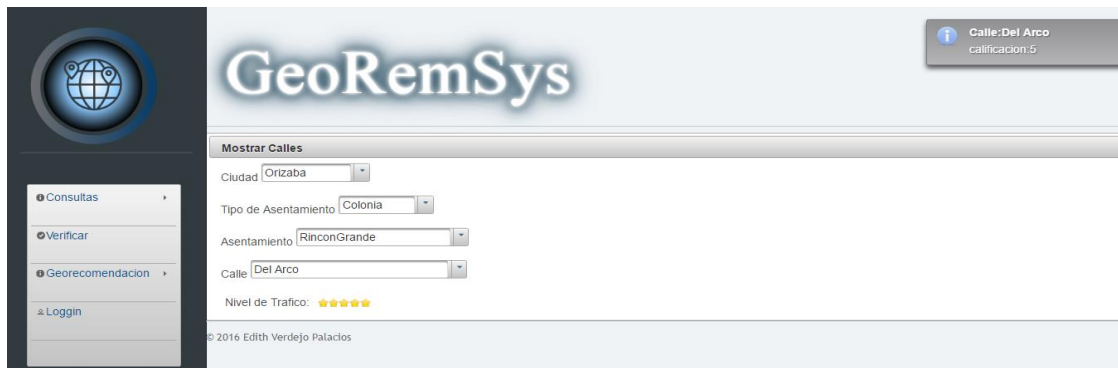


Figura 4. 6 Supervisión del comportamiento

Historial de navegación (NH)= El sistema registra el historial de navegación para ayudar a los usuarios. La figura 4.6 muestra cómo es posible registrar las calificaciones.

Retroalimentación por usuario (UF)= Es la calificación que dan los usuarios para ayudar a otros usuarios en la selección de artículos, para efectos de representación la figura 4.7 ilustra dicha característica.

Nivel de Trafico: ★★★★★

Figura 4. 7 Retroalimentación por usuario

Modelo (VM)= El sistema usa estructuras vectoriales (modelos) para almacenar el perfil de los usuarios. Dichas estructuras son mostradas en la figura 4.8.

```

}

public void generarRecomendaciones() {
    oReco = new Calle();
    int[] ids = null;
    int[] prods = null;
    Long num = new Long(0);
    try {
        DataModel model = new FileDataModel(
            new File("C:\\Users\\Edie\\Documents\\NetBeansProjects"
                + "\\GeoRemSis_V2\\dataset.csv"));
        PearsonCorrelationSimilarity sim =
            new PearsonCorrelationSimilarity(model);
        GenericItemBasedRecommender recommender =
            new GenericItemBasedRecommender(model, sim);
        int i = 0;
        ids = new int[(int) model.getNumItems() * 3];
        prods = new int[(int) model.getNumItems() * 3];
        for (LongPrimitiveIterator items = model.getItemIDs();
            items.hasNext(); ) {

```

Figura 4. 8 Modelo

Filtrado Colaborativo (CF)= Son funciones estadísticas para encontrar gente con perfiles similares.

El objetivo de la evaluación es demostrar que GEOREMSYS cumple con las características seleccionadas y proporcionar una manera para cuantificar el diseño propuesto. La comparación con otros sistemas de recomendación es sólo para ilustrar los beneficios de la GEOREMSYS.

	MYCITYMX	TouristEye	GEOREMSYS
MH	0	1	1
UF	1	1	1
NH	0	0	1
VM	1	1	1
CF	1	0	1

Tabla 4. Evaluación

En la tabla 4.5 se observa que GEOREMSYS y los sistemas comparados presentan algunas de las características seleccionadas. Se observa que algunos sistemas carecen de supervisión de comportamiento el cual se compensa con la retroalimentación que los usuarios dan a los ítems, la cual permite al sistema almacenar la información dentro de un modelo y con los algoritmos

calcular la recomendación. Así como también la existencia de colaboración entre los usuarios para generar información importante con respecto a los niveles de tráfico GEOREMSYS posee todas las características seleccionadas, además de ofrecer mejores recomendaciones solventadas con su capacidad de deducción y con el proceso de evaluación de evaluación de distancias radiales. Dichas características ayudan a GEOREMSYS a ofrecer una respuesta a los usuarios.

Capítulo 5. Conclusiones y recomendaciones

5.1 Conclusiones

En el proyecto de tesis se dio a conocer la necesidad de desarrollar un sistema que permitiera a una empresa la selección de puntos de venta por medio del desarrollo de un sistema de recomendación capaz de recomendar publicaciones. Los sistemas de recomendación son una herramienta en el proceso de selección de un ítem, los cuales permiten ofrecer selecciones más personalizadas a los clientes. Actualmente dichas herramientas están ganando notoriedad en diferentes aplicaciones, ya que permiten conocer mejor a los clientes. Hoy existe una gran tendencia en ofrecer recomendaciones de puntos de interés basadas tanto en el perfil de los usuarios como en la ubicación geográfica. Sin embargo los sistemas tradicionales de recomendación no consideran las características geográficas en la construcción de recomendaciones. Por otra parte se detecta que los sistemas de información geográfica son útiles en el análisis de datos geográficos, desafortunadamente estos sistemas por si solos no son capaces de extraer y analizar el conocimiento de las preferencias de los usuarios. De manera que la ubicación es un factor importante en la construcción de recomendaciones en mejorar la adaptabilidad de los sistemas de recomendación. El uso independiente de modelos geográficos y los sistemas de recomendación para el análisis de puntos de interés son líneas de investigación que están sujetas a limitaciones. Es decir, que la integración de ellos facilitaría en gran parte la creación de estrategias ayuda en la toma de decisión. Por lo que, la propuesta de la arquitectura de integración plantea la fusión de tecnologías que aún no se publican por otros estudios. Dicha propuesta en este capítulo reconoce que el principal objetivo de los sistemas de geo-recomendación es la consiente recomendación de un punto ubicación basada en los intereses individuales de los usuarios así como el historial. El objetivo de los sistemas de geo-recomendación merece ampliarse a partir de más investigaciones necesarias para la solución de este problema existente; ya que se aborda en muchos campos de interés y cambiar el panorama del desarrollo de las recomendaciones, derivando en recomendaciones más eficientes.

5.2 Recomendaciones

GEOREMSYS está orientado a las recomendaciones de puntos de venta de empresas dedicadas a la venta de gasolina y gas de un lugar en específico, por lo cual algunas recomendaciones son:

1. Desarrollar un módulo que permita anexar reglas pertenecientes a los Reglamentos para las acciones de construcción, instalación, conservación y operación de estaciones de servicio en gasolinera y carburación de todos los estados de la república.
2. Detectar los cambios en cuanto a la información territorial, es decir permitir la agregación de más ciudades y con ellos la oportuna búsqueda de sus reglamentos.
3. Incorporar un módulo de dinámica de sistemas que permita conocer el impacto de las posibles ubicaciones en el futuro por medio de simulaciones.

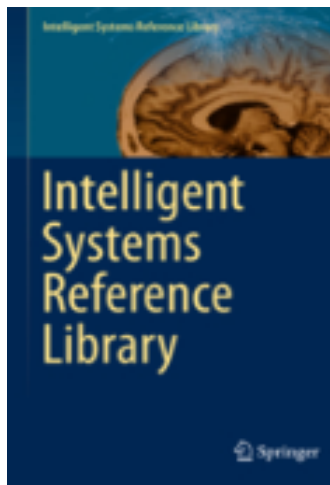
Productos Académicos



1) Edith Verdejo Palacios, Giner Alor Hernández, Cuauhtemoc Sánchez-Ramírez, Jose Luis Sánchez-Cervantes, Lisbeth Rodríguez-Mazahua. “Arquitectura de integración de un sistema de Geo-recomendación para establecer puntos de venta”. 1er Simposio Internacional de Optimización de Procesos Industriales, SIOPI 2015. ISBN: 978-0-692-56750-0.



2) Edith Verdejo Palacios, Giner Alor Hernández, Cuauhtemoc Sánchez-Ramírez, Jose Luis Sánchez-Cervantes, Lisbeth Rodríguez-Mazahua. “Arquitectura de integración para el desarrollo de un sistema de geo-recomendación para establecer puntos de venta”, Congreso Mexicano de Inteligencia Artificial (COMIA), 2016. ISSN:1870-4069.



3) Edith Verdejo Palacios, Giner Alor Hernández, Cuauhtemoc Sánchez-Ramírez, Jose Luis Sánchez-Cervantes, Lisbeth Rodríguez-Mazahua. “Geo-recommender system in industry”. Springer 2016. ISSN: 1868-4394

Bibliografía

- [1] Church, Richard L. "Geographical information systems and location science", *Computers & Operations Research*, vol. 29 no. 6, pp.541-562, 2002
- [2] Goodchild, Michael F. "Geographic information systems." *Journal of Retailing*, vol. 67, no. 1, pp. 3-15, 1991.
- [3] Cox, Allan B., and Fred Gifford. "An overview to geographic information systems." *The journal of academic librarianship*, vol. 23, no. 6, pp. 449-461, 1997.
- [4] Goodchild, Michael F. "Towards an enumeration and classification of GIS functions." *Proc. Int. GIS Symposium*. 1987.
- [5] Harvey, Francis. *Primer of GIS: Fundamental geographic and cartographic concepts*. Guilford Press, 2008.
Available:<http://resources.arcgis.com/es/help/gettingstarted/articles/026n00000014000000.html>
- [6] 0000.html
- [7] Robillard, Martin P., Robert J. Walker, and Thomas Zimmermann. "Recommendation systems for software engineering", *Software, IEEE*, vol. 27, no. 4, pp. 80-86, 2010.
- [8] Cai, Yi, et al. "Typicality-based collaborative filtering recommendation.", *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 3, pp. 766-779, 2014
- [9] Bobadilla, Jesús, et al. "Recommender systems survey", *Knowledge-Based Systems* vol. 46, pp.109-132, 2013
- [10] Carrer-Neto, Walter, et al. "Social knowledge-based recommender system. Application to the movies domain." *Expert Systems with Applications* vol. 39, no. 12, pp.: 10990-11000, 2012.

- [11] Middleton, S. E. Capturing knowledge of user preferences with recommender systems (Doctoral dissertation, university of Southampton), 2003.
- [12] E. Kumar, Artificial Intelligence. I.K. International Publishing House Pvt. Limited, 2008.
- [13] I. Management Association, Cyber Behavior: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications. Premier reference source, IGI Global, 2014.
- [14] M. Montaner, B. L'opez, and J. L. De La Rosa, "A taxonomy of recommender agents on the internet," Artificial intelligence review, vol. 19, no. 4, pp. 285–330, 2003.
- [15] Carrer-Neto, Walter, et al. "Social knowledge-based recommender system. Application to the movies domain." Expert Systems with Applications vol. 39, no. 12, pp.: 10990-11000, 2012.
- [16] Kantor, Paul B., et al. Recommender systems handbook. Springer, 2011.
- [17] Safoury, Laila, and Akram Salah. "Exploiting user demographic attributes for solving cold-start problem in recommender system." Lecture Notes on Software Engineering vol. 1, no. 3, pp.: 303-307, 2013.
- [18] Carrer-Neto, Walter, et al. "Social knowledge-based recommender system. Application to the movies domain." Expert Systems with Applications vol. 39, no. 12, pp.: 10990-11000, 2012.
- [19] Colombo-Mendoza, Luis Omar, et al. "RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes." Expert Systems with Applications vol. 42, no. 3, pp.: 1202-1222, 2015.
- [20] Noguera, José M., et al. "A mobile 3D-GIS hybrid recommender system for tourism" Information Sciences, vol. 215, pp.:37-52, 2012.

- [21] http://www.maori.geek.nz/list_of_recommender_systems/
- [22] Gantner, Zeno, et al. "MyMediaLite: a free recommender system library." Proceedings of the fifth ACM conference on Recommender systems. ACM, 2011.
- [23] <http://www.ibm.com/developerworks/library/os-recommender2/>
- [24] Garg, Deepak, and Khushbu Trivedi. "Fuzzy K-mean clustering in MapReduce on cloud based hadoop." Advanced Communication Control and Computing Technologies (ICACCCT), 2014 International Conference on. IEEE, 2014
- [25] Ingersoll, Grant. "Introducing apache mahout." Scalable, commercial friendly machine learning for building intelligent applications. IBM (2009)
- [26] Anil, Robin, Ted Dunning, and Ellen Friedman. Mahout in action. Shelter Island: Manning, 2011.
- [27] Gibin, M., Singleton, A., Milton, R., Mateos, P., & Longley, P. (2008). An exploratory cartographic visualisation of London through the Google Maps API. Applied Spatial Analysis and Policy, 1(2), 85-97.
- [28] Kanchev, G. M., & Chopra, A. K. (2015, August). Social media through the requirements lens: A case study of Google maps. In Crowd-Based Requirements Engineering (CrowdRE), 2015 IEEE 1st International Workshop on (pp. 7-12). IEEE.
- [29] Fu, W., & Nemesure, S. (2013). Accelerator control data visualization with Google Map (No. BNL--100683-2013-CP). Brookhaven National Laboratory (BNL).
- [30] https://es.wikipedia.org/wiki/Instituto_Nacional_de_Estad%C3%ADstica_y_Geograf%C3%ADa
- [31] www.inegi.org.mx/saladeprensa/boletines/2016/especiales/especiales2016_01_06.pdf

- [32] Batet, M., et al. "Turist@: Agent-based personalised recommendation of tourist activities". *Expert Systems with Applications*, vol 39, no 8, pp: 7319-7329, 2012.
- [33] Li, Yung-Ming, et al "A social recommender mechanism for location-based group commerce." *Information Sciences*, vol. 274, pp.: 125-142, 2014.
- [34] YU, YoungHoon, et al. "Recommendation system using location-based ontology on wireless internet: An example of collective intelligence by using 'mashup' applications". *Expert systems with applications*, vol. 36, no 9, pp.: 11675-11681, 2009.
- [35] Pliakos, Konstantinos, and Constantine Kotropoulos. "PLSA driven image annotation, classification, and tourism recommendation." *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014.
- [36] Yin, Huagang, et al. "Trip mining and recommendation from geo-tagged photos." *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*. IEEE, 2012
- [37] Fu, Cheng-Yao, et al. "Travelbuddy: interactive travel route recommendation with a visual scene interface." *International Conference on Multimedia Modeling*. Springer International Publishing, 2014.
- [38] Wei, Ling-Yin, et al. "Discovering Point-of-Interest Signatures Based on Group Features from Geo-social Networking Data." *2013 Conference on Technologies and Applications of Artificial Intelligence*. IEEE, 2013.
- [39] López-Ornelas, Erick, Rocío Abascal-Mena, and J. Sergio Zepeda-Hernández. "A Geo-collaborative Recommendation Tool to Help Urban Mobility." *International Conference on Human-Computer Interaction*. Springer International Publishing, 2014

- [40] Schedl, Markus, and Dominik Schnitzer. "Location-aware music artist recommendation." International Conference on Multimedia Modeling. Springer International Publishing, 2014.
- [41] Huang, Liwei, Yutao Ma, and Yanbo Liu. "Point-of-interest recommendation in location-based social networks with personalized geo-social influence." China Communications 12.12 (2015): 21-31.
- [42] Castro, Carlos Arturo et. al., "Modelación y simulación computacional usando sistemas de información geográfica con dinámica de sistemas aplicados a fenómenos epidemiológicos." Revista facultad de ingeniería Universidad de Antioquia, vol. 34, pp.: 86-100, 2005.
- [43] Corner, Richard A., et al. "A fully integrated GIS-based model of particulate waste distribution from marine fish-cage sites." Aquaculture, vol. 258, no. 1, pp.:299-311, 2006.
- [44] Vairavamoorthy, Kalanithy, et al. "IRA-WDS: A GIS-based risk analysis tool for water distribution systems." Environmental Modelling & Software, vol. 22, no. 7, pp.: 951-965, 2007.
- [45] Radiarta, I. Nyoman, et. al. "GIS-based multi-criteria evaluation models for identifying suitable sites for Japanese scallop (*Mizuhopecten yessoensis*) aquaculture in Funka Bay, southwestern Hokkaido, Japan." Aquaculture vol. 284, no. 1, pp.: 127-135, 2008.
- [46] Xu, Zhao, and Volker Coors. "Combining system dynamics model, GIS and 3D visualization in sustainability assessment of urban residential development." Building and Environment, vol. 47, pp.: 272-287, 2012.
- [47] Suárez-Vega, Rafael et. al., "Location models and GIS tools for retail site location." Applied Geography, vol. 35, no.1, pp. 12-22, 2012.

- [48] Roig-Tierno, Norat, et al. "The retail site location decision process using GIS and the analytical hierarchy process." *Applied Geography*, vol. 40, pp.: 191-198, 2013.
- [49] Casillas-Higuera, África, et al. "Detección de la Isla Urbana de Calor mediante Modelado Dinámico en Mexicali, BC, México." *Información tecnológica*, vol. 25, no.1, pp.: 139-150, 2014.
- [50] <https://goo.gl/GZMFwO>
- [51] Alag, S. "Collective intelligence in action", New York: Manning, pp. 34-41, 2009
- [52] Rodríguez, E. M. (2007). Aplicación del proceso jerárquico de análisis en la selección de la localización de una PYME. *Anuario jurídico y económico escorialense*, (40), 523-542.
- [53] Berumen, S. A., & Llamazares Redondo, F. (2007). La utilidad de los métodos de decisión multicriterio (como el AHP) en un entorno de competitividad creciente. *Cuadernos de administración*, 20(34), 65-87.
- [54] Álvarez Alonso, M., Arquero Hidalgo, A., & Martínez Izquierdo, E. (2008). Empleo del AHP (Proceso Analítico Jerárquico) incorporado en sig para definir el emplazamiento óptimo de equipamientos universitarios. Aplicación a un abiblioteca. Aplicación a una biblioteca.
- [55] <http://www.ccee.edu.uy/ensenian/catmetad/material/MdA-Scoring-AHP.pdf>
- [56] https://play.google.com/store/apps/details?id=com.mobincube.my_city.sc_3Z694G&hl=es
- [57] Umanets, A., Ferreira, A., & Leite, N. (2014). GuideMe—A tourist guide with a recommender system and social interaction. *Procedia Technology*, 17, 407-414.

