



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

LABORATORIO DE PROCESAMIENTO INTELIGENTE DE INFORMACIÓN GEO-ESPACIAL

**Sistema de recomendación basado en realidad aumentada
para el análisis de datos urbanos**

TESIS

QUE PARA OBTENER EL GRADO DE:

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A:

Lic. Fernando Jesús González Alonso

DIRECTORES DE TESIS:

DR. MIGUEL JESÚS TORRES RUIZ

DR. JOSÉ GIOVANNI GUZMÁN LUGO

CIUDAD DE MÉXICO

DICIEMBRE 2016



INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 10:00 horas del día 01 del mes de diciembre de 2016 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

“Sistema de recomendación basado en realidad aumentada para el análisis de datos urbanos”

Presentada por el alumno:

GONZÁLEZ

Apellido paterno

ALONSO

Apellido materno

FERNANDO JESÚS

Nombre(s)

Con registro:

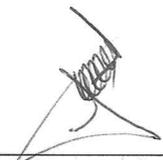
B	1	4	0	4	5	7
----------	----------	----------	----------	----------	----------	----------

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

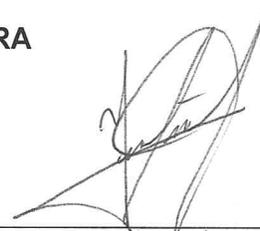
Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis



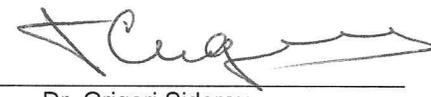
Dr. José Giovanni Guzmán Lugo



Dr. Miguel Jesús Torres Ruiz



Dr. Sergio Suárez Guerra



Dr. Grigori Sidorov



Dr. Francisco Hiram Calvo Castro



Dr. Marco Antonio Moreno Ibarra

PRESIDENTE DEL COLEGIO DE PROFESORES



Dr. Marco Antonio Ramírez Salinas



INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN
EN COMPUTACIÓN
DIRECCIÓN



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 1 del mes de diciembre del año 2016, el (la) que suscribe González Alonso Fernando Jesús alumno (a) del Programa de Maestría en Ciencias de la Computación con número de registro B140457, adscrito a Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. José Giovanni Guzmán Lugo y Dr. Miguel Jesús Torres Ruiz y cede los derechos del trabajo intitulado Sistema de recomendación basado en realidad aumentada para el análisis de datos urbanos, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

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

Fernando Jesús González Alonso
Nombre y firma

Resumen

Los sistemas recomendadores, han demostrado ser de gran utilidad en diversas áreas de investigación, estos sistemas ayudan a la toma de decisiones a los usuarios brindándoles recomendaciones de los artículos o servicios de su interés por medio de las preferencias del mismo; no obstante, frecuentemente cuando se realiza la búsqueda y esta no incluye exactamente las preferencias del usuario, la consulta devuelve resultados vacíos.

Planteemos un ejemplo, una persona desea ver una película y se encuentra en un punto de la ciudad que no sea de su completo conocimiento o desconozca los establecimientos en este caso salas de cine cercanos al usuario. El primer inconveniente con el que se encontraría es que no sabe la localización exacta de la sala o las salas cinematográficas cercanas a su posición actual y posteriormente, en una cartelera la cual puede ser extensa elegir la película que puede ser de su preferencia.

Para resolver esta problemática, se propone una metodología enfocada a la recuperación semántica de información con base en un perfil de usuario y representada con ayuda de la realidad aumentada la cual, por medio de una base de datos espacial podrá recuperar los puntos de interés cercanos mostrándolos al usuario en una aplicación móvil, con base en la coordenada de la posición del dispositivo móvil donde se realiza la consulta y las coordenadas de las salas cinematográficas a las que sea desea llegar. Por otra parte se utilizan las preferencias del usuario para la recomendación de la película.

La metodología del trabajo se divide principalmente en tres etapas: Personalización, Procesamiento semántico y Visualización. La Personalización, consiste en la selección de las preferencias del usuario y su almacenamiento, la cual se utiliza en el filtro de la etapa de Procesamiento semántico junto

con la ontología, y por último, la etapa de Visualización, en la que se despliegan los datos por diferentes medios, esta última pudiéndose auxiliar de la realidad aumentada para su representación.

Como caso de estudio se utiliza la Ciudad de México y las salas cinematográficas que se encuentran dentro de ella. El sistema propuesto, muestra los puntos de interés cercanos dentro de un radio determinado, dependiendo del punto de interés que el usuario seleccione se brinda una recomendación en el caso de estudio una recomendación de las películas que el usuario tenga como preferencia, mostrando primero las que más pudieran ser de su agrado y por último las que sean de su menor preferencia.

Abstract

The recommended systems had proven to be very useful in several areas of research; these systems help the user to make decisions by offering recommendations of the items or services of their interest according to their preferences; However, frequently when the research is performed and it does not include exact user profile preferences, the query gives empty results.

For example, a person wants to see a movie and is somewhere in the city that the establishments near the user are unknown for him in this case, cinemas. The first inconvenience that could be found is that he does not know the exact location of the cinema or cinemas near his current position and subsequently, in a cinema listings which could be extensive he will be able to choose the movie that may be of his preference.

To solve this problem, we propose a methodology focused on the semantic retrieval of information based on a user profile and represented with the help of augmented reality which, through a spatial database can retrieve nearby points of interest by showing them to the user in a mobile application, based on the geographical coordinate of the mobile device where the query is made and the location of the cinematographic rooms which it wishes to arrive. On the other hand the user's preferences are used to recommend a movie.

The methodology of the work is divided mainly in three stages: Personalization, Semantic Processing and Visualization. The Personalization consists in the selection of the user's preferences and his storage, which is used as a filter on the Semantic Processing stage along with the ontology, and finally, the Visualization stage, in which the datas are displayed by different means, also the augmented reality is able to aid for its representation.

As a case of study, Mexico City is used as a general location map and the

cinematographic rooms that are within it. The proposed system shows the nearby points of interest in a given radius, depending on the point of interest that the user selects, as a recommendation is given in the case of study, a list of the movies that the user has as a preference, showing first the most liked.

Agradecimientos

A Dios por todo lo que me ha dado y por dejarme creer.

A mis padres que siempre me han apoyado, por su cariño y compañía.

A mis compañeros de laboratorio y a los amigos que conseguí en el CIC; a mis nuevos amigos en Rusia por cambiar mi percepción del mundo.

A mis asesores de tesis el Dr. Miguel Jesús Torres y el Dr. José Giovanni Guzmán Lugo, al Dr. Grigori Sidorov que me dio la oportunidad de conocer Rusia con mi trabajo de tesis, al Dr. Anton Zimmerling por recibirme en Moscú y ayudarme con mi investigación.

Al instituto Politécnico Nacional por todo lo que me ha permitido lograr, al Centro de Investigación en Computación y al Consejo Nacional de Ciencia y Tecnología por el apoyo en la realización de este trabajo.

¡Gracias!

Índice general

1. Introducción	17
1.1. Antecedentes	17
1.1.1. Sistemas Recomendadores	17
1.2. Planteamiento del problema	18
1.3. Objetivos	18
1.3.1. Objetivo general	18
1.3.2. Objetivos específicos	18
1.4. Justificación	19
1.5. Beneficios esperados	19
1.6. Alcances y limitaciones	19
1.7. Organización de la tesis	19
2. Estado del Arte	21
2.1. Sistema de recomendación basado en películas	21
2.2. Ontologías	22
2.2.1. Sistemas conducidos por ontologías	23
2.2.2. Ontologías en dispositivos móviles	24
2.3. Perfil de usuario	27
2.3.1. Etiquetado de colaboración y perfil de usuario basado en etiquetas	27
2.3.2. Búsquedas personalizadas	28
2.3.3. Búsqueda personalizada en sistemas de etiquetado co- laborativo	29
2.3.4. El usuario y recursos de perfiles de etiquetado colabo- rativo	29
2.3.4.1. Modelado de usuario	29

3. Marco Teórico	31
3.1. Ontologías	31
3.1.1. Construcción de una ontología	32
3.1.1.1. Captura	32
3.1.1.2. Codificación	32
3.1.1.3. Integración de las ontologías existentes	32
3.2. WordNet	33
3.3. NLTK (Natural Language Toolkit)	33
3.3.1. Uso de los Synsets con NLTK	34
3.4. Realidad aumentada	36
3.4.1. Técnicas de visualización para la Realidad Aumentada	49
3.4.2. Integración de los datos	52
3.4.2.1. Percepción de profundidad	52
3.4.2.2. Aumento en las señales de profundidad pic- tóricas	54
3.4.2.3. Manejo de la oclusión	55
3.4.2.4. Características de los bordes	56
3.4.2.5. Características sobresalientes	58
3.5. Appunta	62
3.5.1. Appunta estructura gráfica general	62
3.6. Sistema de recomendación	63
3.6.1. Filtrado por contexto	64
3.6.2. Sistema de recomendación demográfico	66
3.6.3. Sistema de recomendación basado en conocimiento	68
3.6.4. Sistema de recomendación colaborativo	70
3.6.5. Sistemas de recomendación híbridos	73
3.7. PHP	75
3.7.1. Laravel	75
3.8. Android	75
3.8.1. Android - Lollipop	76
3.8.1.1. Material Design	76
3.8.1.2. Multiscreen	77
3.8.1.3. Notificaciones	77
3.8.1.4. Android 5.0 Cambios del comportamiento (API Level 21)	77
3.9. Visualización de los datos en web	77
3.9.1. API de Google Maps	79
3.9.2. OpenLayers	80

<i>ÍNDICE GENERAL</i>	11
3.9.3. Leaflet	80
3.10. PostgreSQL	81
3.10.1. PostGIS	82
4. Metodología	83
4.1. Introducción	83
4.1.1. Etapa de Personalización	84
4.1.1.1. Perfil de usuario	85
4.1.1.2. Vector Característico	86
4.1.2. Etapa de Procesamiento Semántico	87
4.1.2.1. Ontología	87
4.1.2.2. WordNet	90
4.1.2.3. Filtros	92
4.1.2.4. Recuperación de puntos de interés	96
4.1.3. Etapa de Visualización	97
4.1.3.1. Web GIS	97
4.1.3.2. Mobile GIS	98
4.1.3.3. Sistemas operativos móviles	99
4.1.3.4. Realidad aumentada	100
4.1.3.5. Visualización del resultado en una consulta espacial	102
5. Resultados Experimentales	104
5.1. Etapa de Personalización	104
5.1.1. Perfil de usuario	104
5.1.2. Vector característico	108
5.2. Etapa de Procesamiento Semántico	108
5.2.1. Ontología	108
5.2.2. Filtrado	132
5.3. Visualización	137
6. Conclusiones y Trabajos Futuros	143
6.1. Conclusiones	143
6.2. Aportaciones científicas	144
6.3. Limitaciones	145
6.4. Trabajos a futuro	145

Índice de figuras

2.1.	Estructura de la ontología.	25
2.2.	Rango de búsqueda entre una tienda y un usuario.	26
2.3.	Ejemplo de horario de películas y tiempo disponible del usuario.	27
3.1.	Patrones de disparo en software de RA.	38
3.2.	Proceso de realidad aumentada con geolocalización. Fig 1 . . .	39
3.3.	Proceso de realidad aumentada con geolocalización. Fig 2 . . .	40
3.4.	Proceso de realidad aumentada con geolocalización. Fig 3 . . .	41
3.5.	Proceso de realidad aumentada con geolocalización. Fig 4 . . .	42
3.6.	Proceso de realidad aumentada con geolocalización. Fig 5 . . .	43
3.7.	Proceso de realidad aumentada con geolocalización. Fig 6 . . .	44
3.8.	Proceso de realidad aumentada con geolocalización. Fig 7 . . .	45
3.9.	Proceso de realidad aumentada con geolocalización. Fig 8 . . .	46
3.10.	Proceso de realidad aumentada con geolocalización. Fig 9 . . .	47
3.11.	Proceso de realidad aumentada con geolocalización. Fig 10 . . .	48
3.12.	El flujo de datos en un sistema de realidad aumentada común. Un mundo real imaginario es entregado por la alimentación de video del sistema y procesada por algoritmos de seguimiento basados en visión. Para alinear los datos reales y virtuales, los datos de seguimiento derivados se les ha aplicado una trans- formación del contenido virtual. Por último, la prestación se superpone en la parte superior de la secuencia de video.	50

3.13. Aumento indebido. En este ejemplo, los aumentos por descuido de estructuras ocultas causan dos problemas fundamentales: se anula información útil (como puntos de referencia) y carecen de señales de profundidad. (Derecha) escena original antes del aumento, la cruz de color negro indica el punto de inserción de la aguja RFA. (Izquierda) Aumento del hígado con el portal y los árboles de los vasos hepáticos y un tumor. 51

3.14. La interacción visual entre las representaciones reales y virtuales. Las visualizaciones en (a) y (b) utilizan los mismos parámetros. Sin embargo, mientras que la visualización en (a) presenta claramente todos los objetos importantes, los elementos aumentados apenas son visibles en (b). 52

3.15. La sincronización del parámetro de lo virtual y lo real permite a la cámara alinear señales de profundidad pictóricas reales y virtuales. La figura Lego virtual en (a) se percibe correctamente junto a los datos reales, mientras que el virtual en (b) se percibe correctamente detrás de ambos. Este efecto se consigue por señales de profundidad de alineación, tales como la distorsión de la perspectiva y el tamaño relativo. 55

3.16. La importancia de las señales de oclusión (a) A pesar de que existe un número de señales de profundidad diferentes, su orden de profundidad es ambigua y la percepción está mal si las oclusiones han sido ignoradas (b) El mismo renderizado que presentado en (a) con oclusión es resuelto correctamente. Esta visualización es capaz de comunicar la relación espacial entre su contenido real y virtual. 56

3.17. Imagen secundaria basada en imágenes usando la preservación de bordes. La preservación de bordes en toda imagen estorba en su presentación. Una lente plana mágica interactiva permite controlar la saturación en el efecto fantasma del video. Los bordes se han conservado por los operadores en el espacio imagen 2D solamente. Cada trama del video se procesa y los bordes discretos se detectan usando un detector de borde ordinario. 57

3.18. El sistema de realidad aumentada utiliza una reconstrucción de textura trabadas adquiridas de una escena a distancia (centro) a partir de imágenes capturadas y los modelos derivados. La reconstrucción se superpone con textura de vista del usuario del medio ambiente (a la izquierda). En combinación con el resaltado de los bordes de la oclusión de la estructura, únicamente cuando los elementos ocultos pueden ser revelados, se consigue una visualización por rayos x comprensible, mientras que al mismo tiempo se evita el desorden de borde (derecha).	58
3.19. Composición: Las imágenes de origen se procesan a través de una serie de filtros y combinaciones para producir la imagen de salida final. \oplus , \ominus y \otimes denotan la suma, resta y multiplicación de los valores de los píxeles.	61
3.20. Gráfica que muestra el funcionamiento de Appunta.	62
3.21. Palabras representadas en el campo de la investigación de los sistemas de recomendación. Distancias cortas indican similitudes más altas y un color cálido indica una mayor fiabilidad. El tamaño de los nodos es proporcional a la importancia de las palabras.	64
3.22. Componentes generales del proceso tradicional de recomendación.	65
3.23. Paradigmas para la incorporación de contexto en los sistemas de recomendación.	66
3.24. Construcción del conocimiento	69
3.25. Material Design, Multiscreen, Notifications.	76
3.26. Comparación de las técnicas de visualización.	79
3.27. Gráfico que ilustra de manera general los componentes más importantes en un sistema PostgreSQL.	82
4.1. Metodología	84
4.2. Ilustrando el concepto de una matriz léxica	91
4.3. Ejemplo del camino más corto en la jerarquía ontológica.	94
4.4. Ejemplo del ancestro común en la jerarquía.	94
4.5. Arquitectura de Android.	100
4.6. Arquitectura de Android.	102
5.1. Géneros en la aplicación.	106
5.2. Selección de géneros en la aplicación.	107

5.3. Comparativo de similitudes de action contra los demás géneros	122
5.4. Comparativo de similitudes de adventure contra los demás géneros	122
5.5. Comparativo de similitudes de animation contra los demás géneros	123
5.6. Comparativo de similitudes de biography contra los demás géneros	123
5.7. Comparativo de similitudes de comedy contra los demás géneros	124
5.8. Comparativo de similitudes de crime contra los demás géneros	124
5.9. Comparativo de similitudes de documentary contra los demás géneros	125
5.10. Comparativo de similitudes de drama contra los demás géneros	125
5.11. Comparativo de similitudes de family contra los demás géneros	126
5.12. Comparativo de similitudes de fantasy contra los demás géneros	126
5.13. Comparativo de similitudes de history contra los demás géneros	127
5.14. Comparativo de similitudes de horror contra los demás géneros	127
5.15. Comparativo de similitudes de music contra los demás géneros	128
5.16. Comparativo de similitudes de musical contra los demás géneros	128
5.17. Comparativo de similitudes de mystery contra los demás géneros	129
5.18. Comparativo de similitudes de romance contra los demás géneros	129
5.19. Comparativo de similitudes de fiction (sci-fi) contra los demás géneros	130
5.20. Comparativo de similitudes de sport contra los demás géneros	130
5.21. Comparativo de similitudes de thriller contra los demás géneros	131
5.22. Comparativo de similitudes de war contra los demás géneros .	131
5.23. Comparativo de similitudes de western contra los demás géneros	132
5.24. Diagrama de la base de datos de películas y géneros	137
5.25. Resultado de la ubicación 1.	139
5.26. Resultado de la ubicación 2.	139
5.27. Resultado de la ubicación 1 con una rotación diferente de la cámara mostrando una sola coincidencia.	140
5.28. Vista de las películas con su género y similitud.	142

Índice de tablas

1.1. Espacio de diseño técnico	18
2.1. Sistemas de recomendación basados en bases de datos públicas.	22
2.2. Reglas de asociación para la información de ubicación	26
2.3. Reglas de asociación para la información de horarios	27
3.1. Técnicas de Recomendación	74
3.2. Características de PostgreSQL	81
4.1. Matriz de similitudes de C	95
4.2. Ejemplo de representación de datos de puntos de interes	97
4.3. Resultados de la consulta	103
5.1. Tabla de similitudes lch	109
6.1. Tabla de similitudes path	155
6.2. Tabla de similitudes Wup	168

Capítulo 1

Introducción

1.1. Antecedentes

Al día de hoy en México no se cuenta con alguna herramienta que le permita a los usuarios hacer una recomendación para el consumo de un producto o servicio (para nuestro caso de estudio, serán películas de cine en exhibición) con base en un perfil de usuario combinado con alguna métrica de similitud semántica.

1.1.1. Sistemas Recomendadores

Resnick y Varian en su trabajo Resnick & Varian (1997) propusieron un método para la toma de decisiones sin contar con experiencia, obteniendo información a partir de terceros para realizar recomendaciones, esto de un modo de persona a persona. Estas recomendaciones abarcan varios campos como el *e-commerce*, así como guías de restaurantes y atracciones turísticas

Los sistemas de recomendación asisten y proporcionan mayor información para la toma de decisiones para el usuario, aunque existen ciertos casos donde las recomendaciones no son cercanas a las solicitadas por el usuario.

En la Tabla 1.1.1 se muestra la dimensionalidad de un sistema recomendador con base en su diseño técnico.

Tabla 1.1: Espacio de diseño técnico

	Contenido de la recomendación	Entrada explícita?	Anónimo?	Agregación	Uso de las recomendaciones
GroupLens	a) numérico: 1-5 b) segundos	a) explícito b) monitorear el tiempo de lectura	pseudónimo	ponderación personalizada basada en un acuerdo pasado entre los recomendadores	visualización junto con artículos en vistas de resumen existentes
Fab	numérico: 1-7	explícito	pseudónimo	ponderación personalizada; combinado con el análisis de contenido	selección filtrado
ReferralWeb	la mención de una persona o un documento	extraído de las fuentes de datos públicas	atribuida	montar la cadena de referencia a la persona deseada	visualización
PHOAKS	la mención de una URL	extraídos de publicaciones de Usenet	atribuida	una persona un voto (por URL)	visualización ordenada
Siteseer	la mención de una URL	extraído de carpetas de marcadores existentes	anónimo	frecuencia de mención en carpetas superpuestas	visualización

1.2. Planteamiento del problema

Hoy en día, al buscar una cartelera cinematográfica no hay una opción que permita clasificar las películas acorde al género de película. Lo anterior es relevante, porque el interés de cada persona está descrito por una serie de preferencias de cada individuo.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar una metodología enfocada a la recuperación semántica de información, con base en un perfil de usuario y tomando como base, información del usuario, con la finalidad de generar recomendaciones precisas y siendo éstas presentadas mediante realidad aumentada.

1.3.2. Objetivos específicos

- Implementar una ontología de aplicación para el dominio cinematográfico.
- Implementar un método de realidad aumentada para visualizar las salas cinematográficas más cercanas.
- Diseñar e implementar un método para procesar el perfil de usuario mediante consultas semánticas sobre la ontología.

- Implementar un método de similitud semántica para recuperar la información más cercana al perfil de usuario.

1.4. Justificación

Este trabajo brinda una opción al fortalecimiento del área de *e-commerce*, ofreciendo al usuario una herramienta de cómputo que permite elegir una película de una cartelera a partir del análisis de sus preferencias. Asimismo, este sistema admite la difusión de películas no tan taquilleras que otros usuarios puedan preferir y que son considerados de interés, más allá de las películas comerciales.

1.5. Beneficios esperados

Ayudar mediante el uso de la realidad aumentada al sistema de recomendación para que el usuario de una forma más interactiva y fácil pueda tener acceso a sus preferencias, ya que al utilizarse esta técnica combinada con operaciones de geoprocetamiento y dispositivos móviles se facilita la interacción con los resultados y mejoran los mismos.

1.6. Alcances y limitaciones

La metodología es capaz de realizar operaciones de espaciales, utilizando en conjunto la ontología construida. Además, se propone el desarrollo de un perfil de usuario para capturar las preferencias y ponderaciones de los mismos.

1.7. Organización de la tesis

- **Capítulo 2.** Este capítulo se centra en el estado del arte, está compuesto por trabajos relacionados en un sentido científico y tecnológico, acerca de sistemas de recomendación, medidas de similitud semántica y ontologías.

- **Capítulo 3.** Este capítulo denominado marco teórico muestra los conceptos y herramientas que utilizaremos para poder resolver el tema de la tesis.
- **Capítulo 4.** Este capítulo presenta la metodología con la cual se resuelve un caso de estudio definido en esta tesis.
- **Capítulo 5.** Este capítulo presenta los experimentos realizados y los resultados obtenidos, los cuales sirven para realizar un análisis de la metodología propuesta.
- **Capítulo 6.** En este capítulo se presentan las principales contribuciones y conclusiones obtenidas en el proceso de investigación desarrollado, así como algunas propuestas relacionadas con posibles trabajos futuros.

Capítulo 2

Estado del Arte

En este capítulo se describen algunos de los trabajos relacionados con sistemas de recomendación, así como su uso en dispositivos móviles. Se considera un análisis enfocado hacia trabajos relacionados con recuperación semántica de información y fundamentos sobre el perfil de usuario, mejorando la visualización del resultado con ayuda de la realidad aumentada.

2.1. Sistema de recomendación basado en películas

En el trabajo de Carrer-Neto et al. Carrer-Neto et al. (2012) se menciona que con la llegada de la web social y la creciente popularidad de aplicaciones Web 2.0, éstas cobran relevancia y dan un impulso a los sistemas de recomendación. Las recomendaciones que se generan por estos sistemas recomendadores tienen como objetivo proporcionar a los usuarios finales sugerencias acerca de los elementos sociales, elementos de información, productos o servicios que pueden ser de su interés. También se menciona que los sistemas de recomendación basados en técnicas sintácticas tradicionales sufren una serie de deficiencias en las cuales se obstaculiza su eficacia.

Así, conforme las tecnologías semánticas maduran, éstas proporcionan una base coherente y fiable para hacer frente a los datos a nivel de conocimiento, adicionando técnicas semánticamente facultadas para los sistemas de recomendación, con lo cual se pueden mejorar significativamente la calidad general de las recomendaciones.

Tabla 2.1: Sistemas de recomendación basados en bases de datos públicas.

	Sin información social						Información social de GroupLens		
	Movie Lens 1 M	Movie Lens 10 M	Netflix	Jester	Each Movie	Book-Crossing	ML	Last.Fm	Delicious
Puntuaciones	1 millón	10 millones	100 millones	4.1 millones	2.8 millones	1.1 millones	855,598	92,834	104,833
Usuarios	6040	71,567	480,189	73,421	72,916	278,858	2113	1892	1867
Artículos	3592	10,681	17,770	100	1628	271,379	10,153	17,632	69,226
Alcance	{1,...,5}	{1,...,5}	{1,...,5}	-10,10	[0,1]	{1,...,10}	{1,...,5}	Implícito	Implícito
Etiquetas	N/A	N/A	N/A	N/A	N/A	N/A	13222	11946	53388
Etiquetas de asignación	N/A	N/A	N/A	N/A	N/A	N/A	47957	186479	437593
Relaciones de amigos	N/A	N/A	N/A	N/A	N/A	N/A	N/A	25434	15328
Artículos	Películas	Películas	Películas	Chistes	Películas	Libros	Películas	Música	URL's

Christakou y Stafylopatis en Christakou & Stafylopatis (2005) presentan las especulaciones entre los miembros de la comunidad de Inteligencia Artificial en relación de cómo ésta puede ayudar con el problema de la búsqueda de información con éxito. Los sistemas de recomendación proporcionan una solución a este problema, dando recomendaciones individualizadas, las cuales se basan en el contenido y filtrado colaborativo; éstas se aplican generalmente para predecir estas recomendaciones.

La combinación de los resultados de estas dos técnicas fue utilizada en el trabajo, con la finalidad de construir un sistema que proporciona de forma más precisa recomendaciones relativas a las películas.

Bobadilla et al. Bobadilla et al. (2013) señala que la investigación en sistemas de recomendación requiere el uso de un conjunto representativo de bases de datos públicas para facilitar las investigaciones sobre las técnicas, métodos y algoritmos desarrollados por los investigadores en el campo. Por medio de estas bases de datos, la comunidad científica puede replicar experimentos para validar y mejorar sus técnicas. La Tabla 2.1 enumera lo actual con respecto a bases de datos públicas que son citadas más a menudo en la literatura. *Last.Fm* y *Delicious* incorporan calificaciones implícitas e información social; sus datos se generan a partir de las versiones publicadas en los conjuntos de datos de HetRec, 2.011, hospedada por el grupo de investigación GroupLens.

2.2. Ontologías

Existen diversas metodologías para el desarrollo e implementación de ontologías, las cuales se fundamentan en la ingeniería ontológica. Entre las metodologías más utilizadas se tienen: la metodología de Uschold y King Uschold & King (1995), la de Grüninger y Fox Grüninger & Fox (1995),

METHONTOLOGY Fernández-López et al. (1997); López et al. (1999), y On-To-Knowledge Staab et al. (2001). Con respecto a las herramientas de desarrollo de ontologías, se pueden mencionar, entre otras, Protégé Noy et al. (2000), OntoEdit Sure et al. (2002), y KAON Maedche et al. (2003).

2.2.1. Sistemas conducidos por ontologías

El trabajo de Bogdanović et al. Bogdanović et al. (2015) describe que debido a los avances en las tecnologías de información, comunicación y sensores, somos testigos de un constante aumento en la disponibilidad de la información geoespacial. Dado que ésta se genera y está a disposición a través de Internet por diferentes grupos de interés, la heterogeneidad de los datos geoespaciales y fuentes de datos geoespaciales se hace inevitable en ese sentido. Esta heterogeneidad de las fuentes geoespaciales disponibles a nivel mundial presentan un gran reto para los individuos que están tratando de reunir y descubrir información geoespacial de las fuentes de datos geoespaciales distribuidas.

Desde la aparición de las infraestructuras (nacionales) de datos espaciales ((N)SDI) Aditya & Kraak (2005) por sus siglas en inglés, el problema de descubrir y acceder a los datos geoespaciales montados a partir de fuentes de datos geoespaciales heterogéneos y distribuidos gana importancia. El objetivo principal del geoportal es bastante simple: entre todos los datos referenciados (servicios, capas geoespaciales, documentos, etc.), los usuarios que utilizan geoportal necesitan un mecanismo para encontrar (descubrir) fácilmente lo que están buscando utilizando sus propias palabras Tellez-Arenas (2009); Tait (2005).

Recientes reportes indican problemas de usabilidad desde el punto de vista de las personas que no pertenecen al sistema de información geográfica (GIS por sus siglas en inglés); es decir, “no profesionales en GIS” Tait (2005).

El trabajo de Bogdanović et al. Bogdanović et al. (2015) menciona que con el propósito de mejorar el descubrimiento del origen de datos geoespaciales, proponiendo una metodología utilizada en una arquitectura de integración de geoinformación impulsada por ontologías, que implementan la recuperación de información geográfica, a través de conexiones (*mappings*) entre ontologías y las fuentes de información geográfica. El núcleo de la metodología es un proceso de búsqueda de términos extraídos de una descripción en lenguaje

natural de geoinformación, definido por los usuarios finales, con el sentido de los conceptos de ontología (dominio/local).

Los sistemas de información geográfica contemporáneos se basan en fuentes de datos de información geográfica distribuidas para proporcionar a los usuarios un único punto de acceso uniforme sobre los datos refinados, mientras que el punto de acceso único se implementa habitualmente en la forma de un geoportal Tellez-Arenas (2009).

Mientras se realiza la búsqueda de los datos que les interesa, los usuarios del GIS esperan ser provistos con los resultados de búsqueda en forma de conjuntos de datos homogéneos despojados de detalles relacionados con el origen de los datos. Esto crea la necesidad de llevar a cabo la integración de datos y recursos de cómputo que pertenecen a varios sistemas autónomos Calvanese & De Giacomo (2005), lo que conduce al desarrollo de un sistema de geoinformación asociado Sayar (2009). Para poder integrar los datos procedentes de fuentes de datos geoespaciales heterogéneas y distribuidas, los GIS asociados necesitan superar el problema de la heterogeneidad semántica entre diferentes fuentes de datos geoespaciales Hakimpour (2003); Buccella et al. (2009). Dentro de las diferentes arquitecturas de integración de sistemas de geoinformación asociada, los problemas de heterogeneidad semántica se resuelven con mayor frecuencia a través de medios de ontologías Gruber (1993); Buccella et al. (2009).

2.2.2. Ontologías en dispositivos móviles

En el trabajo de Yu et al. Yu et al. (2009) se describe que un servicio de recomendación basado en la ubicación, necesita lugar y hora, adicionalmente en comparación a los servicios de recomendación en general. En la Figura 2.1, se define una ontología, en donde el centro y el lado izquierdo representan la ontología en la información de la ubicación de un usuario y una tienda y el lado derecho representa la ontología basada en la información referente al tiempo. La parte inferior de la figura representa una ontología de otros atributos de la tienda.

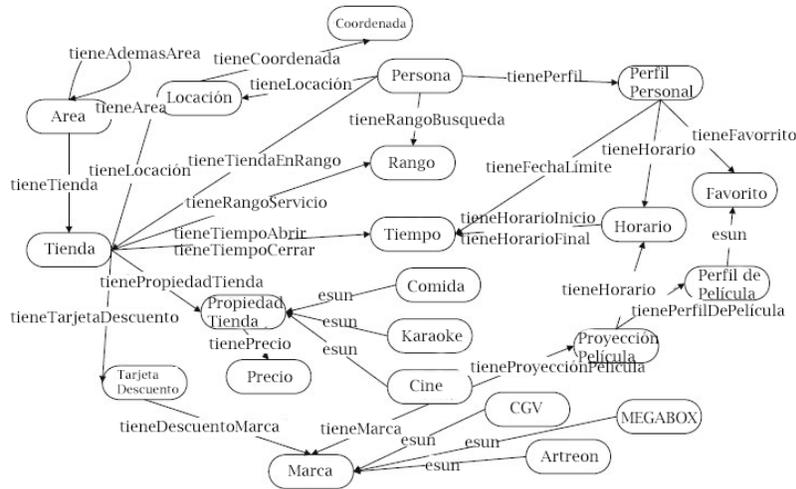


Figura 2.1: Estructura de la ontología.

De acuerdo con en el ejemplo anterior, la clase “*Tienda*” tiene en sus atributos el *horariodelnegocio* y su *ubicación* y como los atributos secundarios, los cuales pueden ser alimentos, cines, así como los precios.

Mediante el uso de esta ontología, un usuario que se aproxima a las tiendas, puede comprobar los servicios de recomendación disponibles en ese momento.

En la Tabla 2.2 se muestran las reglas de asociación relacionadas con la ubicación de un usuario y las tiendas. El atributo de rango de servicio establece el rango de servicio de una tienda.

La Figura 2.2 muestra un mapa con el rango de servicio de una tienda así como el del usuario. Cuando el rango de servicios de la tienda se superpone con el rango de búsqueda del usuario, entonces se genera la recomendación al usuario.

Tabla 2.2: Reglas de asociación para la información de ubicación

Rule Shops_in_user_search_area **is**
if Person(?User) **and** hasArea(?User, ?User_search_area) **and**
 hasBesidesArea(?User_search_area, ?Nearby_area_of_user) **and**
 hasShop(?Nearby_area_of_user, ?Nearby_shops) **and**
 hasServiceRange(?Nearby_shops, ?Shops_in_user_search_area) **or**
 hasInRangeShop(?User, ?Nearby_shops)
then hasServiceShop(?User, ?Shops_in_user_search_area) **and**
 hasServiceShop(?User, ?Nearby_shops);

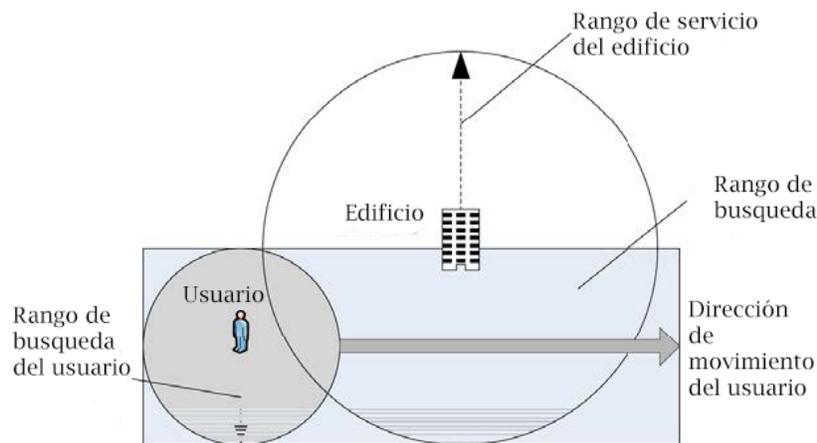


Figura 2.2: Rango de búsqueda entre una tienda y un usuario.

La propiedad *tiempo* también es importante en el sistema propuesto en esta tesis, ya que para la recomendación se deben considerar el horario del establecimiento y la disponibilidad del usuario.

La Tabla 2.3 muestra la disponibilidad tanto del usuario como de los horarios del cine. Esta asociación define la fecha y hora actual, así como el tiempo de que dispone el usuario para ver una película.

Tabla 2.3: Reglas de asociación para la información de horarios

```

Fact fTime is CurrentDateTime(2008-06-24T12:30:00);
Fact fAvailableTime is AvailableTime(PT5H);
Rule RecommendedMovieTheater is
if hasTimeTable(?RunningMoive, ?TimeTableOfMovies) and
    TimeTable_StartTime(?TimeTableOfMovies, ?StartingTimeOfMovie) and
    [func:after(?StartingTimeOfMovie,?CurrentDateTime) = true] and
    [func:before(?StartingTimeOfMovie,[?CurrentDateTime + ?AvailableTime]) =true]
then FinalRecommendationOfTheater(?Theater, ?RunningMovie,?StartingTimeOfMovie,
    ?RecommendedMovie);
    
```

A través de estas reglas, el sistema recomienda, basándose en los horarios del cine y la disponibilidad del usuario donde la película está disponible. La Figura 2.3 muestra un ejemplo de esta asociación de la Tabla 2.3. (En el caso de la Figura 2.3, recomienda la película # 2)

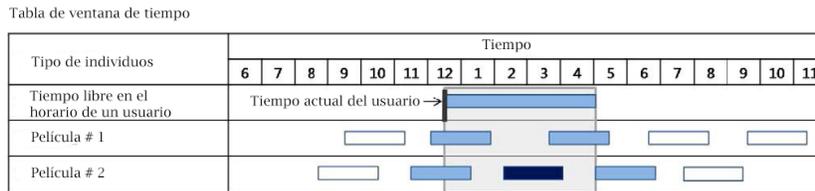


Figura 2.3: Ejemplo de horario de películas y tiempo disponible del usuario.

2.3. Perfil de usuario

En el trabajo de Yi Cai et al. Cai & Li (2010) podemos ver lo siguiente:

2.3.1. Etiquetado de colaboración y perfil de usuario basado en etiquetas

El etiquetado colaborativo, también llamado *folcsonoma*, ganó popularidad con los servicios de la web 2.0 en los años recientes. Se puede dividir en dos aspectos: el primero se enfoca en la investigación y análisis de las características de las etiquetas. Golder and Huberman Golder & Huberman (2006) analizaron las etiquetas usando patrones, las actividades del usuario

y anotando los recursos en los sistemas de etiquetado colaborativo. Bischoff et al. Bischoff et al. (2008) hicieron una encuesta más profunda a sistemas como *Del.icio.us* Delicious , y *Last.FM* Last.fm, para descubrir etiquetas útiles para el acceso a la información.

El segundo aspecto principalmente se enfocó de explorar las anotaciones sociales y enlazar estructuras dentro de la folcsonomía para varias aplicaciones. Las anotaciones sociales son buenas resumiendo los recursos etiquetados, para que éstos puedan ser usados para búsquedas de información. Bao et al. Bao et al. (2007) propone dos algoritmos, llamados *SocialSimRank* y *SocialPageRank* respectivamente, para explorar la latente semántica de las etiquetas para optimizar la búsqueda en la web.

2.3.2. Búsquedas personalizadas

La motivación de proponer búsquedas personalizadas es regresar a los usuarios resultados personalizados acordes a la información que ellos necesitan. La estrategia de la búsqueda personalizada recae en dos categorías:

- Una es la expansión de la búsqueda como Chirita et al. (2007), en donde refiere a modificar la búsqueda original ya sea expandiéndola con otros términos, o asignando diferentes pesos a los términos en la búsqueda.
- Otra categoría que resulta al procesar, primeramente reclasificando, lo cual adapta los resultados de la búsqueda a la preferencia particular del usuario. Más estrategias reclasificadoras intentan construir un perfil de usuario desde un comportamiento histórico, y usar este perfil para filtrar los recursos que no coincidan con sus intereses. Pretschner and Gauch Pretschner & Gauch (1999) propusieron perfiles de usuario estructurados con una ontología, la cual consiste de 4,400 nodos. Chirita et al. Chirita et al. (2005) modelaron perfiles de usuarios y recursos como vectores tópicos de la jeraquía de Dmoz dmoz, así como el emparejamiento entre los intereses del usuario, en donde el contenido puede ser medido por la distancia del vector.

2.3.3. Búsqueda personalizada en sistemas de etiquetado colaborativo

Noll and Meinel Noll & Meinel (2007) proponen una aproximación a la exploración de los usuarios y los recursos relacionados a las etiquetas, basado en la frecuencia de los términos. A su vez, se reclasifican las búsquedas no personalizadas basado en esas etiquetas relacionadas.

Xu et al. Xu et al. (2008) propusieron tópicos personalizados buscados en folcsonomía, en el cual las búsquedas personalizadas son conducidas por una clasificación de recursos basados no solo en los términos de similitud sino también en los tópicos similares.

2.3.4. El usuario y recursos de perfiles de etiquetado colaborativo

2.3.4.1. Modelado de usuario

Considerando que hay dos tipos de etiquetas que el usuario pueda interesarle. La primera clase de etiquetas son las etiquetas que son realizadas por el usuario. Si un usuario ha utilizado una etiqueta para describir un recurso, indica explícitamente que el usuario puede considerar la información de la etiqueta como interesante (o atractiva) para él. Especialmente en los recursos de la receta, de acuerdo con los comentarios de los usuarios del sistema propuesto, a la mayoría de los usuarios les gustaría utilizar esas etiquetas que pueden indicar las características de preferencia de un recurso para el etiquetado. Podemos nombrar el conjunto de etiquetas que se utilizan por un usuario para anotar recursos de manera explícita como un vector de etiqueta con intereses para el usuario.

Por tanto, un perfil de usuario i denotado por \vec{U}_i , es un vector de etiqueta: parejas de valores de la siguiente manera:

$$\vec{U}_i = (t_i, 1 : v_i, 1, t_i, 2 : v_i, 2, \dots, t_i, n : v_i, n), \forall x, t_i, x \in \vec{T}_i \cup \vec{T}_i^H,$$

donde t_i, x es una etiqueta que puede ser interesante para un usuario; i, v_i, x

es el grado de preferencia del usuario i en la etiqueta t_i, x y n es el número de etiquetas en que el usuario i puede estar interesado.

Capítulo 3

Marco Teórico

3.1. Ontologías

La teoría y las aplicaciones de una ontología han sido definidas en Poli (2010), en donde se describe la noción de ontología desde dos puntos de vista: el tradicional de la filosofía y otro orientado a ciencias de la computación, particularmente Inteligencia Artificial. Esta última perspectiva de la ontología se vislumbra como tecnología que se enfoca en crear modelos de ingeniería de la realidad, artefactos que pueden ser utilizados por el software, y tal vez interpretados directamente y razonados por un software denominado motor de inferencia, el cual incluye al software con el nivel de semántica más apegado al ser humano.

En el trabajo de Corcho et al. (2005) se menciona que cuando una aplicación de la tecnología llega a un cierto grado de madurez, se convierte en una ingeniería, la cual proporciona métodos, metodologías y herramientas para la aplicación. Los métodos establecen “cómo” realizar determinada actividad, las metodologías establecen “qué”, “quién” y “cuando” Corcho et al. (2006). Así, las herramientas sirven para dar soporte a dichas actividades. La ingeniería ontológica abarca las actividades relacionadas con el proceso de desarrollos de ontologías, ciclo de vida, métodos y metodologías para su construcción, y con las herramientas y lenguajes que le dan soporte Gómez-Pérez et al. (2006).

3.1.1. Construcción de una ontología

En Uschold & King (1995) se describe que la construcción de una ontología se divide en tres secciones: captura, codificación y la integración de las ontologías existentes.

3.1.1.1. Captura

En Uschold & King (1995) se presenta que la captura de la ontología significa:

1. Identificación de la llave de conceptos y relaciones en el dominio de interés.
2. Producción de definiciones precisas de textos ambiguos para tales conceptos y relaciones.
3. Identificación de términos para referirse a tales conceptos y relaciones.
4. Acordando todo lo anterior.

3.1.1.2. Codificación

En Uschold & King (1995) cita que la codificación es una representación explícita de la conceptualización capturada en la sección anterior en algún lenguaje formal. Esto implicará la comisión de alguna meta-ontología, escogiendo un lenguaje representativo, y creando el código. Algunos lenguajes son *Prolog*, *Conceptual Graphs*, *L-Lilog*, *Otolingua* y muchos lenguajes de la familia de *KL-ONE* (*Back*, *Back++*, *Loom*, *Classic*).

3.1.1.3. Integración de las ontologías existentes

En Uschold & King (1995) se menciona que durante o en las capturas o procesos de codificación, hay una pregunta de cómo usar las ontologías que ya existen. En general, este es un problema difícil y a pesar del progreso que se ha hecho en el contexto de *Otolingua*, el problema general es muy complicado. Proveer la guía y herramientas en esta área puede ser uno de los mayores retos para desarrollar una metodología comprensiva.

3.2. WordNet

En Miller (1995) se puede observar que WordNet® es una gran base de datos léxica del idioma inglés. Nombres, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos (*synsets*), cada uno expresando un concepto distinto. Los *synsets* están vinculados entre sí, por medio de las relaciones conceptuales semánticas y léxicas. WordNet también es libre y públicamente disponible para su descarga. La estructura de WordNet hace que sea una herramienta útil para la lingüística computacional y procesamiento del lenguaje natural.

WordNet superficialmente se parece a un *tesauro*, en el que se agrupan las palabras basándose en sus significados. Sin embargo, hay algunas diferencias importantes. En primer lugar, WordNet crea e interrelaciona no sólo las formas de las palabras -cadenas de letras- pero especifica los sentidos de las palabras. Como resultado de ello, las palabras que se encuentran en estrecha proximidad entre sí en la red son semánticamente desambiguas. En segundo lugar, WordNet etiqueta las relaciones semánticas entre las palabras, mientras que las agrupaciones de palabras en un *tesauro* no sigue ningún patrón explícito que no sea el sentido de similitud.

3.3. NLTK (Natural Language Toolkit)

En NLTK se presenta NLTK como la plataforma líder para la creación de programas de Python para trabajar con los datos del lenguaje humano. Proporciona interfaces fáciles de usar para más de 50 *corpus* y recursos léxicos como WordNet, junto con un conjunto de bibliotecas de procesamiento de texto para la clasificación, tokenización, derivados, etiquetados, análisis y de razonamiento semántico, contenedores para las bibliotecas NLP de potencia industrial, y un foro de discusión activa.

NLTK ha sido llamado como “una maravillosa herramienta para la enseñanza y el trabajo en lingüística computacional, utilizando Python’ y “una biblioteca increíble para jugar con el lenguaje natural”.

3.3.1. Uso de los Synsets con NLTK

En pythonprogramming.net menciona que WordNet es una base de datos léxica para el idioma inglés, que fue creada por Princeton, y es parte del corpus NLTK.

Se puede usar WordNet junto con el módulo NLTK para encontrar los significados de palabras, sinónimos, antónimos y más.

El procedimiento es el siguiente:

Importar wordnet:

```
from nltk.corpus import wordnet
```

Posteriormente, se utiliza el término "programa" para encontrar conjuntos de sinónimos como:

```
syns = wordnet.synsets("program")
```

Un ejemplo de un synset:

```
print(syns[0].name())
```

plan.n.01

Solo la palabra

```
print(syns[0].lemmas()[0].name())
```

plan

Definición de ese primer synset:

```
print(syns[0].definition())
```

a series of steps to be carried out or goals to be accomplished

Ejemplos de la palabra en uso:

```
print(syns[0].examples())
```

['they drew up a six-step plan', 'they discussed plans for a new bond issue']

Para discernir los sinónimos y los antónimos de una palabra, los lemas serán sinónimos, y posteriormente se procede a usar antónimos para encontrar los antónimos en los lemas. Se puede rellenar algunas listas como:

```
synonyms = [ ]
antonyms = [ ]

for syn in wordnet.synsets("good"):
    for l in syn.lemmas():
        synonyms.append(l.name())
        if l.antonyms():
            antonyms.append(l.antonyms()[0].name())

print(set(synonyms))
print(set(antonyms))
```

{'beneficial', 'just', 'upright', 'thoroughly', 'in_force', 'well', 'skilful', 'skillful', 'sound', 'unspoiled', 'expert', 'proficient', 'in_effect', 'honorable', 'adept', 'secure', 'commodity', 'estimable', 'soundly', 'right', 'respectable', 'good', 'serious', 'ripe', 'salutary', 'dear', 'practiced', 'goodness', 'safe', 'effective', 'unspoil', 'dependable', 'undecomposed', 'honest', 'full', 'near', 'trade_good'} {'evil', 'evilness', 'bad', 'badness', 'ill'}

Como se puede ver, se tienen muchos más sinónimos que antónimos, ya que solo se busca el antónimo para el primer lema, pero se puede fácilmente equilibrar esta compra también haciendo exactamente el mismo proceso para el término “bad”.

A continuación, también se usar WordNet para comparar la similitud de dos palabras y sus tiempos, incorporando el método de Wu y Palmer para la

relación semántica.

A continuación se compara el sustantivo de “ship” y “boat”:

```
w1 = wordnet.synset('ship.n.01')
w2 = wordnet.synset('boat.n.01')
print(w1.wup_similarity(w2))
```

0.9090909090909091

```
w1 = wordnet.synset('ship.n.01')
w2 = wordnet.synset('car.n.01')
print(w1.wup_similarity(w2))
```

0.6956521739130435

```
w1 = wordnet.synset('ship.n.01')
w2 = wordnet.synset('cat.n.01')
print(w1.wup_similarity(w2))
```

0.38095238095238093

3.4. Realidad aumentada

En el trabajo de Lara & Villarreal (2004) se presenta el funcionamiento entre la integración de objetos y el mundo real y el virtual. Éstos pueden ser agregados, fusionados, combinados e inclusive intercambiados. Si el entorno del espacio es predominantemente virtual y se le agregan objetos virtuales y reales, estamos hablando de realidad virtual; pero si el entorno dominante es real y se le agregan objetos virtuales, estamos hablando de realidad aumentada.

La realidad aumentada es una tecnología que integra señales captadas del mundo real (generalmente video y audio), con señales generadas por

computadoras (objetos gráficos tridimensionales); las hace corresponder para construir nuevos mundos coherentes, complementados y enriquecidos para coexistir entre sí.

Un sistema de realidad aumentada generalmente inicia con el registro de las señales del mundo real. Estas señales son procesadas por un sistema de realce de orillas de objetos para preparar la imagen para la segmentación o extracción de objetos y el reconocimiento de patrones y marcas fiduciaras. Este proceso permite determinar en dónde hay que remplazar un objeto real por uno virtual, cuál objeto virtual debe colocarse sobre el espacio real (el espacio de video) y en qué posición y perspectiva.

Estos procesos resultan en un sistema con las siguientes características, las cuales definen la realidad aumentada:

- Combina objetos reales y virtuales en nuevos ambientes integrados.
- Las señales y su reconstrucción se ejecutan en tiempo real.
- Las aplicaciones son interactivas.
- Los objetos reales y virtuales son registrados y alineados geométricamente entre ellos y dentro del espacio, para darles coherencia espacial.

En el trabajo presentado en Fombona Cadavieco et al. (2012), relativo a la realidad aumentada se presenta una evolución de las aplicaciones de ésta en dispositivos móviles, en donde se puede observar en la siguiente Figura cómo se reacciona al encontrar un patrón.

<i>1- Marca que desencadena un video</i>		
<i>Marca original</i>	<i>Imagen captada con video incorporado</i>	
		
<i>2- Software que reconoce un esquema icónico y superpone otra imagen</i>		
<i>Imagen original</i>	<i>Esquema icónico y datos almacenados.</i>	<i>Resultado presentad. Ambos iconos aparecen con movimientos solidarios.</i>
		
<i>3- Software reconoce un código icónico predefinido y superpone otra imagen o hipervínculo.</i>		
<i>Código digital o barras</i>	<i>Se muestra un hipervínculo de Internet asociado</i>	
		

Figura 3.1: Patrones de disparo en software de RA.

En el trabajo descrito en Hamynen et al. (2010), se presenta una técnica para desplegar redes de objetos en dispositivos móviles basados en geolocalización. En las imágenes siguientes se puede apreciar como funciona la realidad aumentada combinada con geolocalización.

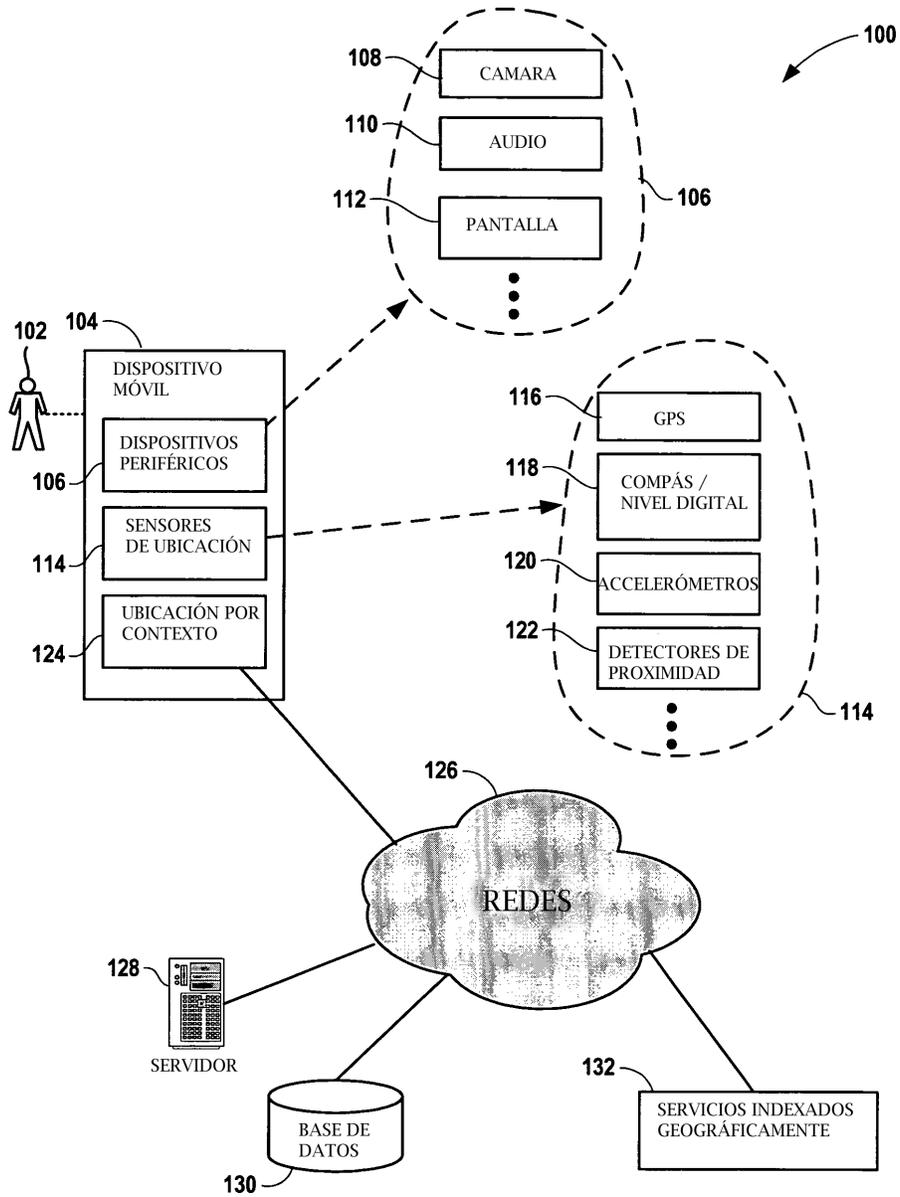


Figura 3.2: Proceso de realidad aumentada con geolocalización. Fig 1

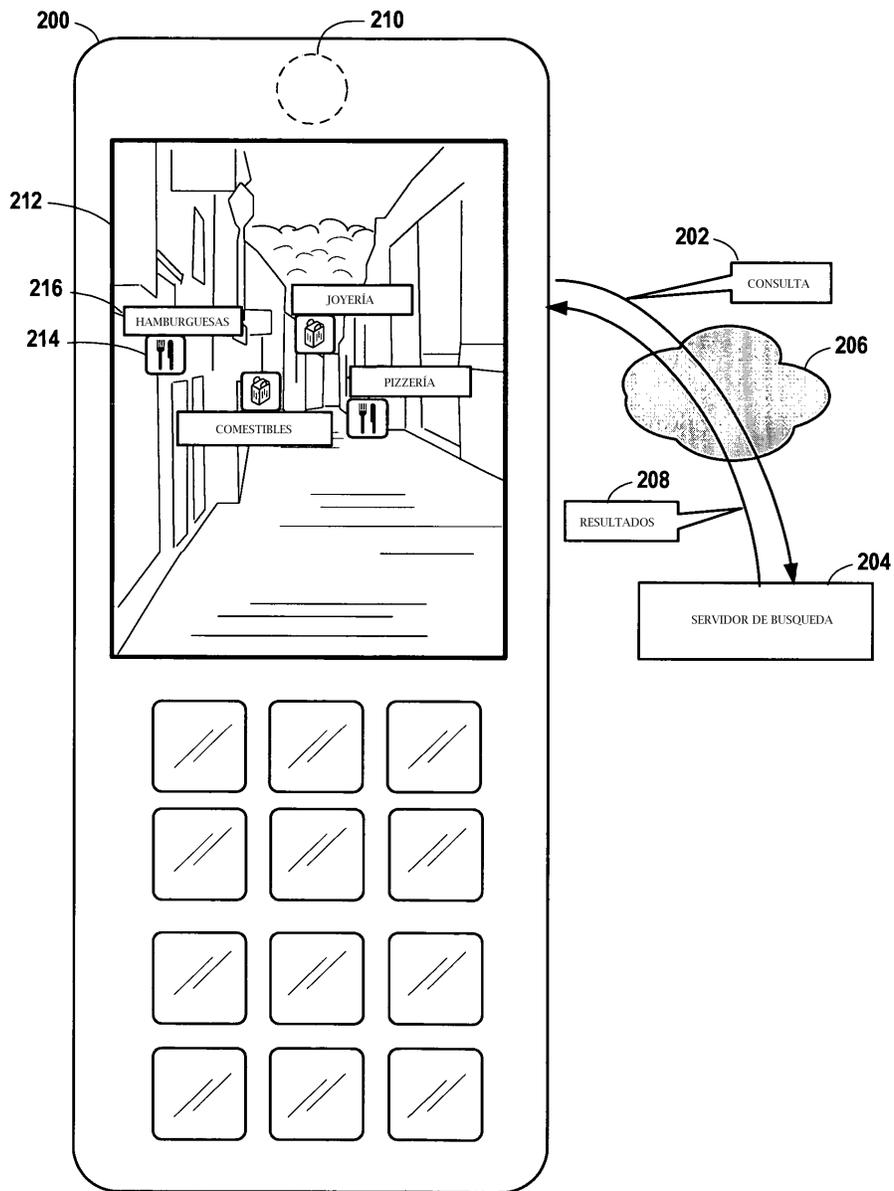


FIG. 2

Figura 3.3: Proceso de realidad aumentada con geolocalización. Fig 2

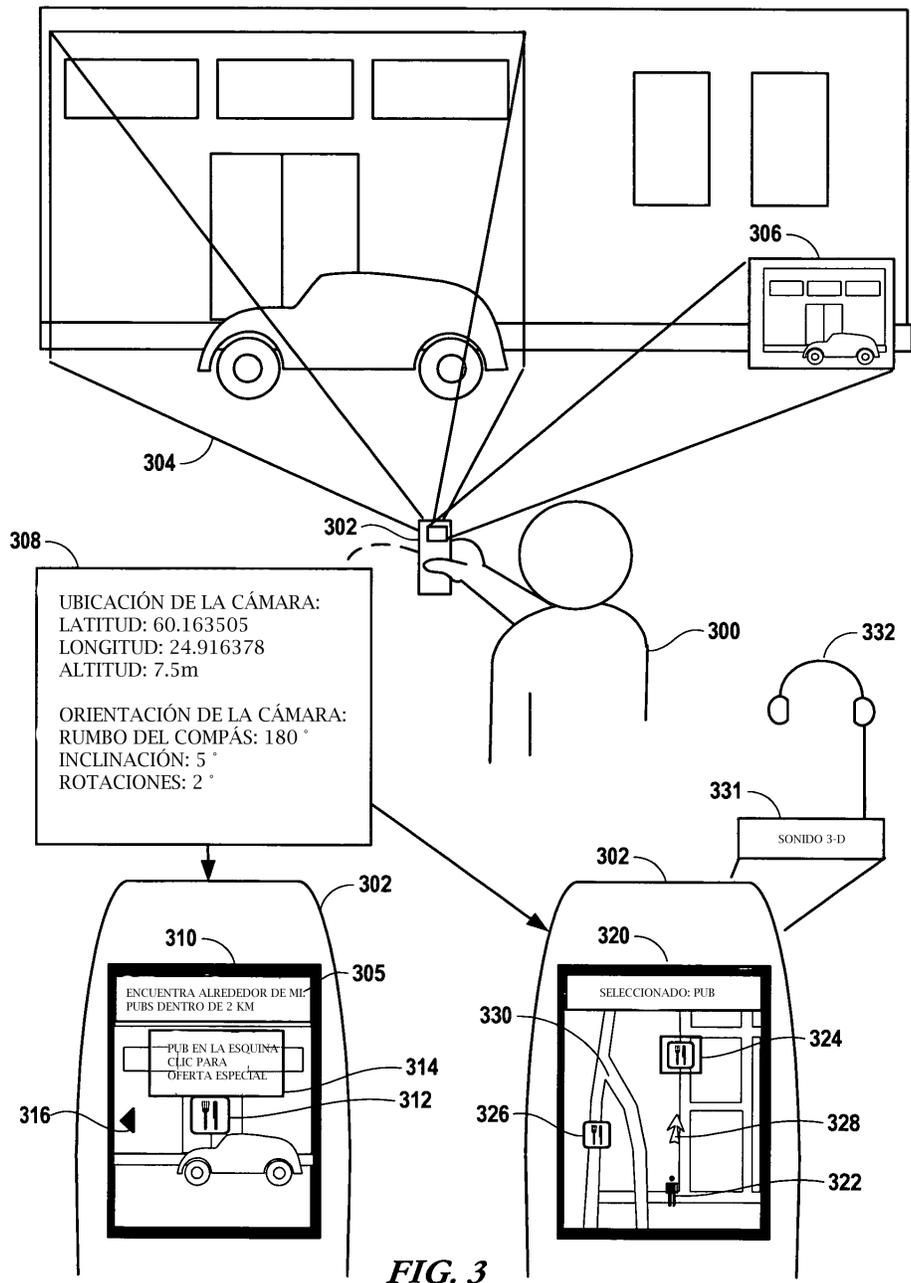


Figura 3.4: Proceso de realidad aumentada con geolocalización. Fig 3

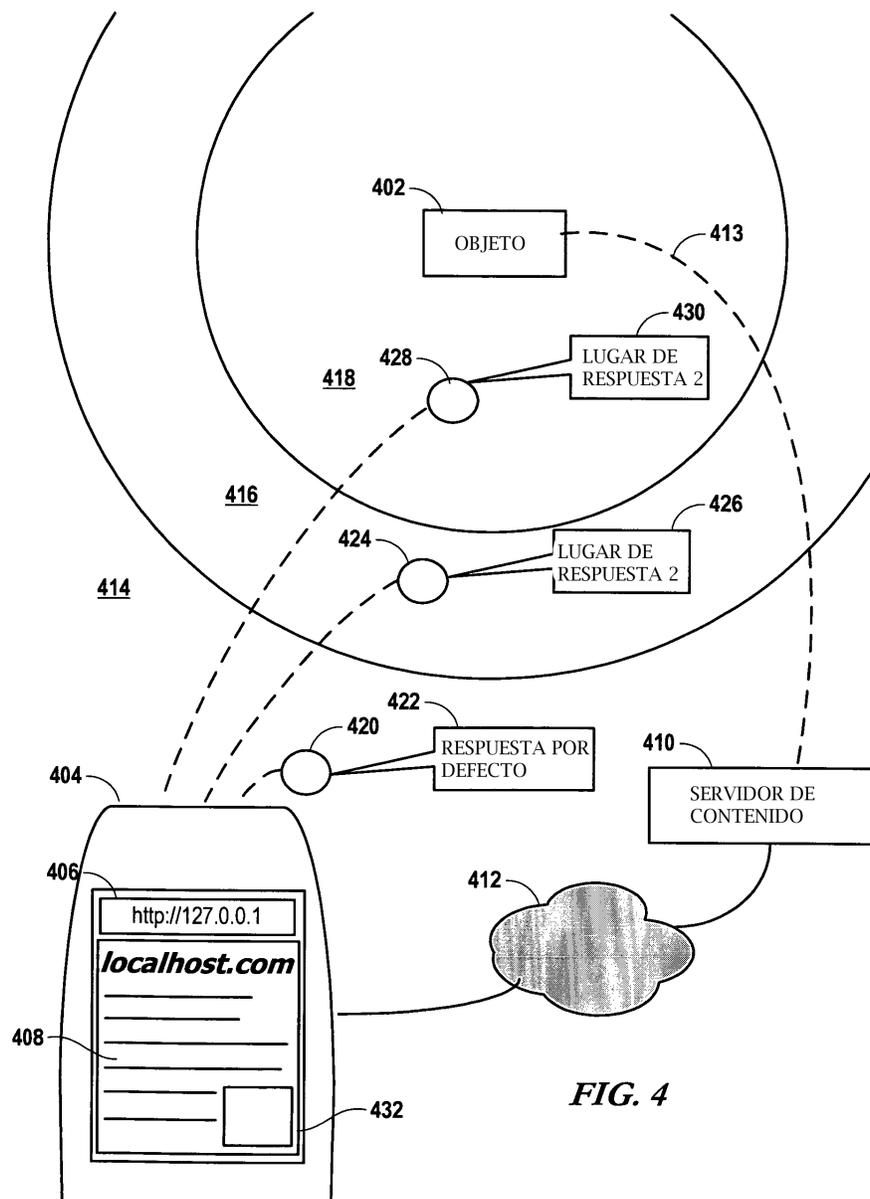


FIG. 4

Figura 3.5: Proceso de realidad aumentada con geolocalización. Fig 4

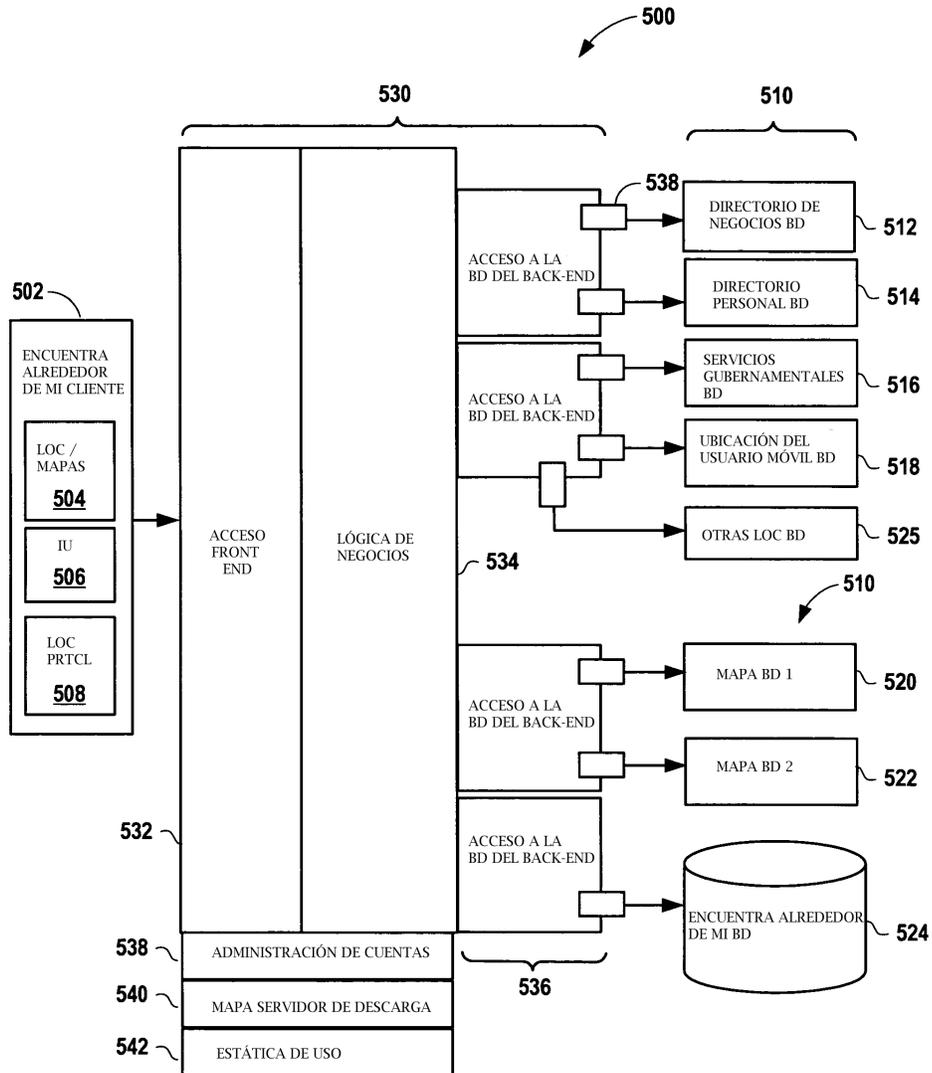


FIG. 5

Figura 3.6: Proceso de realidad aumentada con geolocalización. Fig 5

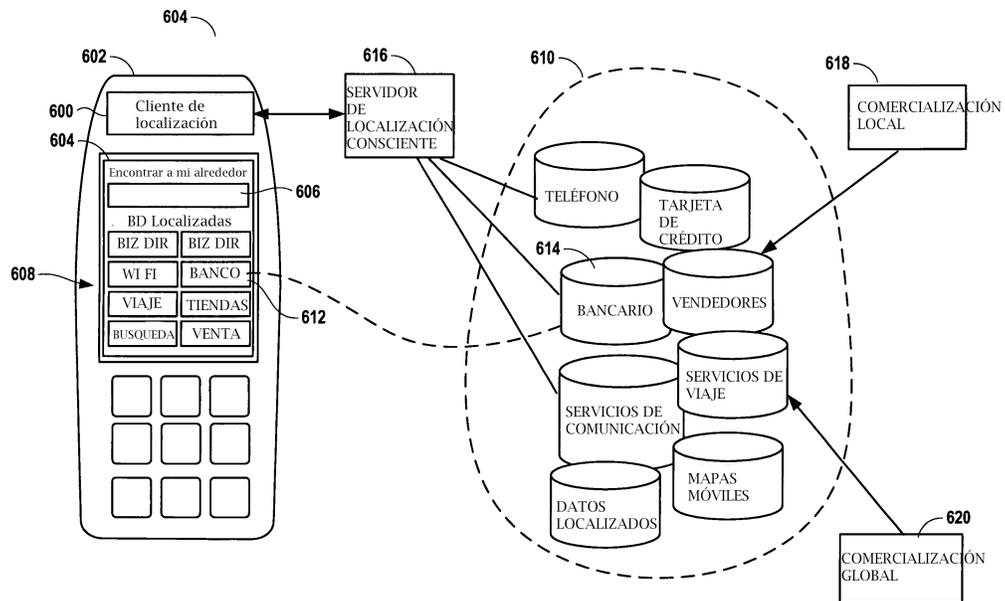


FIG. 6

Figura 3.7: Proceso de realidad aumentada con geolocalización. Fig 6

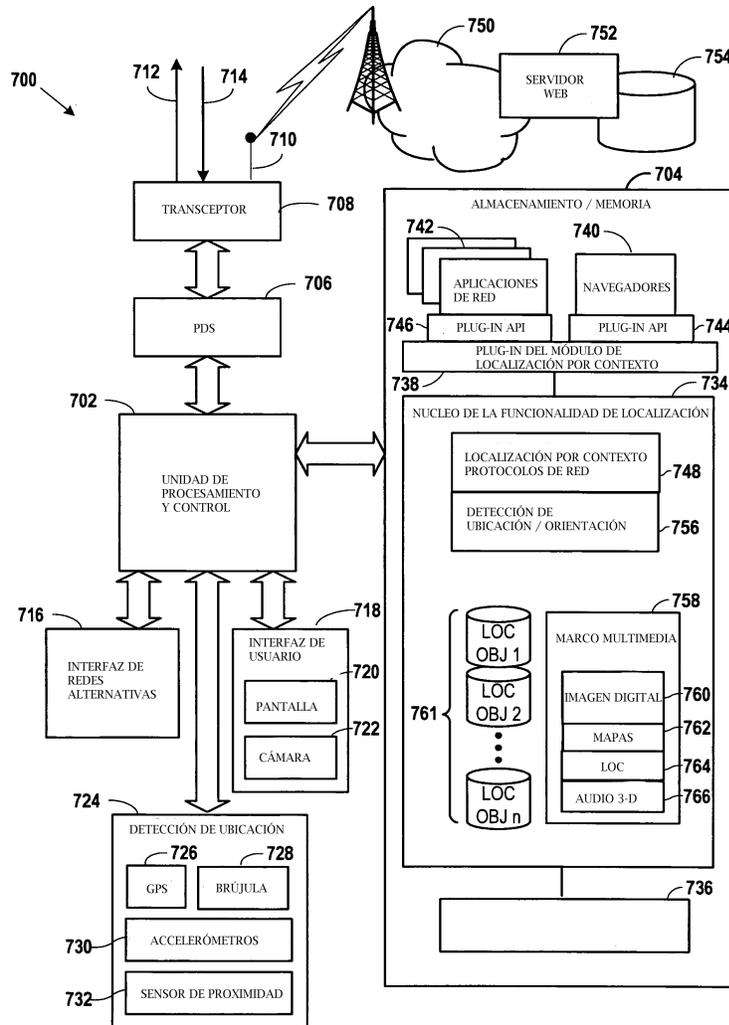


FIG. 7

Figura 3.8: Proceso de realidad aumentada con geolocalización. Fig 7

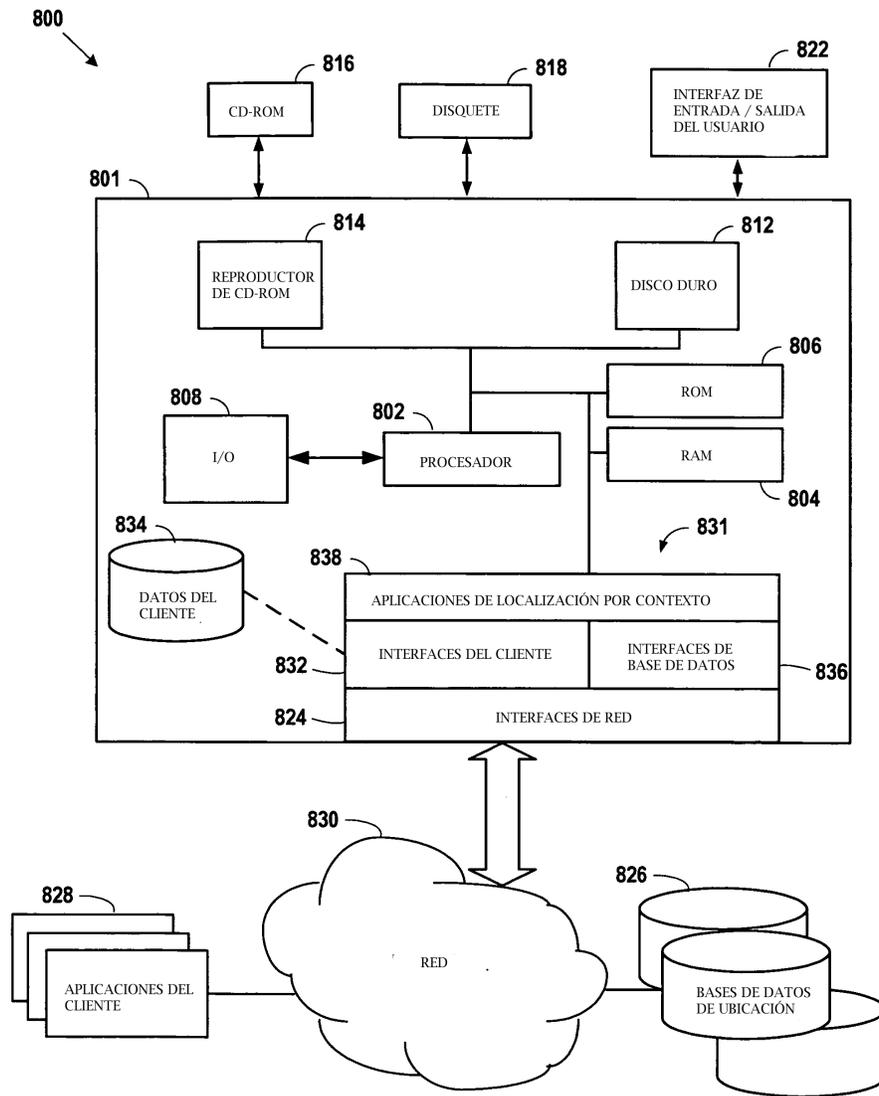
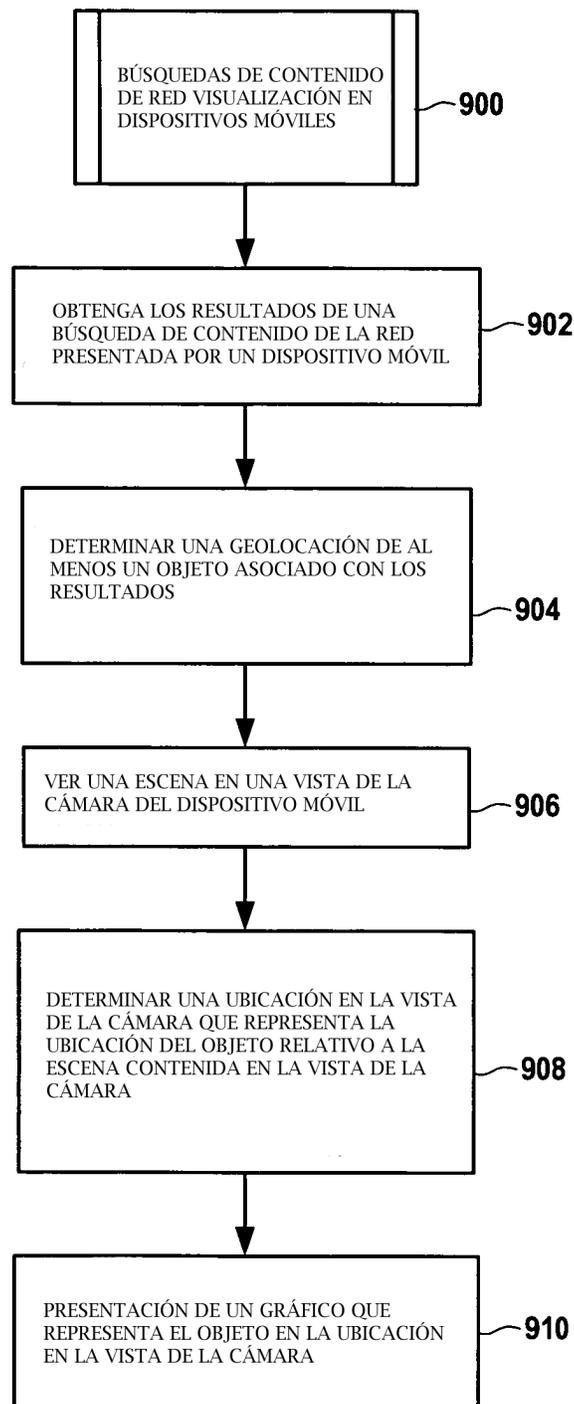
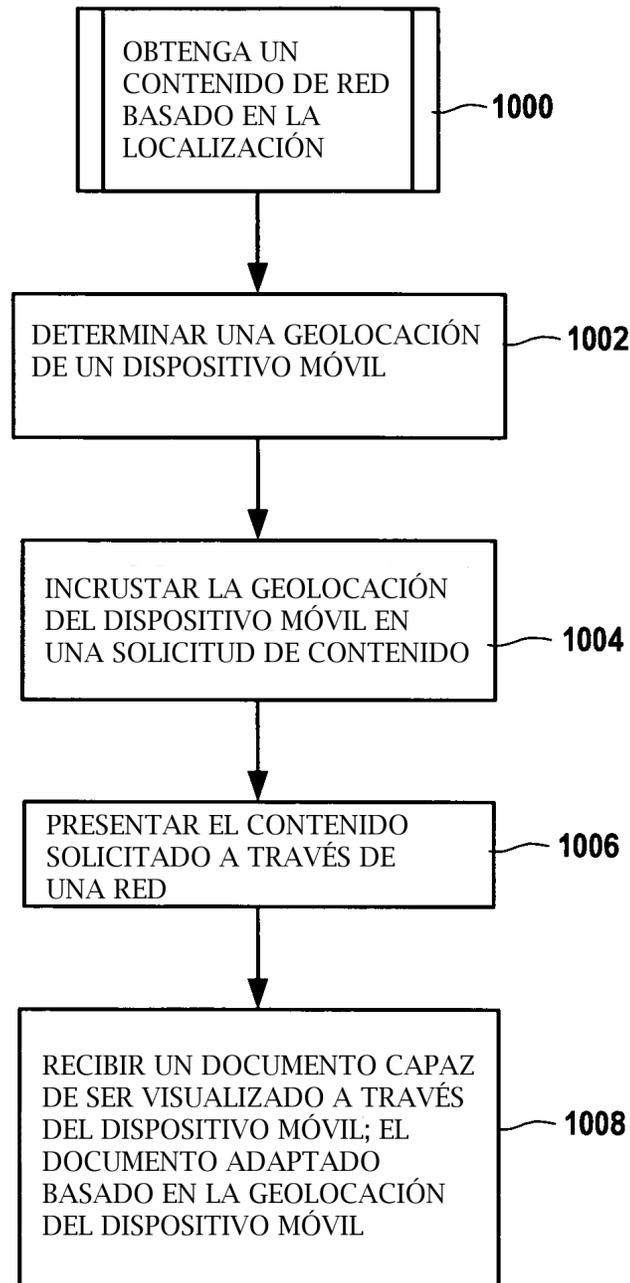


FIG. 8

Figura 3.9: Proceso de realidad aumentada con geolocalización. Fig 8

**FIG. 9**

**FIG. 10**

3.4.1. Técnicas de visualización para la Realidad Aumentada

En Furht (2011) se cita que las aplicaciones de realidad aumentada enriquecen el entorno del mundo real con una cierta cantidad de información sintética, lo ideal es justo lo suficiente para superar las limitaciones del mundo real para una aplicación específica. Azuma et al. Azuma (2013) definen tres requisitos de una aplicación de realidad aumentada. Azuma et al. Azuma (2013) afirmó que además de una mezcla de imágenes reales y virtuales, las aplicaciones de realidad aumentada que se tengan que ejecutar en tiempo real y los objetos virtuales tienen que ser alineados con las estructuras del mundo real. En la Figura 3.12 Furht (2011) se muestra un ejemplo de un sistema de realidad aumentada común y los datos que se adquieren, que se calculan y presentan; con el fin de registrar un monstruo virtual, el sistema de información de seguimiento de realidad aumentada deriva de la entrada de video. Después de hacer que la estructura 3D sea registrada, su superposición permite generar la impresión de una figura virtual de pie en una tarjeta de papel del mundo real.



Figura 3.12: El flujo de datos en un sistema de realidad aumentada común. Un mundo real imaginario es entregado por la alimentación de video del sistema y procesada por algoritmos de seguimiento basados en visión. Para alinear los datos reales y virtuales, los datos de seguimiento derivados se les ha aplicado una transformación del contenido virtual. Por último, la prestación se superpone en la parte superior de la secuencia de video.

Por otra parte, la sustitución por descuido de las partes del mundo real, implica que las imágenes también pueden ocultar estructuras importantes presentes en el entorno real. En la Figura 3.13 se muestra cómo la representación generada por una computadora muestra la anatomía interior de un paciente; sin embargo, se obstruye la vista de las señales que dan gran relevancia y que están presentes en las imágenes del mundo real. En este ejemplo, el objetivo es insertar una aguja en el abdomen de un paciente, utilizando un punto de entrada predefinido (marcas de color negro). Al superponer los órganos virtuales en la parte superior de lo real al mundo imaginario, el usuario es incapaz de ver los puntos de entrada que están marcados en la piel del paciente.

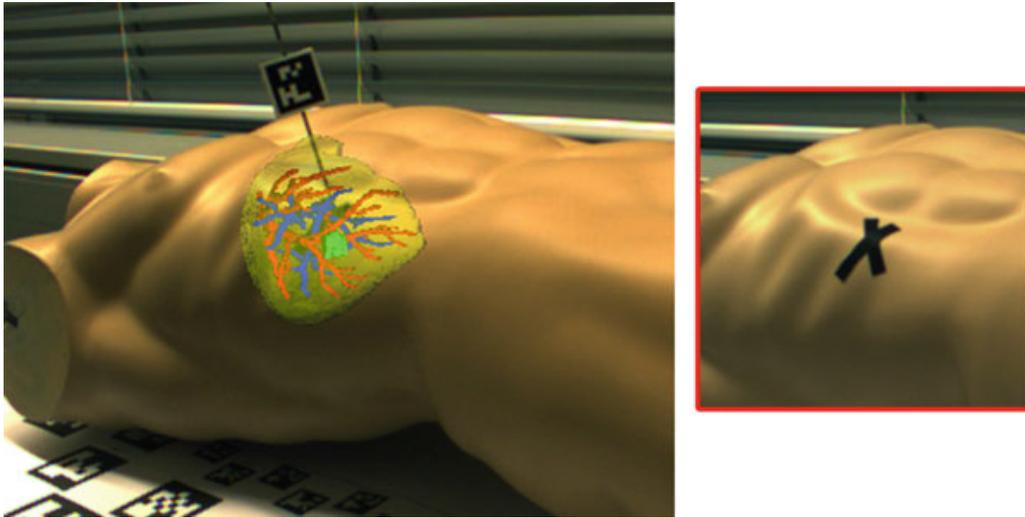


Figura 3.13: Aumento indebido. En este ejemplo, los aumentos por descuido de estructuras ocultas causan dos problemas fundamentales: se anula información útil (como puntos de referencia) y carecen de señales de profundidad. (Derecha) escena original antes del aumento, la cruz de color negro indica el punto de inserción de la aguja RFA. (Izquierda) Aumento del hígado con el portal y los árboles de los vasos hepáticos y un tumor.

Además de los problemas causados por la eliminación de las imágenes del mundo real, la generación y descuido de la visualización en ambientes de realidad aumentada puede dar lugar a interacciones engañosas de colores y tonos que representan lo real y los objetos virtuales. Si la representación de los objetos virtuales no toma en cuenta el entorno del mundo real, la composición de ambos puede fallar en la intención de la visualización. Las visualizaciones comprensibles en entornos de realidad aumentada exigen que la apariencia de los gráficos del ordenador en 3D añadidos encajen en el ambiente del mundo real. Por consiguiente, visualizaciones exitosas tienen que constar de elementos fácilmente distinguibles. Por ejemplo, en la Figura 3.14 se muestra una visualización de rayos x con apariencia similar de ambas estructuras ocultas y oclusión. A pesar de que las señales de profundidad se conservan, es difícil distinguir los elementos de la visualización. Como resultado, las relaciones espaciales son difíciles de percibir.

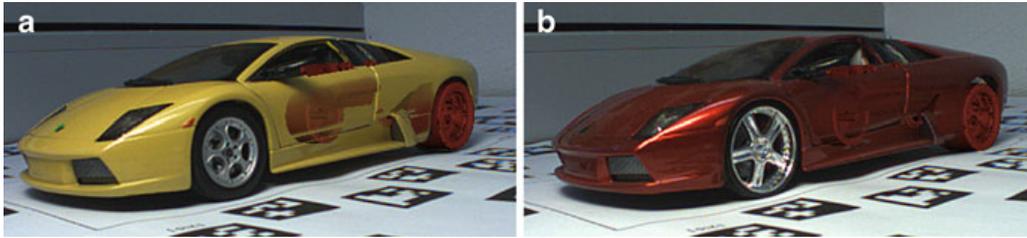


Figura 3.14: La interacción visual entre las representaciones reales y virtuales. Las visualizaciones en (a) y (b) utilizan los mismos parámetros. Sin embargo, mientras que la visualización en (a) presenta claramente todos los objetos importantes, los elementos aumentados apenas son visibles en (b).

3.4.2. Integración de los datos

En Furht (2011) se menciona que una simple superposición de la estructura oculta en la parte superior de la alimentación del video del sistema puede causar una serie de problemas cognitivos, causados por los procesos implicados en la creación de la impresión de profundidad. La comprensión de estas causas permite desarrollar técnicas de renderizado que suman y preservan esa información en visualizaciones de realidad aumentada con éxito.

3.4.2.1. Percepción de profundidad

Furht (2011) describe que nuestro sistema cognitivo tarda aproximadamente de 15-20 diferentes estímulos psicológicos en cuenta, con el fin de percibir las relaciones espaciales entre objetos 3D Wilson & D´ Cruz (2006). Estas llamadas señales de profundidad se pueden dividir en monocular y binocular. Mientras que las señales de profundidad binocular requieren el uso de dos ojos, señales monoculares aparecen incluso cuando un ojo está cerrado.

Las señales de profundidad monocular se pueden dividir en las señales de las imágenes, señales de profundidad dinámicas y señales oculomotores. Las señales de profundidad dinámicas son causadas por el hecho de que los objetos más alejados parecen moverse más lento que los objetos cercanos al mover nuestro punto de vista. Nuestro sistema cognitivo traduce la diferencia de velocidad en una aproximación de las distancias entre los objetos. En consecuencia, esas señales de profundidad aparecen si cualquiera de los objetos

en el entorno 3D o en el movimiento del observador.

Las señales oculomotoras son causadas por la sensación que se produce cuando los ojos convergen y se cambian de forma para enfocar objetos cercanos. La sensación causada por la contracción de los músculos del ojo y la alteración de la forma de su lente se interpreta como la distancia desde el objeto. Además de centrarse en objetos, que están muy cerca, se hace más hincapié en los ojos, se centra en las estructuras más lejanas, relaja los músculos y las lentes; esto hace menor este sentimiento en particular.

Las señales de profundidad ilustradas son aquellas que se pueden encontrar en una sola imagen incluyendo:

- Oclusión: si las proyecciones 2D de dos objetos en el ambiente se superponen, los objetos que están más cerca del observador obstruyen a los objetos que están más lejos.
- Tamaño relativo: los objetos más distantes parecen más pequeños que los objetos cercanos.
- Altura relativa: objetos con bases más altas en la imagen parecen estar más lejos.
- Detalle: objetos que están más cerca ofrecen mayor detalle.
- Perspectiva atmosférica: debido al polvo en la atmósfera, los objetos que están más alejados parecen más borrosos que los que están cerca.
- Sombras: dependiendo en la posición del origen de la luz, las sombras pueden ser emitidas desde un objeto a otro.
- Perspectiva lineal: líneas paralelas convergen al aumentar la distancia. Observe cómo las aceras parecen converger en algún lugar infinito, aunque en realidad parecen ser aproximadamente paralelas.

A pesar de que las señales de profundidad monocular le permiten a uno hacer juicios de profundidad, nuestra impresión típica de un objeto 3D es causada por las señales de profundidad binocular. Este tipo de señal explota la diferencia entre las proyecciones 2D de un punto en el espacio 3D en las retinas de la izquierda y el ojo derecho. Los puntos correspondientes de la

retina se alejan más y más cerca del punto 3D obtenido. Sin embargo, con el fin de proporcionar al usuario una aplicación de realidad aumentada con las señales de profundidad binocular, dispositivos de visualización estereoscópicas especiales tienen que ser utilizados para proporcionar al usuario las dos versiones al mismo tiempo.

3.4.2.2. Aumento en las señales de profundidad pictóricas

En Furht (2011) se describe que la representación de la estructura virtual mediante una cámara que utiliza parámetros que reflejan las características de la cámara real, la fusión de las imágenes del mundo virtual y real proporcionará automáticamente las señales de profundidad pictóricas que se ajustan a los presentes en el entorno del mundo real.

La figura 3.15 Furht (2011) muestra un ejemplo el aumento de una figura Lego virtual que representa (a) por detrás y (b) a la izquierda de dos figuras reales de Lego. Entre otros factores, la perspectiva lineal, altura relativa y el tamaño relativo indican la relación espacial de la figura virtual, en relación con las figuras reales. Dado que los parámetros de la cámara virtual se han alineado con los de la real, las señales de profundidad de ambas presentaciones se alinean. A pesar de que una serie de señales de profundidad están perdidas (por ejemplo, no se proyectan sombras y la iluminación del mundo real y virtual no coinciden), las figuras virtuales de Lego en (a) y (b) se perciben en su posición correcta en el espacio 3D.

La sincronización de otros parámetros puede aumentar aún más la percepción de visualización AR. Por ejemplo, Fischer et al. Wellner, así como Okumura et al. De Courcy (1992) demostraron la alineación de ruido de la imagen y la borrosidad causada por el desenfoque y el movimiento. Más tarde Klein et al. PinkFroot mostró un conjunto más amplio de parámetros sincronizados incluyendo el color y la distorsión de la lente.

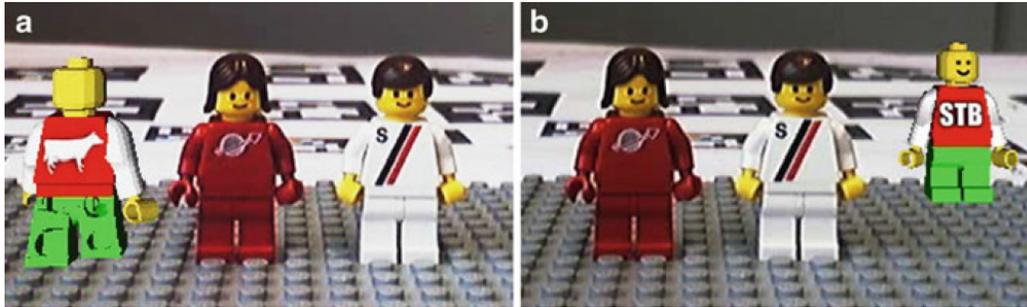


Figura 3.15: La sincronización del parámetro de lo virtual y lo real permite a la cámara alinear señales de profundidad pictóricas reales y virtuales. La figura Lego virtual en (a) se percibe correctamente junto a los datos reales, mientras que el virtual en (b) se percibe correctamente detrás de ambos. Este efecto se consigue por señales de profundidad de alineación, tales como la distorsión de la perspectiva y el tamaño relativo.

3.4.2.3. Manejo de la oclusión

En Furht (2011) se menciona que aunque las representaciones de las cámaras reales y virtuales sincronizadas, ya son capaces de alinear señales de profundidad, tan pronto como aparezcan las oclusiones entre los objetos reales y virtuales, esas señales de profundidad ya no son suficientes para producir los aumentos creíbles. En la Figura 3.16(a) se muestra que todas las demás señales de profundidad se han añadido a la pantalla de realidad aumentada, el objeto virtual se percibe como flotando en frente de la imagen de video. Una integración viable de la estructura virtual en el entorno del mundo real, es posible sólo si se han resuelto las oclusiones entre los objetos reales y virtuales Figura 3.16(b).

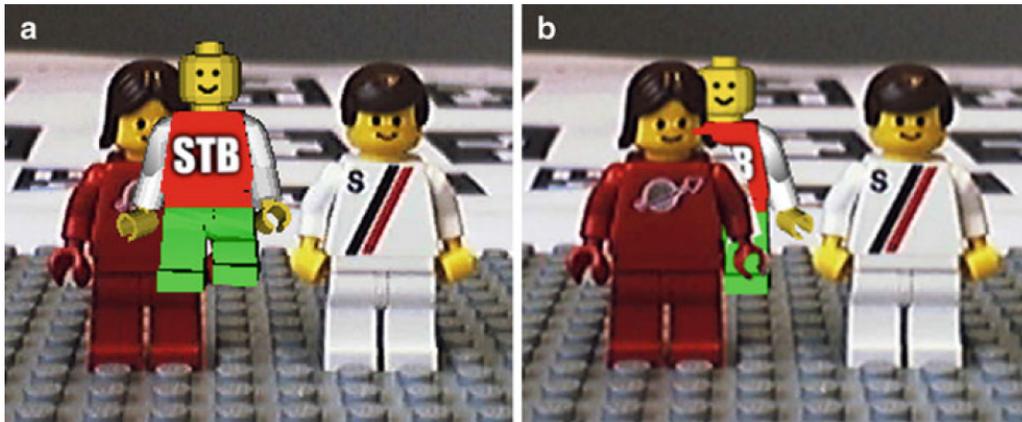


Figura 3.16: La importancia de las señales de oclusión (a) A pesar de que existe un número de señales de profundidad diferentes, su orden de profundidad es ambigua y la percepción está mal si las oclusiones han sido ignoradas (b) El mismo renderizado que presentado en (a) con oclusión es resuelto correctamente. Esta visualización es capaz de comunicar la relación espacial entre su contenido real y virtual.

Algoritmo 1 Manejo de oclusión utilizando objetos fantasmas

1. Dibujar video
 2. Deshabilite la escritura al buffer de color
 3. Renderizar representaciones virtuales de objetos reales (Fantasmas)
 4. Permitir la escritura al buffer de color
 5. Dibujar objetos virtuales
-

3.4.2.4. Características de los bordes

En Furht (2011) se describe que varios investigadores han demostrado el valor de preservar o mejorar el resultado de un detector de bordes en el video de alimentación Avery et al. (2009); Kalkofen et al. (2009) del sistema de realidad aumentada. Por ejemplo, las Figuras 3.17 y 3.18 muestran conservaciones de borde mediante la aplicación de un detector de bordes, con base en la imagen para las imágenes de video actual. Mediante el uso de sólo los datos de video como una fuente para la conservación, se han detectado otras características que pertenecen al objeto de oclusión. En la Figura 3.17 los

bordes en la marioneta, sino también los bordes del marcador ARToolkit y la aguja (que se inserta en un tumor en esta cirugía simulada) fueron detectados y puestos con relieve en negro.

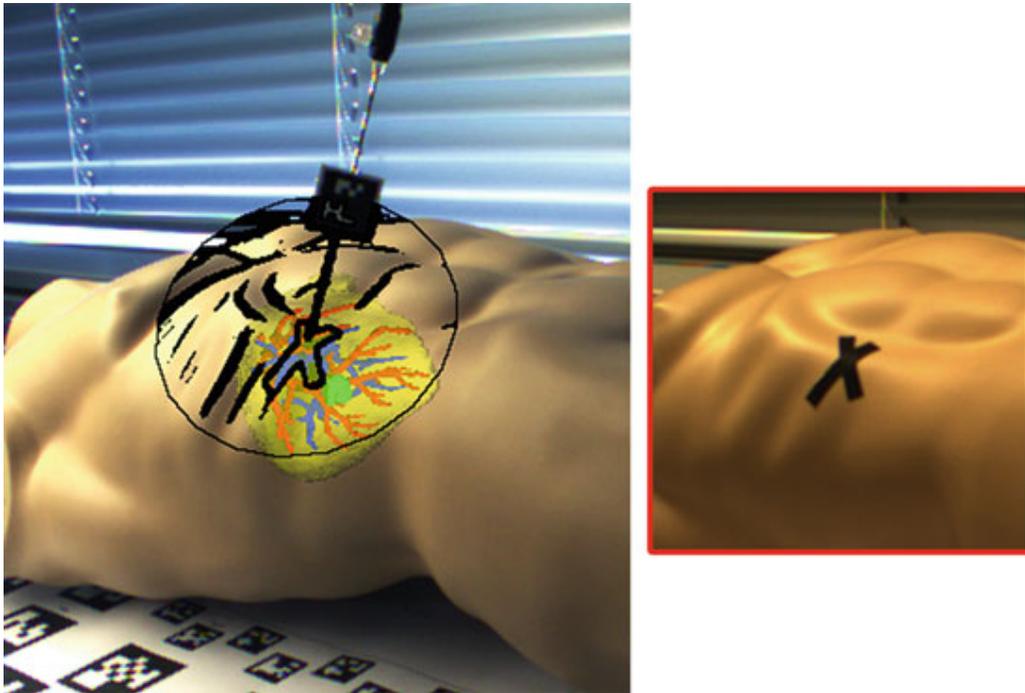


Figura 3.17: Imagen secundaria basada en imágenes usando la preservación de bordes. La preservación de bordes en toda imagen estorba en su presentación. Una lente plana mágica interactiva permite controlar la saturación en el efecto fantasma del video. Los bordes se han conservado por los operadores en el espacio imagen 2D solamente. Cada trama del video se procesa y los bordes discretos se detectan usando un detector de borde ordinario.



Figura 3.18: El sistema de realidad aumentada utiliza una reconstrucción de textura trabadas adquiridas de una escena a distancia (centro) a partir de imágenes capturadas y los modelos derivados. La reconstrucción se superpone con textura de vista del usuario del medio ambiente (a la izquierda). En combinación con el resaltado de los bordes de la oclusión de la estructura, únicamente cuando los elementos ocultos pueden ser revelados, se consigue una visualización por rayos x comprensible, mientras que al mismo tiempo se evita el desorden de borde (derecha).

3.4.2.5. Características sobresalientes

En Furht (2011) se menciona que la preservación de los bordes de los elementos de oclusión ayuda a mantener contexto y mejora la percepción espacial. Sin embargo, la información visual importante sobre los elementos de oclusión todavía puede perderse. Por ejemplo, los contrastes en las características visuales de la imagen, incluyendo el color, la luminosidad, la orientación y el movimiento determinan regiones prominentes y por lo tanto, debe ser conservado también. Las regiones prominentes pueden ser entendidas como las regiones de una imagen que son más propensas a atraer la mirada del espectador Liarokapis et al. (2005). A continuación se describen tres características destacadas: matiz, luminosidad y movimiento. El objetivo de este enfoque es proporcionar a los usuarios información más completa sobre la estructura de oclusión.

El modelo de prominencia presentado se basa en el enfoque de Walther ARToolKit. Las propiedades sensoriales del ojo humano son reconocidas para formar una jerarquía de células receptoras que responden al contraste entre los diferentes niveles para identificar las regiones que se destacan de su entorno. Esta jerarquía se modela mediante submuestreo, una imagen de entrada I en una pirámide diádica de $\sigma = [0..8]$, de manera que la resolución de nivel σ es $1/2^\sigma$ de la resolución de la imagen original. A partir de esta pirámide de

imágenes, P_σ , extraemos las características visuales de luminosidad l , el tono opuesto del color c , y el movimiento t .

Mientras que el movimiento se define como los cambios observados en el canal de luminosidad sobre el tiempo, la luminosidad es el brillo del componente de color, y se define como:

$$M_l = \frac{r + g + b}{3}$$

El tono opuesto del color imita la capacidad del sistema visual para distinguir tonos opuestos de color. La iluminación independiente *rojo-verde* (rg) y *azul-amarillo* (by) opuestos en el mapeo son definidos como sigue antes del mapeo M_{rg} y M_{by} son combinados en un solo mapeo M_c .

$$M_{rg} = \frac{r - g}{\max(r, g, b)} \quad M_{by} = \frac{b - \min(r, g)}{\max(r, g, b)}$$

Los contrastes son modelados en las pirámides de características diádicas a través de la escala de substracción \ominus entre los niveles escalados finos y gruesos de la pirámide. Para cada una de las características, un conjunto de mapeos de características se genera de la siguiente manera:

$$F_{f,p,s} = P_p,$$

donde f representa las característica visual $f \in \{l, c, m\}$. p y s refieren a los niveles de la pirámide y son aplicados como $p \in \{2, 3, 4\}$, $s = p + S$, y $S \in \{3, 4\}$. Los mapeos de características son combinados usando adición a través de la escala \oplus para producir mapas de visibilidad C . Finalmente, todos los mapeos de visibilidad son combinados para formar el mapeo sobresaliente S .

$$C = \bigoplus_{p=2}^4 \bigoplus_{s=p+3}^{p+4} F_{p,s} \quad S = \frac{1}{3} \sum_{k \in \{l, c, t\}} C_k$$

En la siguiente etapa, las regiones ocluidas y oclusores tienen que estar compuestos utilizando la información de la prominencia para crear la visualización final de rayos x. La Figura 3.19 Furht (2011) muestra la composición de los diferentes mapas en la visualización final de rayos x. Los mapas de

prominencia S_o y S_d se generan tanto para las imágenes ocluser I_o y ocluidas I_d respectivamente. Además de esto, destacamos bordes en el ocluser para enfatizar la estructura. Un mapa de borde E se genera a partir de la región de oclusión y se pondera con el mapa del ocluser de prominencia.

$$E = \gamma(I_o) \times S_o \times \varepsilon$$

Donde γ es una función de borde Sobel y ε es una constante de ponderación. Este mapa de borde se combina con el mapa ocluser de salida como una adición, $S_{o'} = S_o + E$. Así, se combinan $S_{o'}$ y S_d para crear el mapa de salida final que indica la transparencia del ocluser. Suponemos que las regiones más destacadas del ocluser deben tener prioridad sobre las regiones más destacadas de la oclusión. Además, se genera una máscara M y una máscara inversa M' para revelar sólo la parte de la región ocluida que nos interesa. Dado lo anterior, la imagen final se compone como sigue:

$$I_c = S_{o'} \times M + P_o \times M + P_d \times M'$$

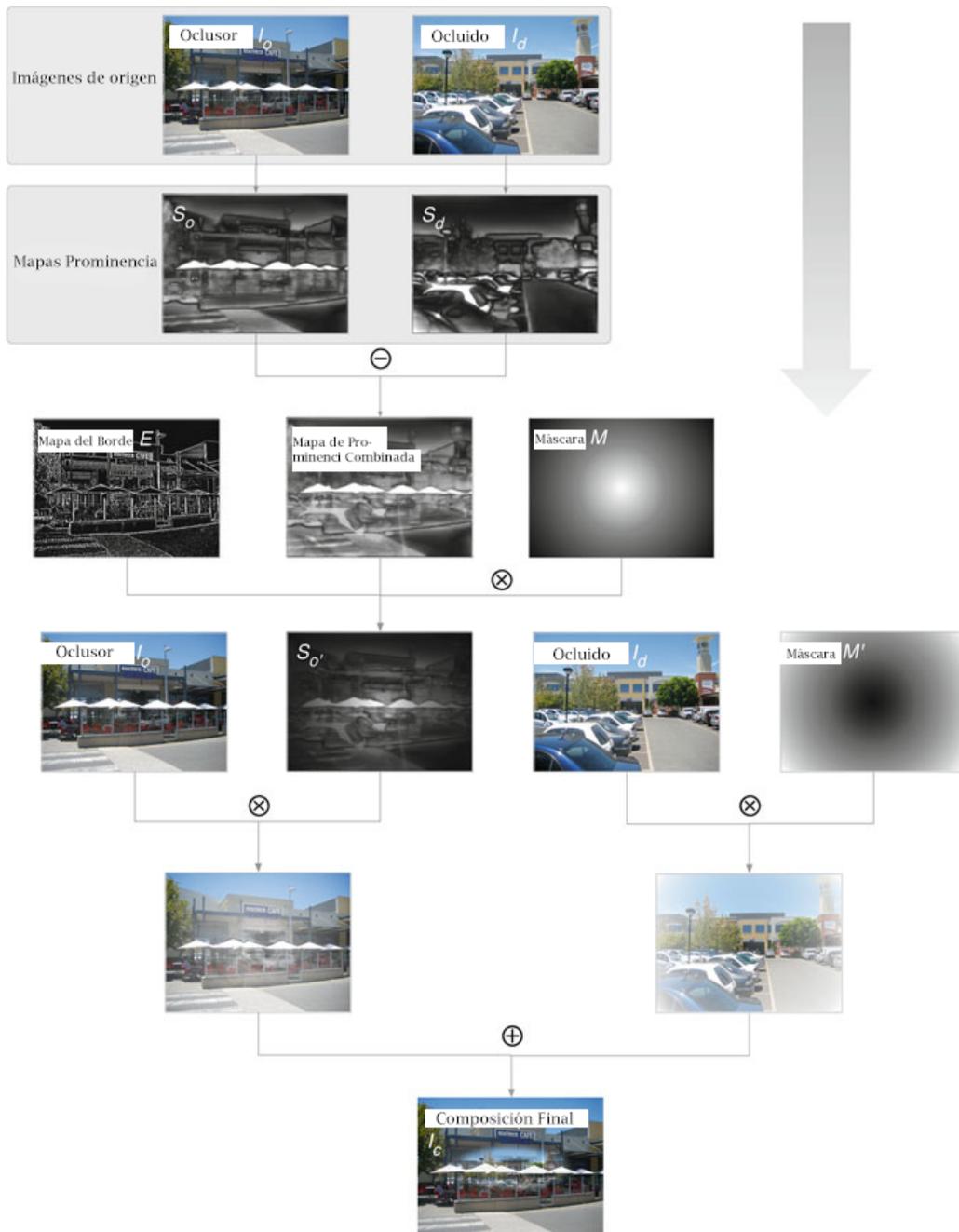


Figura 3.19: Composición: Las imágenes de origen se procesan a través de una serie de filtros y combinaciones para producir la imagen de salida final. \oplus , \ominus y \otimes denotan la suma, resta y multiplicación de los valores de los píxeles.

3.5. Appunta

Appunta Martínez (2012) es un marco de trabajo para la plataforma Android, el cual no solamente permite mostrar fácilmente la información de la posición geoespacial para el usuario, sino además auxilia a crear nuevas formas para mostrar esta información o modificar la existente.

Básicamente, se tiene un conjunto de POIs (Points Of Interest por sus siglas en inglés), ubicado en un mapa (por lo tanto, con una latitud, longitud y opcionalmente una altitud) se tienen que mostrar estos POIs y la información asociada con el usuario.

Appunta permite fuera de la caja, representar esta información de dos maneras diferentes: un radar y una vista de realidad aumentada. Sin embargo, puede modificar estos componentes para mostrar los datos de otras maneras o crear nuevas formas de visualizar esta información.

3.5.1. Appunta estructura gráfica general

En Martínez (2012) se presenta una figura que permite apreciar cómo es el funcionamiento de Appunta:

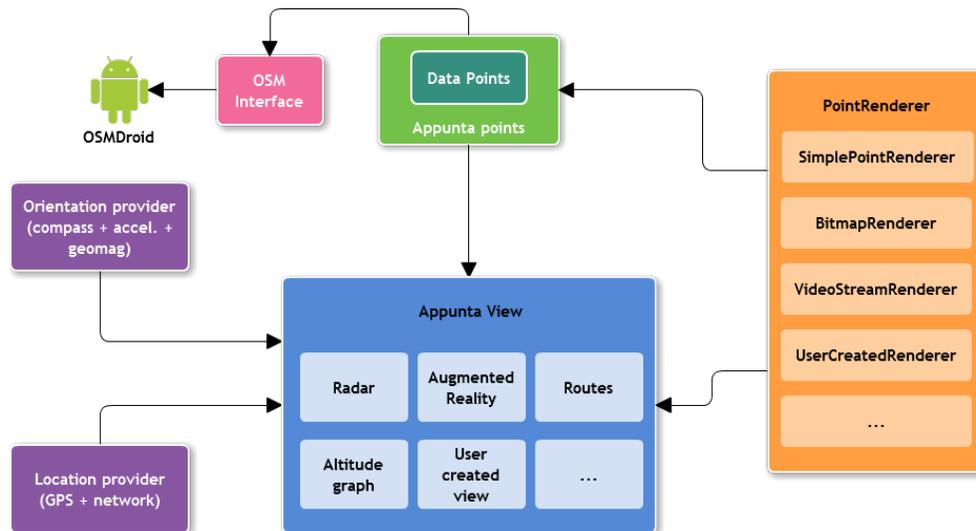


Figura 3.20: Gráfica que muestra el funcionamiento de Appunta.

3.6. Sistema de recomendación

Bobadilla et al. Bobadilla et al. (2013) define que los sistemas de recomendación han tenido un desarrollo paralelo con la red. Estos sistemas inicialmente se basaban en el filtrado de contenido demográfico y colaborativo. Sin embargo, en un futuro cercano, se utilizará información implícita, local y personal de Internet.

En la Figura 3.21 se representa en la parte más significativa del grafo, las distancias cortas entre palabras, las cuales indican los valores de similitudes más altas. Los colores cálidos indican una mayor fiabilidad para las relaciones. El tamaño de los nodos indica la importancia de las palabras en función de los parámetros N^k , N^t , N^a (número de palabras significativas en las palabras “*keywords*”, “*title*” y “*abstract*”) y N_w^k , N_w^t , N_w^a (número de veces que la palabra w aparece en “*keywords*”, “*title*” y “*abstract*”).

La ecuación utilizada para determinar la importancia de cada palabra w es la siguiente:

$$fw = \frac{1}{3} \left(\frac{N_w^k}{N^k} + \frac{N_w^t}{N^t \log \frac{N^a}{N^t}} + \frac{N_w^a}{N^a \frac{N^a}{N^t}} \right)$$



Figura 3.22: Componentes generales del proceso tradicional de recomendación.

Los sistemas de recomendación son construídos con base en el conocimiento de *preferencias parciales del usuario*. En contraste, los sistemas de recomendación basados en el contexto se componen del conocimiento de las preferencias contextuales parciales del usuario, y por lo general se ocupan de los registros de datos de la forma $\langle \text{usuario}, \text{artículo}, \text{contexto} \rangle$, donde cada registro específico incluye no solo la cantidad de un determinado usuario y el gusto por un artículo específico, sino también la información contextual en la que el artículo fue consumido por este usuario.

Además de la información descriptiva sobre los usuarios, artículos y clasificaciones, los sistemas de recomendación basados en contexto pueden hacer uso de atributos contextuales adicionales, tales como las jerarquías de contexto.

En la presencia de información contextual disponible, siguiendo los diagramas de la Figura 3.23 se toman los datos que tienen la forma $U \times I \times C \times R$, donde C es la dimensión contextual adicional y termina con una lista de recomendaciones contextuales $i_1, i_2, i_3 \dots i_n$ para cada usuario.

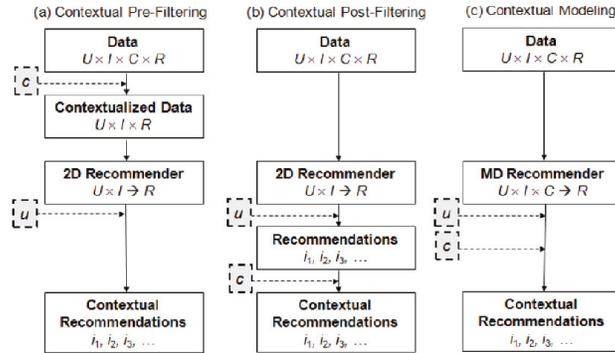


Figura 3.23: Paradigmas para la incorporación de contexto en los sistemas de recomendación.

- Pre-filtrado contextual:** En este paradigma de recomendación (presentado en la Figura 3.23a), la información contextual impulsa los datos de selección o de construcción hacia un contexto específico. En otras palabras, la información sobre el actual contexto C se utiliza para seleccionar o construir un conjunto relevante de registros (calificaciones). Las calificaciones se pueden predecir utilizando cualquier sistema de recomendación 2D tradicional sobre los datos seleccionados.
- Pos-filtrado contextual:** En este paradigma de recomendación (presentado en la Figura 3.23b), la información contextual es inicialmente ignorada y las calificaciones son previstas usando cualquier sistema de recomendación 2D tradicional en todos los datos. Entonces, el conjunto resultante de recomendaciones se ajusta (contextualiza) para cada usuario, utilizando la información contextual.
- Modelado contextual:** En este paradigma de recomendación (presentado en la Figura 3.23c), la información contextual se utiliza directamente en la técnica de modelado como parte de la estimación de calificaciones.

3.6.2. Sistema de recomendación demográfico

En Ricci (2011) se menciona que en este tipo de sistema recomienda elementos con base en el perfil demográfico del usuario. El supuesto es que diferentes recomendaciones se deben generar para diferentes nichos demográficos.

Muchos sitios web adoptan soluciones sencillas y eficaces de personalización basados en la demografía.

En Al-Shamri (2016) menciona que es un sistema estereotípico, ya que categoriza a los usuarios sobre la base de sus atributos demográficos. Posteriormente, Los sistemas de recomendación demográfico utiliza las opiniones de los usuarios para los elementos del sistema como base para las recomendaciones. Vale la pena señalar que tanto los sistemas de recomendación demográfico como los sistemas de recomendación colaborativo utilizan correlaciones de usuario a usuario pero basadas en datos diferentes. Por lo tanto, las ventajas de los sistemas de recomendación demográfico son casi similares a las de los sistemas de recomendación colaborativo en términos de su capacidad única en la identificación de nichos de género cruzado, atraer a los usuarios a saltar fuera de lo familiar, y su capacidad para mejorar a sí mismos con el tiempo.

En Al-Shamri (2016) se menciona que formalmente, el sistema de recomendación demográfico tiene M usuarios $U = \{u_1, \dots, u_M\}$ con N atributos demográficos, $D = \{a_1, \dots, a_N\}$. Por lo general, el sistema de recomendación demográfico recoge los atributos demográficos durante el proceso de registro utilizando un cuestionario acerca de los datos demográficos del usuario y las características del usuario. A través de la interacción con el sistema, se le pide al usuario que explicita o implícitamente califique los elementos de K , $S\{s_1, \dots, s_K\}$, como noticias, páginas Web, libros, películas o CDs. Inicialmente, cada usuario u_i puede clasificar un subconjunto de elementos S_i . La clasificación declarada si se encuentra disponible del usuario u_c para un elemento s_k se denota por $r_{c,k}$.

Después de la construcción del perfil de usuario, el sistema de recomendación demográfico calcula el valor de similitud entre el usuario activo actual y los usuarios de entrenamiento restantes usando una medida de similitud adecuada. Este valor indica la estrecha relación entre los dos usuarios en consideración. En consecuencia, se selecciona un conjunto de vecinos para este usuario activo de la lista clasificada de los usuarios de formación. Después de que el sistema de recomendación demográfico asigna una clasificación prevista a todos los elementos vistos por el conjunto de vecindad y no por el usuario activo. La clasificación prevista, $pr_{x,k'}$ indica la interesabilidad esperada del elemento s_k para el usuario \mathbf{u}_x . La clasificación prevista, $pr_{x,k'}$, se calcula normalmente como un agregado de las clasificaciones de $\mathbf{u}_{x,S}$ vecinales establecidas para el mismo elemento S_k :

$$pr_{x,k} = \frac{\sum_{\mathbf{u}_y \in N_x} sim(\mathbf{u}_x, \mathbf{u}_y) \times r_{y,k}}{\sum_{\mathbf{u}_y \in N_x} sim(\mathbf{u}_x, \mathbf{u}_y)}$$

donde N_x denota el conjunto de vecinos para \mathbf{u}_x que han clasificado el elemento S_k . El sistema de recomendación demográfico no requiere una lista de calificaciones para el perfil de usuario que son requeridas por otros sistemas recomendadores.

3.6.3. Sistema de recomendación basado en conocimiento

En Ricci (2011) se menciona que los sistemas basados en el conocimiento recomiendan artículos basados en el conocimiento del dominio específico acerca de cómo ciertas características de elementos satisfacen las necesidades y preferencias de los usuarios y en última instancia, cómo el artículo es útil para el usuario.

En estos sistemas una función de similitud estima cuánto las necesidades de los usuarios coinciden con las recomendaciones. Aquí, la puntuación de similitud puede interpretarse directamente como la utilidad de la recomendación para el usuario.

Los sistemas basados en conocimiento tienden a funcionar mejor que otros al inicio de su despliegue, pero si no están equipados con componentes de aprendizaje pueden ser superados por otros métodos poco profundos que pueden explotar los registros de la interacción humano / máquina.

En Baizal et al. (2016) menciona que Construir una base de conocimiento es un aspecto importante en el desarrollo del sistema. Los pasos en el proceso de la construcción del conocimiento se ilustran en Figura 3.24

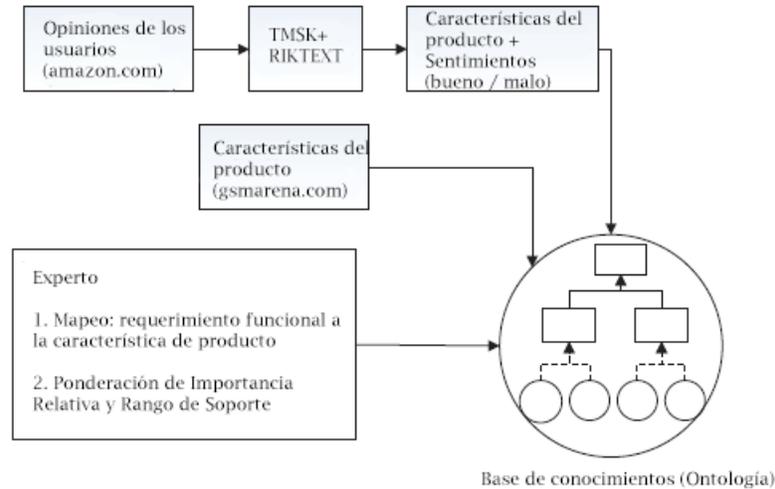


Figura 3.24: Construcción del conocimiento

La Figura 3.24 muestra el proceso de la construcción de la base del conocimiento.

Tres elementos esenciales que construyen el conocimiento son los siguientes:

- a) *Información relacionada con las características del producto*: Información acerca de las características de los productos son obtenidas de gsmarena.com. Esta información es nombre del producto, marca, precio, especificaciones, sistemas operativos, etc.
- b) *Información relacionada con las características del producto y sus sentimientos*: El sentimiento de las características del producto se obtiene de la extracción de revisiones que llevaron a cabo TMSK y RIKTEXT.

La revisión es tomada de www.amazon.com. Además, la revisión (texto) se convierte en XML, como entrada para TMSK. Entonces, los resultados del proceso son sentencias de diccionario. Este diccionario se procesa en RIKTEXT para la clasificación de sentimientos. El resultado final son las palabras que comprenden las especificaciones del producto (característi-

cas del producto) que ya están clasificadas con base en los sentimientos "bueno" "malo".

- c) *Información relacionada con el mapeo de requisitos funcionales con las características del producto:* Para el mapeo y ponderación de las características del producto, se necesita información del experto del dominio para compartir algún nivel de característica del producto existente. Cada nivel de característica (grados de especificación) se pondrá de acuerdo con su nivel de soporte para cada requisito funcional.

3.6.4. Sistema de recomendación colaborativo

En Ricci (2011) se menciona que en este tipo de sistema recomienda elementos en función de las preferencias de los amigos de los usuarios. Esta técnica sigue el epigrama “Dime con quién andas y te diré quien eres”. La evidencia sugiere que las personas tienden a confiar más en las recomendaciones de sus amigos que en las recomendaciones de individuos similares pero anónimas.

Esta observación, junto con la creciente popularidad de las redes sociales abiertas, está generando un creciente interés en los sistemas basados en la comunidad.

La recomendación se basa en las calificaciones que fueron proporcionadas por los amigos del usuario. De hecho, estos sistemas de recomendación van de la mano con el auge de las redes sociales y permiten una adquisición sencilla y completa de los datos basados en las relaciones sociales de los usuarios.

En el enfoque de las películas Fatemi & Tokarchuk (2013) menciona que se recomiendan en función de su pertenencia a la comunidad, el valor de utilidad del grado, los nodos adyacentes y el ranking de estrellas. El valor de utilidad del rank determina cuán importante o popular es un nodo en una red, mientras que los nodos adyacentes muestran cómo coinciden las dos películas en términos de interés de los revisores y la fuerza de su vinculación está determinada por el peso de su enlace (Número de revisores comunes). El grado de un nodo también muestra el número de revisiones que recibió y por lo tanto su popularidad e importancia. La clasificación por estrellas también

indica lo buena que es una película según el punto de vista de los críticos. Por lo tanto, cuando el Sistema de recomendación colaborativo está generando las recomendaciones, considera los siguientes factores: La popularidad de una película dentro de la comunidad (valor de utilidad), la coincidencia más cercana de los intereses del crítico de la película (nodos adyacentes) y la calidad general de una película (clasificación en estrella).

Membresía en la comunidad: se define como un conjunto de todas las comunidades en las que un nodo es miembro:

$$L_c(m) = \{C_1, \dots, C_{n_m}\}$$

Un nodo puede ser un miembro de más de una comunidad. La comunidad principal $C_M(m)$ es definido como la comunidad en la cual la película m tiene la suma agregada más alta de pesos de borde y grados:

$$C_M(m) = \forall (c_i \in L_c(m) \& x \in c_i) : \\ \max_{1 \leq i \leq n_m} (edge(x_{c_i}) + degree(x_{c_i}))$$

Conjunto de grado de utilidad: Es la lista ordenada de películas basada en el grado de sus valores de utilidad. El grado de valor de utilidad se define como el grado de los nodos en la comunidad a los grados totales de todos los nodos en la red:

$$u(x) = \frac{degree(x)}{\sum_{j=1}^N degree(x_j)}$$

N : número total de nodos en la red

Cuanto más alto sea el valor de utilidad, más importante será el nodo. Y lista de utilidad de grado para la comunidad C_i , $L_U(C_i)$ es la lista de las películas x_j en la comunidad C_i , clasificados según sus utilidades de grado $u(x_j)$ como:

$$L_U(C_i) = \{x_{1_u}, \dots, X_{n_u}\}, \text{ Donde : } x_j \in C_i$$

Conjunto de adyacentes clasificado: se define como un conjunto de nodos que están vinculados por un borde dentro de la misma comunidad. La clasificación está determinada por el peso del borde que los conecta.

$$L_A(m) = \{X_{1_a}, \dots, X_{n_a}\},$$

$$\text{Donde : } (m \& x_j \in C_i) \& (\text{Edge}(m, x_j) = 1)$$

La adyacencia de nodos de una película es importante, ya que indica que las dos películas conectadas son del mismo interés para los revisores. Por lo tanto, cuanto mayor sea el peso, mayor será la fuerza de conexión entre los dos nodos.

Conjunto de clasificación estrella: Es la lista ordenada de películas dentro de la misma comunidad C_i basada en la clasificación media de usuarios $r(x_j)$ para cada película. Se utiliza para decidir qué película tiene una mayor calidad y popularidad según el punto de vista de los revisores.

$$L_R(C_i) = \{x_1, \dots, x_{n_r}\}, \text{ Donde : } x_j \in C_i$$

Es lógico recomendar los nodos adyacentes más otros nodos prominentes por el mérito de su valor de grado de utilidad alto en la comunidad de la película m específica. Las recomendaciones finales, M es un conjunto de 10 películas. Se compone de películas en el conjunto M_1 y M_2 , donde M_1 es el top 10 películas comunes de todas las listas y M_2 está vacío a menos que el número de películas en M_1 es inferior a 10, cuando M_2 este poblado. Por la simplicidad se decide elegir un número par de películas de cada lista, para asegurarse de que un conjunto diverso de películas se elija para la recomendación.

$$M = M_1 \cup M_2$$

$$M_1(x_j) = \{\text{Top}_{10}x_j(L_U(C_M(m)) \cap L_A(m) \cap L_R(C_M(m)))\}$$

$$M_2 = \{\text{Top}_z x_j(L_U(C_M(m))) \cup \text{Top}_z x_j(L_A(m)) \cap \text{Top}_z x_j(L_R(C_M(m)))\},$$

$$\text{donde : } z = (10 - \text{Len}(M_1)/3) \& (x_j \notin M_1)$$

Como cada conjunto representa una dimensión, (popularidad, mayor interés y calidad) eligiendo de diferentes números de películas de cada conjunto que dominan las recomendaciones hacia las características asociadas con esa lista. Sin embargo, si los usuarios prefieren recomendaciones más diversas, incluyendo películas menos populares a las recomendaciones más cercanas a su interés, sin diversidad en los elementos o simplemente prefieren la calidad a la popularidad, es un estudio que está muy relacionado con el gusto de los usuarios.

3.6.5. Sistemas de recomendación híbridos

En Ricci (2011) se menciona que estos sistemas de recomendación se basan en la combinación de las técnicas antes mencionadas. Un sistema híbrido que combina técnicas de A y B intenta utilizar las ventajas de A para solucionar los inconvenientes de B. Por ejemplo, en los artículos que no tienen calificaciones, esto no limita los enfoques basados en contenidos de predicción para nuevos artículos, sino que se basa en su descripción que se encuentra fácilmente disponible.

El contexto del usuario, cuando se está buscando una recomendación se puede utilizar para personalizar mejor la salida del sistema.

Por tanto, se discuten tres paradigmas algorítmicos diferentes para la incorporación de la información contextual en el proceso de recomendación: con base en la reducción (pre-filtrado), filtrado contextual de correos y el modelado de contexto.

En filtrado contextual posterior, el algoritmo de la recomendación ignora la información de contexto. La salida del algoritmo se filtra o ajusta para incluir sólo las recomendaciones que son relevantes en el contexto de destino. En el modelado contextual, el más sofisticado de los tres enfoques, los datos del contexto se utilizan de manera explícita en el modelo de predicción.

En Burke (2002) menciona que las técnicas de recomendación tienen una serie de posibles clasificaciones (Resnick y Varian, 1997, Schafer, Konstan & Riedl, 1999, Terveen y Hill, 2001). De interés en esta discusión no es el tipo de

interfaz o las propiedades de la interacción del usuario con el recomendador, sino más bien las fuentes de datos sobre las que se basa la recomendación y el uso al que se aplican dichos datos. Específicamente, los sistemas recomendadores tienen (i) datos de antecedentes, la información que el sistema tiene antes de que comience el proceso de recomendación, (ii) datos de entrada, la información que el usuario debe comunicar al sistema para generar una recomendación, y (iii) Un algoritmo que combina datos de fondo y de entrada para llegar a sus sugerencias.

Sobre esta base, podemos distinguir cinco técnicas de recomendación diferentes, como se muestra en la Tabla 3.1. Supongamos que \mathbf{I} es el conjunto de elementos sobre los que se pueden hacer recomendaciones, \mathbf{U} es el conjunto de usuarios cuyas preferencias son conocidas, \mathbf{u} es el usuario para quien las recomendaciones Necesidad de ser generado, e \mathbf{i} es un elemento para el que nos gustaría predecir la preferencia de \mathbf{u} .

Tabla 3.1: Técnicas de Recomendación

Técnica	Fondo	Entrada	Proceso
Colaborativo	Valoraciones de \mathbf{U} de elementos en \mathbf{I} .	Valoraciones de \mathbf{u} de elementos en \mathbf{I} .	Identificar usuarios en \mathbf{U} similar a \mathbf{u} , y extrapolar de sus calificaciones de \mathbf{i} .
Basado en Contenido	Características de elementos en \mathbf{I}	calificaciones de \mathbf{u} de los elementos en \mathbf{I}	Genera un clasificador que se ajusta al comportamiento de rating de \mathbf{u} y lo usa en \mathbf{i} .
Demográfico	Información demográfica sobre \mathbf{U} y sus calificaciones de elementos en \mathbf{I} .	Información demográfica sobre \mathbf{u}	Identificar usuarios que son demográficamente similares a \mathbf{u} , y extrapolar de sus calificaciones de \mathbf{i} .
Basado en la utilidad	Características de los elementos en \mathbf{I}	Una función de utilidad sobre elementos en \mathbf{I} que describe las preferencias de \mathbf{u} .	Aplicar la función a los elementos y determinar el rango de \mathbf{i}
Basado en Conocimiento	Características de los elementos en \mathbf{I} . Conocimiento de cómo estos elementos satisfacen las necesidades de un usuario.	Una descripción de las necesidades o intereses de \mathbf{u}	Inferir una coincidencia entre la necesidad de \mathbf{i} y de \mathbf{u} .

3.7. PHP

En php se menciona que PHP, es el acrónimo de “PHP: Hypertext Pre-processor”. Éste es un lenguaje de ‘scripting’ de propósito general y de código abierto que está especialmente pensado para el desarrollo web y que puede ser embebido en páginas HTML. Su sintaxis recurre a C, Java y Perl, siendo así sencillo de aprender. El objetivo principal de este lenguaje es permitir a los desarrolladores web escribir dinámica y rápidamente páginas web generadas; aunque se puede hacer mucho más con PHP.

3.7.1. Laravel

Se eligió Laravel como marco de trabajo, porque utiliza la siguiente filosofía: Laravel es un framework de aplicaciones web con la sintaxis expresiva y elegante. El desarrollo debe ser una experiencia agradable y creativa para ser verdaderamente satisfactoria. Laravel intenta eliminar las molestias del desarrollo, facilitando las tareas comunes que se utilizan en la mayoría de los proyectos web, como la autenticación, enrutamiento, sesiones y el almacenamiento en caché Laravel.

Laravel apunta hacer el proceso de desarrollo placentero para los desarrolladores, sin sacrificar la funcionalidad de la aplicación. Con este fin, han intentado combinar lo mejor de lo que han visto en otros frameworks web, incluidos los marcos implementados en otros idiomas, como el Ruby on Rails, ASP.NET MVC, y Sinatra.

Laravel es accesible y poderoso, además proporciona potentes herramientas necesarias para aplicaciones grandes y robustas. Asimismo, proporciona una excelente inversión de contenedor de control, sistema de migración expresiva y el soporte estrechamente integrado de las pruebas unitarias. También brinda las herramientas necesarias para construir cualquier aplicación que sea requerida.

3.8. Android

Android es el sistema operativo más utilizado en dispositivos móviles, un dato curioso de Android es que las versiones son nombradas como dulces

y éstas coinciden con el abecedario iniciando desde la “D” (Donut, Eclair, Froyo, Gingerbread, Honeycomb, Ice Cream Sandwich, Jelly Bean, KitKat, Lollipop [Versión actual al momento de la escritura de esta tesis]) android.

Esta sistema operativo fue elegido como plataforma para el desarrollo, utilizando un lenguaje de programación *open – source*, por su gran compatibilidad con diferentes dispositivos móviles, a continuación se detalla un poco de los orígenes de Android y la versión que se utiliza para esta metodología.

Los orígenes sobre este sistema operativo remontan a los inicios de la década anterior, a través del trabajo del fundador Andy Rubin, y Google androidcentral.

3.8.1. Android - Lollipop

En android se mencionan los principales cambios en esta versión, los cuales son “Material Design”, “Multiscreen”, “Notifications”.



Figura 3.25: Material Design, Multiscreen, Notifications.

3.8.1.1. Material Design

Android tiene un nuevo estilo y es más ligero para navegar en el dispositivo, ya que se basa en sombras y movimientos. Material Design une los

principios básicos de buen diseño con innovación de lo que es posible a través de la tecnología. android.

3.8.1.2. Multiscreen

La versión Lollipop permite movimientos en pantalla sin problemas en dispositivos móviles, tales como, teléfonos, tabletas, relojes inteligentes con Android, o Android TV android.

Desde que Lollipop funciona en todos los dispositivos, es posible continuar desde donde se quedó una canción, navegación de fotos y búsquedas recientes.

3.8.1.3. Notificaciones

Las notificaciones para esta versión se encuentran ahora en la pantalla, aunque ésta se encuentre bloqueada, donde se despliegan de manera limpia en forma de tarjetas segmentadas android.

3.8.1.4. Android 5.0 Cambios del comportamiento (API Level 21)

En Android se mencionan los cambios como reemplazar Dalvik como plataforma predeterminada, por Android Runtime (ART), la cual tiene mejoras en la compilación, en el recolector de basura entre otras cosas. Además, se utilizan los estilos de Material Design, se modificó la forma de mostrar notificaciones con sonido y vibración, se puede agregar un nivel de visibilidad (en cuestión de privacidad) en las notificaciones, se hizo obsoleto el método *getRecentTasks()* para sustituirlo por el método *getAppTasks()*, con la finalidad de aumentar el control en la privacidad. Android 5.0 introduce el soporte para sistemas basados en plataformas de 64 bits, entre otras mejoras relacionadas con la privacidad.

3.9. Visualización de los datos en web

La visualización de los datos geoespaciales utiliza un tipo especial de datos que especifican la ubicación de un objeto o fenómenos Ward et al. (2010). Por lo general, esto es posible gracias a la información de latitud y longitud en cada registro. Ejemplos de datos geoespaciales son modelos globales del

clima, los registros ambientales, medidas económicas y sociales y los indicadores para el análisis de clientes y datos sobre la delincuencia. La estrategia utilizada para representar este tipo de datos es mapear los atributos espaciales directamente a las dos dimensiones de la pantalla física, lo que resulta en el mapa visualizaciones Ledur et al. (2015).

En la creación de una visualización de datos, es importante saber acerca de la entrada, estructura y tipo de datos que estamos manejando. Cuando estos datos vienen de una fuente externa, es necesario llevar a cabo una preparación o adecuación de los mismos para la selección, el filtrado y la limpieza. Después, es posible asignar los datos en representaciones visuales, de acuerdo con los atributos seleccionados previamente. Ward et al. (2010).

Hay algunas librerías que permiten a los usuarios crear mapas de visualización de datos. Al utilizar estas librerías, como Google Maps API, OpenLayers, y Leaflet, corresponde al usuario realizar un pre-procesamiento de los datos en el formato correcto. Cuando se trata de grandes conjuntos de datos, los usuarios tendrán que consumir más tiempo para construir un mapa. Por lo general, es recomendable crear un software para la automatización del procesamiento de dichos datos. En la Figura 3.26(a) Ledur et al. (2015) se muestra el flujo de trabajo de las librerías tradicionales para generar una visualización. La línea de puntos alrededor de *la generación de la visualización y la librería del formato de datos* demuestra el alcance de estas librerías de soporte, sin la necesidad de realizar programación adicional por parte del usuario.

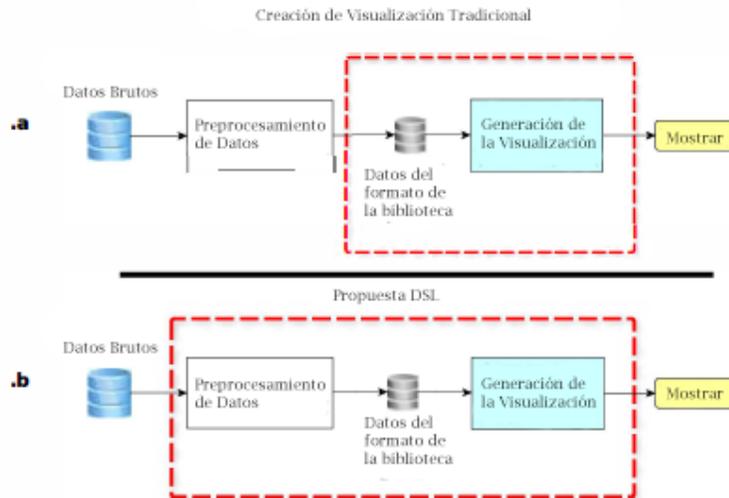


Figura 3.26: Comparación de las técnicas de visualización.

3.9.1. API de Google Maps

En Ledur et al. (2015) se describe la API que fue desarrollada por Google para revolucionar los mapas en la web, lo que permite a los usuarios arrastrar e interactuar con la visualización para encontrar la información esperada.

Google Maps API funciona utilizando HTML, CSS, JavaScript, a su vez opera en conjunto con estas tecnologías para diseñar e implementar ambientes de desarrollo. Los mosaicos de mapas son piezas de imágenes que se cargan en el fondo con las llamadas Ajax y luego se insertan en un `<div>` en la página HTML. Al navegar por el mapa, la API envía información sobre las nuevas coordenadas y los niveles de zoom del mapa en el Ajax, el cual devuelve nuevas imágenes Svennerberg (2010).

Para la creación de mapas de visualización de datos utilizando la API de Google Maps, el usuario necesita tener conocimientos en JavaScript para crear las variables, objetos y funciones de uso. Inicialmente, esto requiere la inclusión de la biblioteca en el archivo HTML y asociación de un objeto de mapa a una variable y la información privilegiada, considerando detalles del mapa como la posición inicial, zoom y la capa. Después, para cada marcador

se puede crear un límite de tres líneas de código para generar un objeto con información de latitud y longitud. Si se desea una clasificación, como para cambiar los colores de marcador, los usuarios también pueden insertar en el código JavaScript de una etiqueta para una declaración de icono.

3.9.2. OpenLayers

De acuerdo con Ledur et al. (2015), *OpenLayers* es una librería de código abierto de JavaScript que proporciona características para la visualización de datos de mapas en los navegadores web.

Además, esta tecnología proporciona una API para crear aplicaciones geográficas basadas en la web. Por otra parte, esto representa un conjunto de componentes, tales como mapas, capas o controles. OpenLayers ofrece acceso a un gran número de fuentes de datos que utilizan muchos formatos de datos diferentes, e implementa muchos estándares de Open Geospatial Consortium OpenLayers.

Además, se requiere la inserción de la librería en el archivo HTML y la creación de una librería antes de añadir un nuevo marcador. La capa es una característica de OpenLayers que permite la aplicación de diferentes tipos de visualización de datos en una única visualización de datos.

3.9.3. Leaflet

Con base en Ledur et al. (2015), Leaflet se describe como una librería de código abierto de JavaScript para la creación de mapas interactivos. Esta tecnología funciona aprovechando HTML5 y CSS3 Leaflet. También permite la creación de mapas a partir de los datos geoespaciales. A su vez, proporciona capas de mosaico, marcadores, ventanas emergentes, capas vectoriales como polilíneas, polígonos, círculos, rectángulos, capas marcadores, GeoJSON, imágenes superpuestas, capas WMS y grupos de capas.

La creación y el marcador de mapa es similar a otras librerías. Inicialmente se produce la inserción de la librería en el archivo HTML y un conjunto de códigos es responsable de crear el mapa, así como los marcadores se insertan más adelante. Una opción permite la especificación de la información dentro de los marcadores cuando se selecciona mediante un clic. Al igual que en otras

librerías, el usuario debe tener conocimiento en programación JavaScript para crear un mapa de visualización de datos y la replicación de código para cada marcador insertado.

3.10. PostgreSQL

De acuerdo con PostgreSQL, PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, distribuido bajo licencia BSD y con su código fuente disponible libremente. Es el sistema de gestión de bases de datos de código abierto más potente del mercado y en sus últimas versiones no tiene nada que envidiarle a otras bases de datos comerciales.

PostgreSQL utiliza un modelo cliente/servidor y usa multiprocesos en vez de multi-hilos para garantizar la estabilidad del sistema. Un fallo en uno de los procesos no afectará el resto y el sistema continuará funcionando. Algunas de las características de PostgreSQL son las siguientes:

Tabla 3.2: Características de PostgreSQL

Límite	Valor
Máximo tamaño base de dato	Ilimitado (Depende de tu sistema de almacenamiento)
Máximo tamaño de tabla	32 TB
Máximo tamaño de fila	1.6 TB
Máximo tamaño de campo	1 GB
Máximo número de filas por tabla	Ilimitado
Máximo número de columnas por tabla	250 - 1600 (dependiendo del tipo)
Máximo número de índices por tabla	Ilimitado

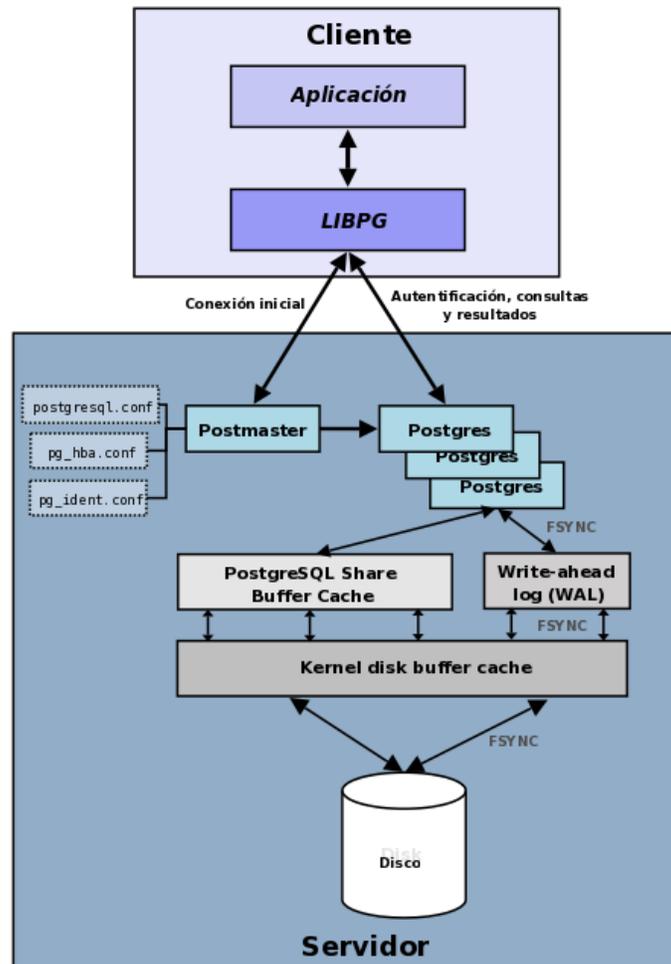


Figura 3.27: Gráfico que ilustra de manera general los componentes más importantes en un sistema PostgreSQL.

3.10.1. PostGIS

Asimismo, PostGIS menciona que PostGIS es base de datos espacial, la cual corresponde como una extensión de PostgreSQL. Esta extensión incorpora soporte para objetos geográficos, el cual permite desarrollar consultas espaciales, ejecutándose en SQL.

Capítulo 4

Metodología

En este capítulo, se describe la metodología propuesta, y de forma exhaustiva las etapas que la componen, presentando las técnicas, estándares y protocolos que fueron implementados en cada etapa.

4.1. Introducción

La metodología se divide en tres secciones principales; la personalización, en la cual se incluye un perfil de usuario y un vector característico, en el perfil de usuario se tienen todas las preferencias del usuario las cuales sirven de base para poder comparar y generar las medidas de similitud, el vector característico será utilizado para almacenar las preferencias del usuario, su longitud será la cantidad de elementos que conformen la categoría $\langle C_1, C_2, \dots, C_n \rangle$, las preferencias serán denotadas con 1 cuando sea de la preferencia del usuario y con un 0 cuando no sea de su preferencia.

La segunda sección se conforma del procesamiento semántico en el cual, por medio de una ontología que es la que incluye toda la información y por medio del filtrado que se da por medio del vector característico que se construye en el proceso de personalización, utilizando técnicas o algoritmos para poder medir la similitud semántica como por ejemplo la distancia entre los nodos de la ontología resultando más cercanos los que tengan el significado más similar.

La tercera sección se conforma de la visualización, esta es la represen-

tación de los resultados obtenidos después de la medición de la similitud semántica, podemos representar los resultados en forma de tablas por medio de una base de datos espacial, en la cual por medio de tablas tendremos los resultados de las similitudes y las ubicaciones de donde se originaron o de donde serán utilizadas, por medio de consultas, podemos recuperar la lista de los resultados ranquedados los cuales pueden ser representados tanto en dispositivos móviles (Mobile GIS) o en contenido web (Web GIS), para una representación gráfica de los mismos.

Las secciones previamente mencionadas se muestran en la Figura 4.1.

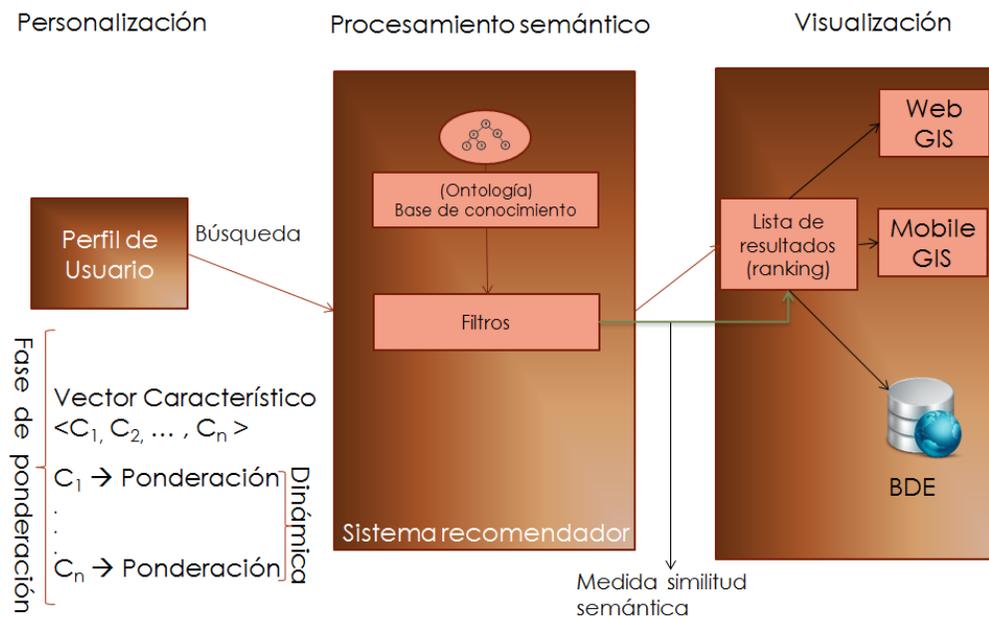


Figura 4.1: Metodología

4.1.1. Etapa de Personalización

La etapa de personalización es esta representada por el perfil de usuario el cual esta conformado por las preferencias del usuario sobre una o varias categorías predeterminadas, el perfil de usuario es uno de los componentes más importantes, ya que por medio de este el sistema podrá realizar las re-

comendaciones o filtrado del contenido, la información puede ser general, o de los intereses del usuario.

El perfil de usuario puede contener también la ubicación actual del usuario, la cual se puede utilizar para dar a conocer ubicaciones cercanas que sean de su interés.

El perfil de usuario puede ser construido del lado del cliente o del servidor, pero al estar del lado del cliente se cuenta con la personalización por ejemplo en un dispositivo móvil con el cual el usuario puede interactuar en cada momento con su perfil.

Un vector característico es el que se encarga de guardar las preferencias seleccionadas en el perfil de usuario, su longitud será la cantidad de elementos que conformen la categoría $\langle C_1, C_2, \dots, C_n \rangle$, estas son almacenadas de forma binaria (0 y 1) las preferencias serán guardadas en la posición correspondiente en el vector con 1 cuando sea de la preferencia del usuario y sera almacenada con un 0 cuando no sea de su preferencia.

4.1.1.1. Perfil de usuario

En este apartado se presenta información sobre los perfiles de usuario y lo relacionado a la construcción de los mismos.

En los sistemas de recomendación el perfil de usuario es uno de los componentes más importantes, ya que el sistema necesita la información del usuario para poder recomendar o filtrar el contenido para cada uno de los usuarios por separado. Un sistema necesita saber algo acerca de cada uno de los usuarios para poder proporcionar una experiencia a la medida del usuario.

Generalmente, el perfil de usuario se crea con base en la información que el sistema necesita o debe saber del usuario. Esta información puede ser la información general del usuario (por ejemplo, nombre, edad, sexo), el comportamiento que tiene el usuario, el contexto y los intereses.

El perfil de usuario debe contener la mayor cantidad de información posible sobre los intereses de un usuario y otra información contextual. Si el perfil de usuario contiene la ubicación actual del usuario, esto se puede uti-

lizarse para proporcionar las instalaciones cercanas. Esto podría ser mucho más relevante para el usuario que los anuncios que podrían no estar disponibles para el mismo.

Los perfiles de usuario pueden ser construídos del lado del cliente o bien, o del lado del servidor para que reciba la información acerca del usuario. Si los requisitos del sistema lo permiten almacenar el perfil de usuario del lado del cliente tiene ventajas. El perfil puede contener información confidencial sobre el usuario si estos datos estuvieran del lado del servidor y el usuario no tendría control directo sobre ellos. Una restricción para almacenar el perfil de usuario del lado del cliente podría ser la limitación del tamaño de los clientes o los requisitos computacionales con respecto a las operaciones que se necesite realizar. Por otro lado, almacenar el perfil de usuario en un servidor, ofrece mayor durabilidad de los datos y mayor velocidad y cantidad de operaciones de los mismos.

Los intereses del usuario son la información más importante en los sistemas de recomendación. Las preferencias del usuario se pueden utilizar para localizar contenidos que se adapten mejor a las necesidades del usuario (Gao et al, 2010).

El papel de la modelación del perfil de usuario y el acceso a información personalizada se ha incrementando debido a la gran cantidad de información disponible en Internet, bibliotecas digitales y otros medios. Los usuarios necesitan una atención personalizada en la búsqueda, a través de una gran cantidad de información recuperada de acuerdo con sus intereses. Los sistemas de recuperación adaptan su funcionamiento a usuarios individuales aprendiendo sus preferencias durante la interacción, con el propósito de construir un perfil de usuario que pueda ser aprovechado en el proceso de búsqueda.

4.1.1.2. Vector Característico

En (Bouneffouf, 2013) se definen las características del perfil de usuario, sus indicadores de interés y describen algunos enfoques para modelar un perfil de usuario. Estos enfoques son los siguientes:

- 1) Representación vectorial, la cual se basa en el modelo clásico de espacio vectorial, donde el perfil se representa como un vector m -dimensional,

cada dimensión corresponde a un término distinto y m es el total de términos que contiene el perfil de usuario.

- 2) Representación por conexión, se basa en una interconexión asociativa de los nodos que representan el perfil de usuario.
- 3) Representación basada en ontologías, para representar las relaciones semánticas entre unidades de información del perfil de usuario. Este modelo representa el perfil de usuario como una jerarquía de conceptos, donde cada clase en la jerarquía representa el conocimiento en un área de interés del perfil de usuario. La relación (*generalización/especificación*) entre los elementos de la ontología refleja un interés más realista del usuario.
- 4) Representación multidimensional, tiene la ventaja de proporcionar una mejor interpretación de la semántica del perfil de usuario.

Como se menciona en la Subsubsección 2.3.4.1 un perfil de usuario i denotado por \vec{U}_i es un vector de etiqueta, compuesto por parejas de valores de la siguiente manera:

$$\vec{U}_i = (t_i, 1 : v_i, 1, t_i, 2 : v_i, 2, \dots, t_i, n : v_i, n), \forall x, t_i, x \in \vec{T}'_i \cup \vec{T}''_i$$

Donde t_i, x es una etiqueta que puede ser interesante para un usuario i , v_i, x es el grado de preferencia del usuario i en la etiqueta t_i, x y n es el número de etiquetas en que el usuario i puede estar interesado.

4.1.2. Etapa de Procesamiento Semántico

4.1.2.1. Ontología

Como se mencionó en la Sección 3.1 el proceso de crear una ontología se divide en tres tareas:

- Captura
- Codificación
- Integración de las ontologías existentes

En Arano (2004) menciona que una ontología es una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son explicitadas mediante formalismos en un determinado dominio. Su función más frecuente en la Lingüística es de apoyo para sistemas de Traducción Automática Basada en el Conocimiento y para la Terminografía (práctica de la Terminología). En ambos casos, la ontología es una representación formal y explícita de la estructura conceptual del campo sobre el que se trabaja. Este recurso lingüístico incluye como mecanismo de inferencia a la herencia, que implica una economía en la codificación de la información: los conceptos superiores transmiten sus características a los conceptos inferiores.

Por consiguiente, la ontología es uno de los módulos asociados a un sistema de conocimiento donde su función es la de apoyo semántico para las unidades léxicas, es decir que las unidades léxicas son descritas como objetos lingüísticos en una base de datos léxica y son relacionadas con una jerarquía conceptual localizada en una ontología.

Una base de conocimiento es una forma avanzada de base de datos que no solo pretende almacenar, recuperar y modificar grandes cantidades de información, sino también, plasmar elementos de conocimiento (generalmente en forma de hechos y reglas de inferencia), así como la forma en que éste ha de ser utilizado.

Este recurso lingüístico es un modelo de un mundo/empresa/sección de la realidad, como declara Mylopoulos (1982) citado por Pérez Hernández (2002), en donde se considera el mundo/universo como una colección de individuos o entidades y el conjunto de relaciones que existen entre esos individuos. La colección de individuos, más las relaciones entre éstos, constituye un estado, cuyos cambios causan la creación o modificación de individuos o de las relaciones entre ellos.

La base de conocimiento utiliza al esquema de representación como notación precisa para representar el conocimiento que contiene. El tipo de esquema de representación de mayor difusión son las redes semánticas que estructuran sus datos en nodos que simbolizan a los conceptos unidos por arcos que representan las relaciones conceptuales. La red semántica también utiliza la herencia como método de inferencia, donde los nodos inferiores heredan las características de los nodos superiores permitiendo así una economía de

codificación.

Los tres tipos tradicionales de redes semánticas son: las redes IS-A, los grafos conceptuales y las redes de marcos (frames). El esquema basado en marcos (frames) es el más explorado y utilizado por los investigadores dada su mayor flexibilidad, y en mayor o menor medida, generan las ontologías de conceptos.

En Lingüística, siguiendo a Feliu et al. (2002), Moreno Ortiz (2001), Cabré et al. (2004), la base de conocimiento se concibe como un recurso modular que integra recursos tales como bases textuales, bases bibliográficas, bases terminológicas y ontologías.

‘WordNet is an online lexical database...’ (Miller 1995: 39)

Una base de datos léxica es un sistema de almacenamiento de información lingüística organizada según un determinado modelo de datos que posibilita el almacenamiento, recuperación y modificación de los mismos. El modelo de datos puede tener una estructura jerárquica, de red o relacional, esta última la de más amplia difusión. Una base de datos léxica tiene la función de responder a consultas sobre los datos que contiene, ya sea desde prestaciones propias o a partir de aplicaciones externas, permitiendo la reutilización de la información contenida. El comportamiento de una base de datos léxica es pasivo, pues las operaciones sobre sus datos son realizadas por aplicaciones que deben ser iniciadas explícitamente.

Formalmente las bases de datos no están planeadas para guardar información compleja sino grandes volúmenes de datos, con lo cual su aplicación en la representación de la información léxica dificulta la visión de conjunto debido a la tendencia a la atomicidad de los datos almacenados. El concepto de atomicidad implica que un elemento sea atómico (o escalar), cuando no puede fragmentarse en partes más pequeñas. Por ejemplo, en la codificación de un número telefónico, si se opta por codificar la información en tres valores separados –prefijo internacional, prefijo interprovincial y número del abonado–, se posibilita una gestión impensable en caso de que todo el número se tomara como un único valor. Si el número telefónico es segmentado según los distintos prefijos que lo componen, un programa de comunicaciones que marque el número y establezca la conexión en forma automática puede

utilizarlo. Por tanto, el concepto de atomicidad no es consecuente con las características de la información léxica.

Las bases de datos léxicas son utilizadas en la Lingüística como fuentes de información léxica a reutilizar por otros recursos, por ejemplo un lexicón computacional o una base de datos terminológica.

4.1.2.2. WordNet

WordNet lexical database for English es una base de datos léxica diseñada sobre la base de las teorías psicolingüísticas acerca del lexicón mental, cuya finalidad era agilizar las búsquedas de los diccionarios en línea para la lengua inglesa.

Esta base de datos léxica se construye sobre la base de las categorías sintácticas de nombre, verbo, adjetivo y adverbio. Dichas categorías se organizan en distintas estructuras léxicas: los nombres en jerarquías léxicas sobre la base de relaciones de hiponimia y meronimia; los verbos en base a relaciones de implicación (entailment), y finalmente, los adjetivos y adverbios se organizan como hiperespacios N-dimensionales. Sin embargo, este tipo de organización produce una redundancia de información en los casos en que una unidad léxica pertenece a más de una categoría.

WordNet se basa en el supuesto teórico de matriz léxica (lexical matrix) integrada por los elementos de forma léxica (word form) y significado léxico (word meaning) que se corresponden con la expresión física que se escribe o pronuncia, por un lado, y con el concepto lexicalizado que se expresa por medio de una forma léxica, por otro.

Ilustrando el Concepto de una Matriz Léxica:
 F_1 y F_2 son sinónimos y F_2 es Polisémico

Significados Palabra	Formas de la palabra			
	F_1	F_2	F_3	... F_n
M_1	$E_{1,1}$	$E_{1,2}$		
M_2		$E_{2,2}$		
M_3			$E_{3,3}$	
\vdots				
M_m				$E_{m,n}$

Figura 4.2: Ilustrando el concepto de una matriz léxica

En la matriz léxica el encabezamiento de las columnas (F_1) corresponde a las unidades léxicas de una lengua y el encabezamiento de las filas (M_1) a los conceptos. Una entrada en una celda de la matriz ($E_{1,1}$) implica que esa forma (F_1) puede ser utilizada para expresar el concepto (M_1). Esta presentación en columnas y filas permite observar gráficamente dos de los principales temas de la semántica léxica: la polisemia (en caso de que la misma columna contara con dos entradas, $E_{1,2}$ - $E_{2,2}$) y la sinonimia (en caso de que la misma fila contara con al menos dos entradas, $E_{1,1}$ - $E_{1,2}$).

En esta matriz léxica los conceptos son representados por la lista de unidades léxicas que pueden ser usadas para expresarlo (todas las entradas que pertenezcan a una misma fila), es decir, el conjunto de sinónimos (synset) no explica al concepto sino que simplemente indica que el concepto existe. No obstante, la representación propuesta por esta matriz léxica no puede trabajar directamente con los conceptos, sino que lo hace con las unidades léxicas. La relación léxica principal en WordNet es la sinonimia, pero también están presentes la antonimia, la hiperonimia, la hiponimia, la meronimia y las relaciones morfológicas.

WordNet está organizado con base en relaciones semánticas y como los significados están representados por medio de synsets, las relaciones semánticas están representadas como punteros (pointers) entre synsets. Es decir,

a la serie de sinónimos que constituye un synset se le suman símbolos que indican el tipo de relación que existe entre un synset y otro, por ejemplo @ para indicar los nombres superordinados y para los subordinados de otros synsets. Cabe agregar que los tipos de relaciones semánticas varían según la categoría sintáctica en que la que se establezcan.

WordNet, como base de datos léxica, es aplicada a la expansión de búsquedas (query expansion) como también a la desambiguación del sentido de las unidades léxicas (WSD) en el área de la recuperación y extracción de información.

4.1.2.3. Filtros

Con base en la ontología utilizada, se realiza un proceso de filtrado para obtener los resultados que se requieren para el dominio que se está utilizando.

Se utiliza el filtrado basado en conocimiento, el cual puede ser descrito como una *función*, que toma datos parciales de preferencias del usuario como parámetro de entrada y produce una lista de recomendaciones para cada usuario como salida, se utiliza este tipo de filtrado porque es el que se adapta al vector característico.

Primeramente, se tiene un conjunto con todos los elementos que lo componen:

$$E = \{e_1, e_2, e_n\},$$

donde:

E es el conjunto que contiene a los elementos

e_i es cualquier subconjunto del conjunto E

Por tanto, se se obtiene la siguiente función:

$$f(e_i \in E) \rightarrow c_j$$

En la anterior función se muestra que de cada elemento (e_i) dentro del conjunto E se devuelve un elemento c_j , el cual puede ser comparado con un elemento c_i del conjunto C .

$$C = \{c_1, c_2, c_m\},$$

donde:

C es el conjunto de características

c_i es cualquier subconjunto del conjunto C

Donde $1 \leq m$

En NLTK se comenta que dependiendo la técnica utilizada se pueden obtener las similitudes, una de éstas puede estar basada en el camino más corto que conecta las palabras como se puede apreciar en la Figura 4.3, en la cual las palabras definidas en la taxonomía más alejadas son menos similares, *es-un* (hipernimia/hiponimia), su rango es entre 0 y 1 (a menos que no se encuentre el camino) y comparando una palabra consigo misma da 1. Se cuenta también con otra forma que no dispone necesariamente del camino más corto que conecta los dos sentidos; por definición es el ancestro común más profundo en la taxonomía, no más cerca de los dos sentidos.

La probabilidad de que una palabra sea una instancia del concepto c :

$$P(c) = \frac{\sum_{w \in words(c)} count(w)}{N},$$

donde $words(c)$ son las palabras que se incluyen bajo el concepto c y N es la cantidad de palabras en el *corpus* usado. En la Figura 4.4 se muestran las probabilidades $p(c)$ de cada concepto, la palabra raíz es igual 1, ya que contiene a todas las palabras, el concepto más abajo en la jerarquía es igual a la probabilidad más baja, el contenido de información de un concepto se determina por c $IC(c) = -\log P(c)$.

Otra manera de obtener la similitud, es considerando el contenido de información (IC), el cual toma en cuenta las diferencias entre conceptos:

Comunalidad: $IC(\text{común}(a, b))$

Diferencia: $IC(\text{descripción}(A, B) - IC(\text{común}(a, b)))$

Por tanto, se utiliza en *Teorema de la similitud* NLTK, el cual dice que la similitud entre A y B se mide mediante la razón entre la cantidad de información necesaria para la comunidad de A y B y la información necesaria para describir completamente A y B :

$$sim_{Lin}(A, B) = \frac{common(A, B)}{description(A, B)}$$

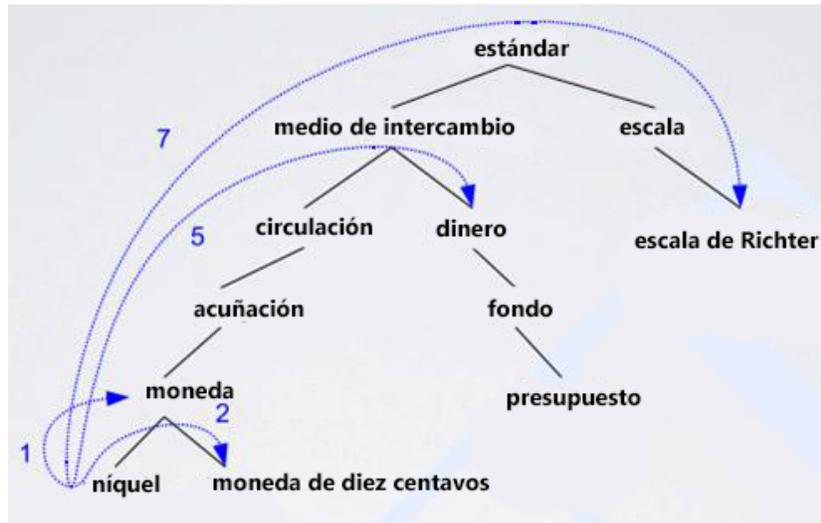


Figura 4.3: Ejemplo del camino más corto en la jerarquía ontológica.

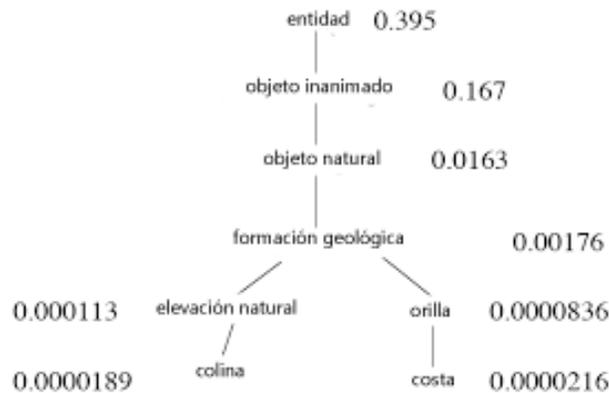


Figura 4.4: Ejemplo del ancestro común en la jerarquía.

Una vez que ha sido seleccionada la técnica de recuperación de similitud, se genera una tabla de similitudes donde se muestra la diagonal en la cual va a ser comparado ese valor con su mismo valor, por lo que el resultado será igual a 1 ($S(c_j, c_j) = 1$), mientras que en las demás celdas se reflejará la similitud para cada caso. Un ejemplo se puede observar en la Tabla 4.1.

Algunos ejemplos de técnicas de recuperación de similitud lexical database for English pueden ser `wup_similarity` (Wu & Palmer) esta calcula la profundidad de los dos conceptos en la UMLS, junto con la profundidad del LCS (menos común) La formula es:

$$score = 2 * depth(lcs) / (depth(s1) + depth(s2))$$

Esto significa que $0 < score \leq 1$. La puntuación (score) no puede ser cero porque la profundidad del LCS nunca es cero (la profundidad de la raíz de una taxonomía es uno). La puntuación es uno si los dos conceptos de entrada son los mismos.

Otro ejemplo de técnicas de recuperación de similitud es `path_similarity` la cual esta basada en el camino más corto que conecta a las palabras su rango es entre 0 y 1 (a menos que no se encuentre en el camino) y comparándose consigo misma el resultado es 1

$$sim_path(c1, c2) = -\log(len_path(c1, c2))$$

como tercer ejemplo esta `lch_similarity` medida de similitud propuesta por Leacock y Chodorow

$$-\log(length / (2 * D))$$

donde `length` es la longitud del camino mas corto entre los dos conceptos (usando conteo de nodos) y `D` es la profundidad máxima de la taxonomía.

Tabla 4.1: Matriz de similitudes de C .

	c_1	c_2	...	c_m
c_1	1			
c_2		1		
...			...	
c_m				1

La similitud tendrá un valor entre 0 a 1 siempre y cuando los valores de i, j sean menores o iguales a m .

$$0 \leq S(c_i, c_j) \leq 1 \quad \forall \quad i, j \leq m$$

A continuación se genera un vector correspondiente a nuestro perfil de usuario $P(U_i) = \{c_1, c_2, \dots, c_k\}$, el cual debe ser mayor o igual a 1 pero menor o igual a la cantidad de datos en el conjunto C .

$$1 \leq k \leq m \quad \forall \quad c_i \in C$$

$$R(U_i) = \{e_1, e_2, \dots, e_l\}$$

4.1.2.4. Recuperación de puntos de interés

La recuperación de los puntos de interés es mediante una base de datos espacial, en la cual los puntos se encuentran registrados con los datos que pueden estar indexados por un número único (*id*), puede incluir el nombre, el dominio al cual pertenece ese valor, y una ubicación geográfica representada por un punto que contiene longitud y latitud; ese punto puede ser representado dentro de un dominio como se muestra a continuación:

$$O_g = \{o_1, o_2, \dots, o_s\}$$

Se pueden obtener estos puntos desde diferentes instancias como por ejemplo Google Maps, OpenStreetMap, Bing Maps, Here Maps, Yandex Maps, entre otros o se puede crear una base de datos con los sitios de interés.

Referente a los puntos de interés si un sitio fue cerrado recientemente y sus bases de datos no han actualizadas se mostraran resultados incorrectos o antiguos.

En la Tabla 4.2 se puede ver como puede ser representada la información para su recuperación.

Tabla 4.2: Ejemplo de representación de datos de puntos de interes

Id	Tipo	Nombre	Posición Geográfica
1796526675	cinema	Cinepolis	0101000020E6100000CB1B458397D058C0F6E3BB5E355E3340
2682499791	school	Colegio Cristóbal Colón	0101000020E6100000BFB476458ED058C0A6ED5B9F0A 823340
851950234	university	Tecnológico de Monterrey, Campus Santa Fe	0101000020E610000032B19D898CD058C0F10AC0700E5C3340
1489012293	school	Instituto Tecnológico de Estudios Superiores de Monterrey Campus Santa Fe	0101000020E61000005283094686D058C0CB7F700DF15B3340
1835887102	school	Primaria Montessori	0101000020E610000069475D5B77D058C0B85BAAC24F7A 3340
2682477325	bank	Ban Bajío	0101000020E61000001027B0F80BD058C09F6E2262A77F3340
1975929160	hospital	Star Médica	0101000020E6100000B5FB392D0BD058C0E137DF6B87F3340
1975911013	fuel	PEMEX	0101000020E610000070826C77F9CF58C0A 3D20F56237F3340
1041569607	restaurant	Vips Axomiata	0101000020E6100000D8300527EACF58C0D073CA 3F75573340
1041550040	fast_food	Comida rápida Plaza Lilas	0101000020E6100000C741078FDA CF58C0B29A E30677633340
3750094501	cafe	Dunkin Donuts	0101000020E61000001272A9EAD5CF58C022C29C28C4E3340
331139119	parking	Free Parking	0101000020E610000082AAB801D1CF58C080CD6E0B8F6A 3340
1975915768	cinema	Cinemark	0101000020E61000008902D660D0CF58C06C3A9FE3E1 7E3340
2682518329	fast_food	McDonald's	0101000020E6100000545D9968C1CF58C0CD284940927E3340
2768795271	fuel	Pemex	0101000020E61000005087D63BCCF58C0C9F001C239613340
1836060878	university	FES Acatlan	0101000020E6100000CC0B47B6B3CF58C05E8CF596577B3340
2383430377	townhall	Delegación Magdalena Contreras	0101000020E6100000EFA F3A5774CF58C024F9772D0B4E3340
2324761563	hospital	Hospital Materno Infantil Magdalena Contreras	0101000020E6100000D76264144 CC F58C0252B0D74C 5513340
330798971	place_of_worship	Parroquia de Czestojova	0101000020E61000002C20ABB02FCF58C062A1C234286B3340
330800673	fountain	Fuente	0101000020E6100000A6491E8722CF58C0555C998DEA 6A 3340
1098346877	school	Escuela Primaria: Manuel E. Alvarez	0101000020E6100000CD41412C17CF58C04B3356FBDE743340
1836064691	theatre	Teatro Ciudad Bicentenario	0101000020E610000012DF9DD703CF58C00F03E7F5667A 3340
1359936894	bar	La Tradicional	0101000020E6100000DA07E18CFDCE58C0D189FCF8F6813340

4.1.3. Etapa de Visualización

Los resultados obtenidos con esta metodología se pueden representar de diferentes maneras:

1. Web GIS.
2. Mobile GIS (en esta visualización es posible utilizar la realidad aumentada).
3. Visualización del resultado en una consulta espacial.

4.1.3.1. Web GIS

Como se mencionó en la Sección 3.9 es posible gracias a la información de latitud y longitud, representar los datos geoespaciales. A su vez, se describió también que la estrategia utilizada para representar este tipo de datos es mapear los atributos espaciales directamente a dos dimensiones de la pantalla física, lo que resulta en el mapa de las visualizaciones. A su vez, se debe saber el tipo de entrada, las estructuras y los datos que se manejan.

Si los datos son externos, se requiere de un proceso de limpieza de los mismos.

En la Figura 3.26(a) se muestra el flujo de trabajo de las librerías tradicionales. Estas librerías tradicionales son las siguientes:

- Google Maps API,
- OpenLayers, y
- Leaflet

Una explicación más detallada de las librerías anteriores se presenta en la Sección 3.9.

4.1.3.2. Mobile GIS

Esta subsección presenta la división de los sistemas operativos móviles como Android android e iOS Apple, así como la realidad aumentada que es auxiliar para cualquiera de estos sistemas.

Mobile GIS tiene ventajas ya que como su nombre lo indica al ser un dispositivo móvil y con la ayuda de los sensores Android incluidos en el mismo dispositivo como:

- Sensores de movimiento: Estos sensores miden las fuerzas de aceleración y las fuerzas de rotación a lo largo de tres ejes. Esta categoría incluye acelerómetros, sensores de gravedad, giroscopios y sensores de vectores rotativos.
- Sensores ambientales: Estos sensores miden diversos parámetros ambientales, tales como temperatura y presión del aire ambiente, iluminación y humedad. Esta categoría incluye barómetros, fotómetros y termómetros.
- Sensores de posición: Estos sensores miden la posición física de un dispositivo. Esta categoría incluye sensores de orientación y magnetómetros.

con el uso de los sensores de movimiento y los sensores de posición se puede obtener la ubicación actual, la velocidad a la que se está moviendo, la dirección hacia donde está apuntando el dispositivo, los cuales son utilizados en esta metodología.

4.1.3.3. Sistemas operativos móviles

El sistema operativo Android es el más utilizado en dispositivos móviles, éste cuenta con una máquina virtual propia que puede ser Dalvik o Android Runtime, dependiendo de su versión.

El framework de Android está basado en el lenguaje de programación Java, el cual permite utilizar esa sintaxis para el desarrollo del mismo. Este lenguaje de programación cuenta con diferentes API's las cuales pueden ser integradas para implementar la realidad aumentada y con *layouts* propios para poder mostrar las diferentes vistas en las cuales se pueden desplegar los resultados.

Se elige el sistema operativo Android por su gran portabilidad, este al ser código abierto se puede desarrollar libremente en él, cuenta con todas las funciones necesarias para manipular los sensores de los dispositivos, todas sus aplicaciones son escritas en lenguaje de programación java, en la Figura 4.5 se muestra la arquitectura del sistema operativo Android.

Cuenta también con gráficos optimizados, los cuales son suministrados por medio de una librería de gráficos 2D. Los gráficos 3D se basan en la especificación Open GL ES 1.0, con soporte para aceleración gráfica por hardware (esta es opcional).

Para estructurar el almacenamiento de datos se cuenta con SQ Lite.

Es un Sistema Operativo móvil basado en Linux, que junto con aplicaciones middleware está enfocado para ser utilizado en dispositivos móviles como teléfonos inteligentes, tabletas, Google TV y otros dispositivos.



Figura 4.5: Arquitectura de Android.

Para una descripción más detallada de estos sistemas operativos móviles como el caso de Android, ésta se muestra en la Sección 3.8.

4.1.3.4. Realidad aumentada

Como se presentó en la 3.4.1, las aplicaciones de realidad aumentada enriquecen el entorno del mundo real con cierta cantidad de información sintética, se menciona que deben de existir tres requisitos básicos:

- Mezcla de imágenes reales y virtuales.
- Ejecución en tiempo real.
- Alineación de los objetos virtuales con estructuras del mundo real.

Se debe tener especial cuidado a la hora de sustituir las estructuras como se muestra en la Figura 3.13, ya que los aumentos por descuido de estructuras pueden causar la pérdida de información útil.

Otro problema puede ser la interacción visual entre las representaciones reales y virtuales como se muestra en la Figura 3.14, ya que si la utilización de los parámetros es igual, pero el objeto en el mundo real es diferente, los resultados pudieran llegar a ser no óptimos.

Se debe tener también precaución en la percepción de la profundidad, como se mencionó en la sección 3.4.1, las señales de profundidad ilustradas son aquellas que se pueden encontrar en una sola imagen incluyendo:

- Oclusión
- Tamaño relativo
- Altura relativa
- Detalle
- Perspectiva atmosférica
- Sombras
- Perspectiva lineal

También se muestra que la sincronización del parámetro de lo virtual y real permite a la cámara alinear señales de profundidad pictóricas reales y virtuales.

Otros aspectos a señalar son el manejo de la oclusión como se muestra en la Figura 3.16, en la cual los resultados pueden mostrarse de forma incorrecta, mostrándose superpuestos. En la misma sección se muestra el algoritmo de manejo de oclusión utilizando objetos fantasmas, el cual resuelve esta situación, también es necesario tener un buen control de las características de los bordes para mejorar el resultado con un detector de bordes para conservar los que sean necesarios como se muestra en la Figura 3.17. Otro aspecto a tomar en cuenta son las características sobresalientes relacionadas a la preservación

de los bordes de los elementos de ocultación, lo cual ayuda a mantener el contexto y mejora la percepción espacial.

El diagrama de clases que se utiliza se muestra en la Figura 4.6

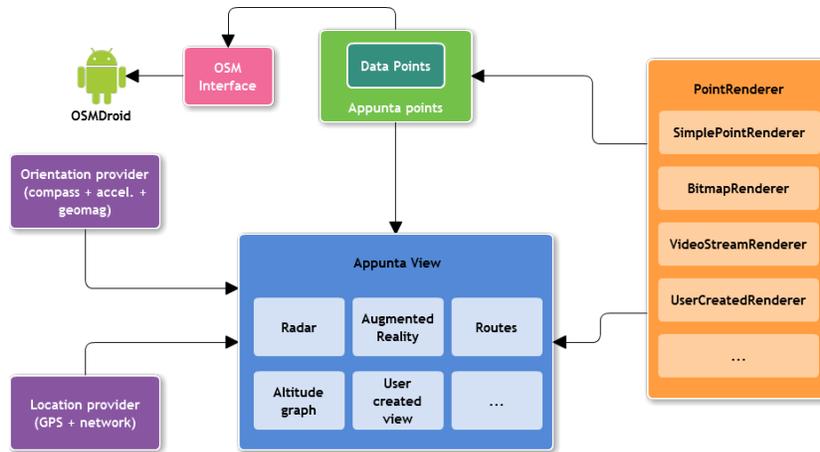


Figura 4.6: Arquitectura de Android.

El modelo matemático de las técnicas de visualización y la integración de los datos se describe en la Sección 3.4.1.

4.1.3.5. Visualización del resultado en una consulta espacial

Por medio de una base de datos con soporte de datos espaciales, se puede realizar una consulta espacial. Éstas difieren de las consultas SQL que no son espaciales porque permiten el uso de datos geométricos como puntos, líneas y polígonos y sus relaciones son geométricas y topológicas.

Algunos tipos de consultas se presentan en PostGIS:

- Distance(geometry, geometry) : number
- Touches(geometry, geometry) : boolean
- Contains(geometry, geometry) : boolean
- Crosses(geometry, geometry) : boolean

- Overlaps(geometry, geometry) : boolean
- Equals(geometry, geometry) : boolean
- Disjoint(geometry, geometry) : boolean
- Intersects(geometry, geometry) : boolean
- Area(geometry) : number

La consulta puede incluir un buffer el cual puede brindar resultados más apropiados ya que estos se encuentran dentro de un radio de la posición actual del usuario, un ejemplo en el cual solo cambiando la categoría que nosotros necesitemos podemos obtener el resultado deseado, es el siguiente:

```
select name, ST_ASTEXT(way)
from planet_osm_point
where ST_DWithin(way, ST_GeographyFromText('SRID=4326;POINT(-99.143887 19.510518)')
and name like '%Cine%'
```

el cual obtiene la salida que se muestra en la Tabla 4.3 la cual se basa en en posición actual (POINT(-99.143887 19.510518)) que se utiliza en la consulta:

Tabla 4.3: Resultados de la consulta

Nombre	Posición Geografica
Cinemex Ticoman	POINT(-99.1420001237274 19.5165946697402)
Cinapolis Torres Lindavista	POINT(-99.1436773681944 19.506422236764)
Cinemex Tenayuca	POINT(-99.1555828304864 19.5298817625377)

Capítulo 5

Resultados Experimentales

5.1. Etapa de Personalización

En la Figura 4.1 se muestra que lo primero que se necesita en el proceso de personalización es un perfil de usuario y un vector característico.

5.1.1. Perfil de usuario

En el caso de estudio para este trabajo, nos hemos enfocado en el dominio de la cinematografía, por lo cual el perfil de usuario incluye los géneros y su definición de las películas, estos se obtienen de la ontología de WordNet, por lo cual su significado se obtiene en inglés, al ser este el idioma de la ontología, las definiciones son las que a continuación se detallan:

- Action: the series of events that form a plot
- Adventure: a wild and exciting undertaking (not necessarily lawful)
- Animation: the making of animated cartoons
- Biography: an account of the series of events making up a person's life
- Comedy: light and humorous drama with a happy ending
- Crime: (criminal law) an act punishable by law; usually considered an evil act

- Documentary: a film or TV program presenting the facts about a person or event
- Drama: an episode that is turbulent or highly emotional
- Family: an association of people who share common beliefs or activities
- Fantasy: fiction with a large amount of imagination in it
- History: a record or narrative description of past events
- Horror: intense and profound fear
- Music: an artistic form of auditory communication incorporating instrumental or vocal tones in a structured and continuous manner
- Musical: a play or film whose action and dialogue is interspersed with singing and dancing
- Mystery: something that baffles understanding and cannot be explained
- Romance: a relationship between two lovers
- Sci-fi: a literary work based on the imagination and not necessarily on fact
- Sport: an active diversion requiring physical exertion and competition
- Thriller: a suspenseful adventure story or play or movie
- War: the waging of armed conflict against an enemy
- Western: a film about life in the western United States during the period of exploration and development

En la Figura 5.1 se presenta la lista completa desplegada en la aplicación en Android, la cual tiene la opción de seleccionar cada uno de los géneros (debiendo al menos uno ser seleccionado para su correcto funcionamiento), generando diferentes combinaciones. En la Figura 5.2 se puede observar el funcionamiento de la aplicación en la vista de selección de géneros y la selección de los mismos con base en los gustos del usuario, los cuales serán considerados para el proceso de recomendación.

Movie genres	
Action	<input type="checkbox"/>
Adventure	<input type="checkbox"/>
Animation	<input type="checkbox"/>
Biography	<input type="checkbox"/>
Comedy	<input type="checkbox"/>
Crime	<input type="checkbox"/>
Documentary	<input type="checkbox"/>
Drama	<input type="checkbox"/>
Family	<input type="checkbox"/>
Fantasy	<input type="checkbox"/>
History	<input type="checkbox"/>
Horror	<input type="checkbox"/>
Music	<input type="checkbox"/>
Musical	<input type="checkbox"/>
Mystery	<input type="checkbox"/>
Romance	<input type="checkbox"/>
Sci-Fi	<input type="checkbox"/>
Sport	<input type="checkbox"/>
Thriller	<input type="checkbox"/>
War	<input type="checkbox"/>
Western	<input type="checkbox"/>

Figura 5.1: Géneros en la aplicación.



Figura 5.2: Selección de géneros en la aplicación.

En Android se crea un recurso xml, el cual es un *PreferenceScreen*. Primero se genera un *PreferenceCategory* que sirve para almacenar el nombre del usuario, posteriormente se crea otro *PreferenceCategory*, el cual sirve para generar la vista que permite la selección de los géneros y guardar las preferencias de los mismos. A continuación se detalla como está estructurada el *PreferenceCategory* correspondiente a los géneros:

```
<CheckBoxPreference
android:key="action"
android:title="Action"
android:defaultValue="false"/>
```

```
<CheckBoxPreference
android:key="adventure"
android:title="Adventure"
android:defaultValue="false"/>
```

Posteriormente, se genera una clase que hereda de *PreferenceActivity* y el método *onCreate* envía el *PreferenceScreen* de la siguiente manera:

```
protected void onCreate(Bundle savedInstanceState) {
    super.onCreate(savedInstanceState);
    addPreferencesFromResource(R.xml.settings);
}
```

5.1.2. Vector característico

Para generar el vector que permite obtener los géneros, la forma de recuperar las preferencias del usuario en Android es con el método *getSharedPreferences*, recuperándolo de siguiente manera:

```
preferences.getBoolean("action", false);
```

Este método se repite para cada caso, con lo cual se puede llenar el vector con las preferencias del usuario.

5.2. Etapa de Procesamiento Semántico

5.2.1. Ontología

Con el uso de la ontología de WordNet Fellbaum (1998) y mediante el uso de los *synsets* lexical database for English, se puede obtener el resultado deseado. Para ello, se utiliza la técnica de similitud semántica, considerando el significado de las palabras. Así, sin tener información previa, se pueden obtener similitudes que se emplean para los resultados de la recomendación. A continuación se detallan los métodos utilizados en Python con el uso de la ontología WordNet y la librería NLTK para la recuperación de las similitudes por medio de los siguientes métodos *wup similarity*, *lch similarity* y *path similarity*.

```
def simwup(a, b):  
return wn.wup_similarity(wn.synset(a), wn.synset(b))
```

```
def simlch(a, b):  
return wn.lch_similarity(wn.synset(a), wn.synset(b))
```

```
def simpath(a, b):  
return wn.path_similarity(wn.synset(a),  
wn.synset(b))
```

Teniendo los métodos y utilizando las listas de comprensión, se puede obtener el listado con todas las combinaciones y su similitud entre ellas para poder darles uso para la recomendación. A continuación se muestra el código utilizado para la recuperación de dichas listas:

```
listawup = [(a, b, simwup(a, b)) for a in generos
            for b in generos]
listapath = [(a, b, simpath(a, b)) for a in generos
             for b in generos]
listalch = [(a, b, simlch(a, b)) for a in generos
            for b in generos]
```

A continuación se muestran las similitudes lexical database for English de *lch* (Leacock y Chodorow) en la Tabla 5.1. Las tablas de similitudes *wup* (Wu & Palmer) y *path* (Path length) lexical database for English se anexan al apéndice:

Tabla 5.1: Tabla de similitudes lch

ID	GENERO 1	GENERO 2	SIMILITUD
1	action	action	3.63758616
2	action	adventure	0.864997437
3	action	animation	0.747214402
4	action	biography	1.072636802
5	action	comedy	1.15267951
6	action	crime	0.864997437
7	action	documentary	0.929535959
8	action	drama	0.99852883
9	action	family	0.99852883
10	action	fantasy	2.028148247
11	action	history	1.15267951
12	action	horror	1.15267951
13	action	music	1.335001067
14	action	musical	0.929535959
15	action	mystery	2.251291799
16	action	romance	1.84582669
17	action	fiction	2.251291799
18	action	sport	0.929535959

19	action	thriller	2.028148247
20	action	war	0.99852883
21	action	western	0.864997437
22	adventure	action	0.864997437
23	adventure	adventure	3.63758616
24	adventure	animation	1.440361582
25	adventure	biography	0.99852883
26	adventure	comedy	1.072636802
27	adventure	crime	1.691676011
28	adventure	documentary	1.335001067
29	adventure	drama	1.440361582
30	adventure	family	1.072636802
31	adventure	fantasy	0.99852883
32	adventure	history	1.072636802
33	adventure	horror	1.239690887
34	adventure	music	1.239690887
35	adventure	musical	1.335001067
36	adventure	mystery	0.929535959
37	adventure	romance	0.929535959
38	adventure	fiction	1.072636802
39	adventure	sport	1.84582669
40	adventure	thriller	0.864997437
41	adventure	war	1.558144618
42	adventure	western	1.239690887
43	animation	action	0.747214402
44	animation	adventure	1.440361582
45	animation	animation	3.63758616
46	animation	biography	0.864997437
47	animation	comedy	0.929535959
48	animation	crime	1.440361582
49	animation	documentary	1.15267951
50	animation	drama	1.239690887
51	animation	family	0.929535959
52	animation	fantasy	0.864997437
53	animation	history	0.929535959
54	animation	horror	1.072636802

55	animation	music	1.072636802
56	animation	musical	1.15267951
57	animation	mystery	0.804372816
58	animation	romance	0.804372816
59	animation	fiction	0.929535959
60	animation	sport	1.558144618
61	animation	thriller	0.747214402
62	animation	war	1.335001067
63	animation	western	1.072636802
64	biography	action	1.072636802
65	biography	adventure	0.99852883
66	biography	animation	0.864997437
67	biography	biography	3.63758616
68	biography	comedy	1.335001067
69	biography	crime	0.99852883
70	biography	documentary	1.072636802
71	biography	drama	1.15267951
72	biography	family	1.15267951
73	biography	fantasy	1.239690887
74	biography	history	2.944438979
75	biography	horror	1.335001067
76	biography	music	1.558144618
77	biography	musical	1.072636802
78	biography	mystery	1.15267951
79	biography	romance	1.15267951
80	biography	fiction	1.335001067
81	biography	sport	1.072636802
82	biography	thriller	1.072636802
83	biography	war	1.15267951
84	biography	western	0.99852883
85	comedy	action	1.15267951
86	comedy	adventure	1.072636802
87	comedy	animation	0.929535959
88	comedy	biography	1.335001067
89	comedy	comedy	3.63758616
90	comedy	crime	1.072636802

91	comedy	documentary	1.15267951
92	comedy	drama	1.239690887
93	comedy	family	1.239690887
94	comedy	fantasy	1.335001067
95	comedy	history	1.440361582
96	comedy	horror	1.440361582
97	comedy	music	1.691676011
98	comedy	musical	1.15267951
99	comedy	mystery	1.239690887
100	comedy	romance	1.239690887
101	comedy	fiction	1.440361582
102	comedy	sport	1.15267951
103	comedy	thriller	1.15267951
104	comedy	war	1.239690887
105	comedy	western	1.072636802
106	crime	action	0.864997437
107	crime	adventure	1.691676011
108	crime	animation	1.440361582
109	crime	biography	0.99852883
110	crime	comedy	1.072636802
111	crime	crime	3.63758616
112	crime	documentary	1.335001067
113	crime	drama	1.440361582
114	crime	family	1.072636802
115	crime	fantasy	0.99852883
116	crime	history	1.072636802
117	crime	horror	1.239690887
118	crime	music	1.239690887
119	crime	musical	1.335001067
120	crime	mystery	0.929535959
121	crime	romance	0.929535959
122	crime	fiction	1.072636802
123	crime	sport	1.84582669
124	crime	thriller	0.864997437
125	crime	war	1.558144618
126	crime	western	1.239690887

127	documentary	action	0.929535959
128	documentary	adventure	1.335001067
129	documentary	animation	1.15267951
130	documentary	biography	1.072636802
131	documentary	comedy	1.15267951
132	documentary	crime	1.335001067
133	documentary	documentary	3.63758616
134	documentary	drama	1.558144618
135	documentary	family	1.15267951
136	documentary	fantasy	1.072636802
137	documentary	history	1.15267951
138	documentary	horror	1.335001067
139	documentary	music	1.335001067
140	documentary	musical	2.538973871
141	documentary	mystery	0.99852883
142	documentary	romance	0.99852883
143	documentary	fiction	1.15267951
144	documentary	sport	1.440361582
145	documentary	thriller	0.929535959
146	documentary	war	1.558144618
147	documentary	western	2.251291799
148	drama	action	0.99852883
149	drama	adventure	1.440361582
150	drama	animation	1.239690887
151	drama	biography	1.15267951
152	drama	comedy	1.239690887
153	drama	crime	1.440361582
154	drama	documentary	1.558144618
155	drama	drama	3.63758616
156	drama	family	1.239690887
157	drama	fantasy	1.15267951
158	drama	history	1.239690887
159	drama	horror	1.440361582
160	drama	music	1.440361582
161	drama	musical	1.558144618
162	drama	mystery	1.072636802

163	drama	romance	1.072636802
164	drama	fiction	1.239690887
165	drama	sport	1.558144618
166	drama	thriller	0.99852883
167	drama	war	1.691676011
168	drama	western	1.440361582
169	family	action	0.99852883
170	family	adventure	1.072636802
171	family	animation	0.929535959
172	family	biography	1.15267951
173	family	comedy	1.239690887
174	family	crime	1.072636802
175	family	documentary	1.15267951
176	family	drama	1.239690887
177	family	family	3.63758616
178	family	fantasy	1.15267951
179	family	history	1.239690887
180	family	horror	1.440361582
181	family	music	1.440361582
182	family	musical	1.15267951
183	family	mystery	1.072636802
184	family	romance	1.072636802
185	family	fiction	1.239690887
186	family	sport	1.15267951
187	family	thriller	0.99852883
188	family	war	1.239690887
189	family	western	1.072636802
190	fantasy	action	2.028148247
191	fantasy	adventure	0.99852883
192	fantasy	animation	0.864997437
193	fantasy	biography	1.239690887
194	fantasy	comedy	1.335001067
195	fantasy	crime	0.99852883
196	fantasy	documentary	1.072636802
197	fantasy	drama	1.15267951
198	fantasy	family	1.15267951

199	fantasy	fantasy	3.63758616
200	fantasy	history	1.335001067
201	fantasy	horror	1.335001067
202	fantasy	music	1.558144618
203	fantasy	musical	1.072636802
204	fantasy	mystery	2.251291799
205	fantasy	romance	2.251291799
206	fantasy	fiction	2.944438979
207	fantasy	sport	1.072636802
208	fantasy	thriller	2.028148247
209	fantasy	war	1.15267951
210	fantasy	western	0.99852883
211	history	action	1.15267951
212	history	adventure	1.072636802
213	history	animation	0.929535959
214	history	biography	2.944438979
215	history	comedy	1.440361582
216	history	crime	1.072636802
217	history	documentary	1.15267951
218	history	drama	1.239690887
219	history	family	1.239690887
220	history	fantasy	1.335001067
221	history	history	3.63758616
222	history	horror	1.440361582
223	history	music	1.691676011
224	history	musical	1.15267951
225	history	mystery	1.239690887
226	history	romance	1.239690887
227	history	fiction	1.440361582
228	history	sport	1.15267951
229	history	thriller	1.15267951
230	history	war	1.239690887
231	history	western	1.072636802
232	horror	action	1.15267951
233	horror	adventure	1.239690887
234	horror	animation	1.072636802

235	horror	biography	1.335001067
236	horror	comedy	1.440361582
237	horror	crime	1.239690887
238	horror	documentary	1.335001067
239	horror	drama	1.440361582
240	horror	family	1.440361582
241	horror	fantasy	1.335001067
242	horror	history	1.440361582
243	horror	horror	3.63758616
244	horror	music	1.691676011
245	horror	musical	1.335001067
246	horror	mystery	1.239690887
247	horror	romance	1.239690887
248	horror	fiction	1.440361582
249	horror	sport	1.335001067
250	horror	thriller	1.15267951
251	horror	war	1.440361582
252	horror	western	1.239690887
253	music	action	1.335001067
254	music	adventure	1.239690887
255	music	animation	1.072636802
256	music	biography	1.558144618
257	music	comedy	1.691676011
258	music	crime	1.239690887
259	music	documentary	1.335001067
260	music	drama	1.440361582
261	music	family	1.440361582
262	music	fantasy	1.558144618
263	music	history	1.691676011
264	music	horror	1.691676011
265	music	music	3.63758616
266	music	musical	1.335001067
267	music	mystery	1.440361582
268	music	romance	1.440361582
269	music	fiction	1.691676011
270	music	sport	1.335001067

271	music	thriller	1.335001067
272	music	war	1.440361582
273	music	western	1.239690887
274	musical	action	0.929535959
275	musical	adventure	1.335001067
276	musical	animation	1.15267951
277	musical	biography	1.072636802
278	musical	comedy	1.15267951
279	musical	crime	1.335001067
280	musical	documentary	2.538973871
281	musical	drama	1.558144618
282	musical	family	1.15267951
283	musical	fantasy	1.072636802
284	musical	history	1.15267951
285	musical	horror	1.335001067
286	musical	music	1.335001067
287	musical	musical	3.63758616
288	musical	mystery	0.99852883
289	musical	romance	0.99852883
290	musical	fiction	1.15267951
291	musical	sport	1.440361582
292	musical	thriller	0.929535959
293	musical	war	1.558144618
294	musical	western	2.251291799
295	mystery	action	2.251291799
296	mystery	adventure	0.929535959
297	mystery	animation	0.804372816
298	mystery	biography	1.15267951
299	mystery	comedy	1.239690887
300	mystery	crime	0.929535959
301	mystery	documentary	0.99852883
302	mystery	drama	1.072636802
303	mystery	family	1.072636802
304	mystery	fantasy	2.251291799
305	mystery	history	1.239690887
306	mystery	horror	1.239690887

307	mystery	music	1.440361582
308	mystery	musical	0.99852883
309	mystery	mystery	3.63758616
310	mystery	romance	2.028148247
311	mystery	fiction	2.538973871
312	mystery	sport	0.99852883
313	mystery	thriller	2.251291799
314	mystery	war	1.072636802
315	mystery	western	0.929535959
316	romance	action	1.84582669
317	romance	adventure	0.929535959
318	romance	animation	0.804372816
319	romance	biography	1.15267951
320	romance	comedy	1.239690887
321	romance	crime	0.929535959
322	romance	documentary	0.99852883
323	romance	drama	1.072636802
324	romance	family	1.072636802
325	romance	fantasy	2.251291799
326	romance	history	1.239690887
327	romance	horror	1.239690887
328	romance	music	1.440361582
329	romance	musical	0.99852883
330	romance	mystery	2.028148247
331	romance	romance	3.63758616
332	romance	fiction	2.538973871
333	romance	sport	0.99852883
334	romance	thriller	1.84582669
335	romance	war	1.072636802
336	romance	western	0.929535959
337	fiction	action	2.251291799
338	fiction	adventure	1.072636802
339	fiction	animation	0.929535959
340	fiction	biography	1.335001067
341	fiction	comedy	1.440361582
342	fiction	crime	1.072636802

343	fiction	documentary	1.15267951
344	fiction	drama	1.239690887
345	fiction	family	1.239690887
346	fiction	fantasy	2.944438979
347	fiction	history	1.440361582
348	fiction	horror	1.440361582
349	fiction	music	1.691676011
350	fiction	musical	1.15267951
351	fiction	mystery	2.538973871
352	fiction	romance	2.538973871
353	fiction	fiction	3.63758616
354	fiction	sport	1.15267951
355	fiction	thriller	2.251291799
356	fiction	war	1.239690887
357	fiction	western	1.072636802
358	sport	action	0.929535959
359	sport	adventure	1.84582669
360	sport	animation	1.558144618
361	sport	biography	1.072636802
362	sport	comedy	1.15267951
363	sport	crime	1.84582669
364	sport	documentary	1.440361582
365	sport	drama	1.558144618
366	sport	family	1.15267951
367	sport	fantasy	1.072636802
368	sport	history	1.15267951
369	sport	horror	1.335001067
370	sport	music	1.335001067
371	sport	musical	1.440361582
372	sport	mystery	0.99852883
373	sport	romance	0.99852883
374	sport	fiction	1.15267951
375	sport	sport	3.63758616
376	sport	thriller	0.929535959
377	sport	war	1.691676011
378	sport	western	1.335001067

379	thriller	action	2.028148247
380	thriller	adventure	0.864997437
381	thriller	animation	0.747214402
382	thriller	biography	1.072636802
383	thriller	comedy	1.15267951
384	thriller	crime	0.864997437
385	thriller	documentary	0.929535959
386	thriller	drama	0.99852883
387	thriller	family	0.99852883
388	thriller	fantasy	2.028148247
389	thriller	history	1.15267951
390	thriller	horror	1.15267951
391	thriller	music	1.335001067
392	thriller	musical	0.929535959
393	thriller	mystery	2.251291799
394	thriller	romance	1.84582669
395	thriller	fiction	2.251291799
396	thriller	sport	0.929535959
397	thriller	thriller	3.63758616
398	thriller	war	0.99852883
399	thriller	western	0.864997437
400	war	action	0.99852883
401	war	adventure	1.558144618
402	war	animation	1.335001067
403	war	biography	1.15267951
404	war	comedy	1.239690887
405	war	crime	1.558144618
406	war	documentary	1.558144618
407	war	drama	1.691676011
408	war	family	1.239690887
409	war	fantasy	1.15267951
410	war	history	1.239690887
411	war	horror	1.440361582
412	war	music	1.440361582
413	war	musical	1.558144618
414	war	mystery	1.072636802

415	war	romance	1.072636802
416	war	fiction	1.239690887
417	war	sport	1.691676011
418	war	thriller	0.99852883
419	war	war	3.63758616
420	war	western	1.440361582
421	western	action	0.864997437
422	western	adventure	1.239690887
423	western	animation	1.072636802
424	western	biography	0.99852883
425	western	comedy	1.072636802
426	western	crime	1.239690887
427	western	documentary	2.251291799
428	western	drama	1.440361582
429	western	family	1.072636802
430	western	fantasy	0.99852883
431	western	history	1.072636802
432	western	horror	1.239690887
433	western	music	1.239690887
434	western	musical	2.251291799
435	western	mystery	0.929535959
436	western	romance	0.929535959
437	western	fiction	1.072636802
438	western	sport	1.335001067
439	western	thriller	0.864997437
440	western	war	1.440361582
441	western	western	3.63758616

A partir de la Figura 5.3 a la Figura 5.23 se muestra la similitud de cada género comparado contra su mismo género y comparado contra los demás géneros, utilizando la medida de similitud de Leacock & Chodorow (lch), la medida de similitud de Path length (path) y la medida de similitud de Wu & Palmer (wup).

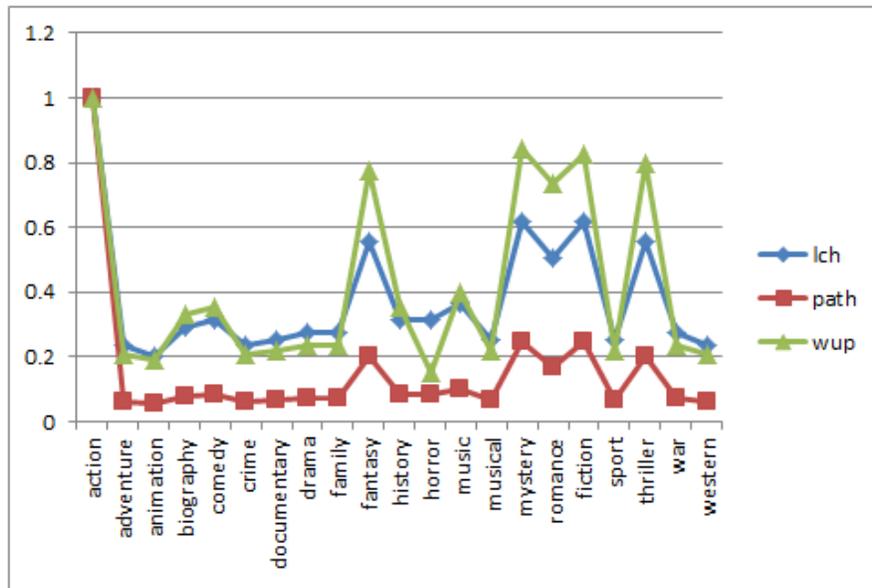


Figura 5.3: Comparativo de similitudes de **action** contra los demás géneros

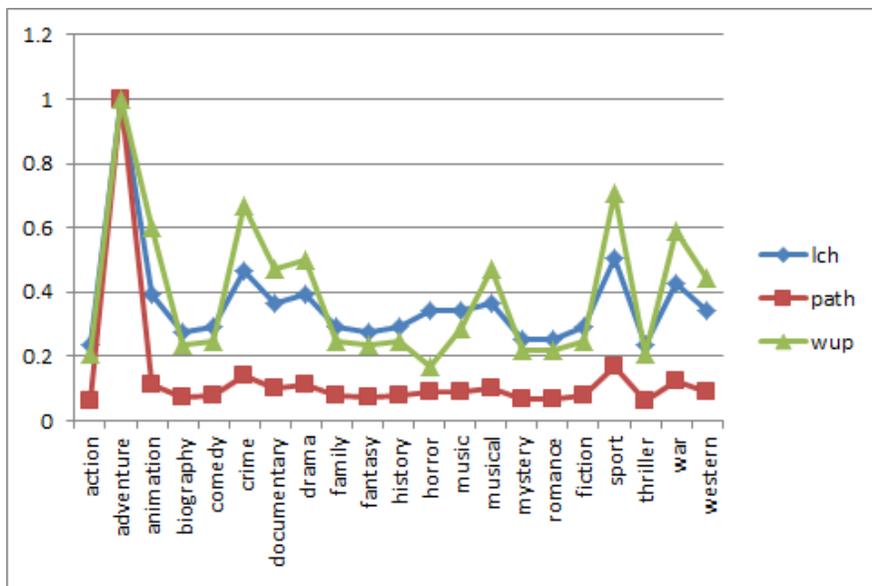


Figura 5.4: Comparativo de similitudes de **adventure** contra los demás géneros

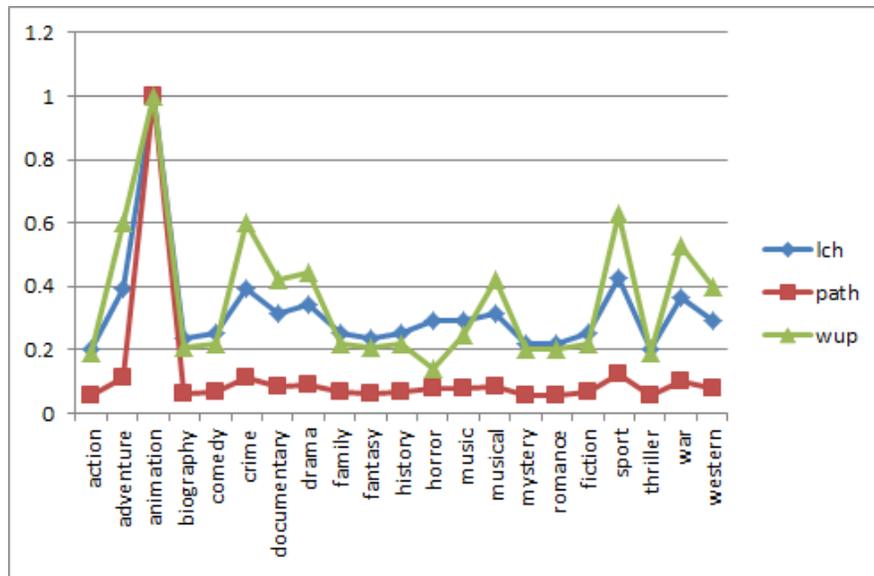


Figura 5.5: Comparativo de similitudes de **animation** contra los demás géneros

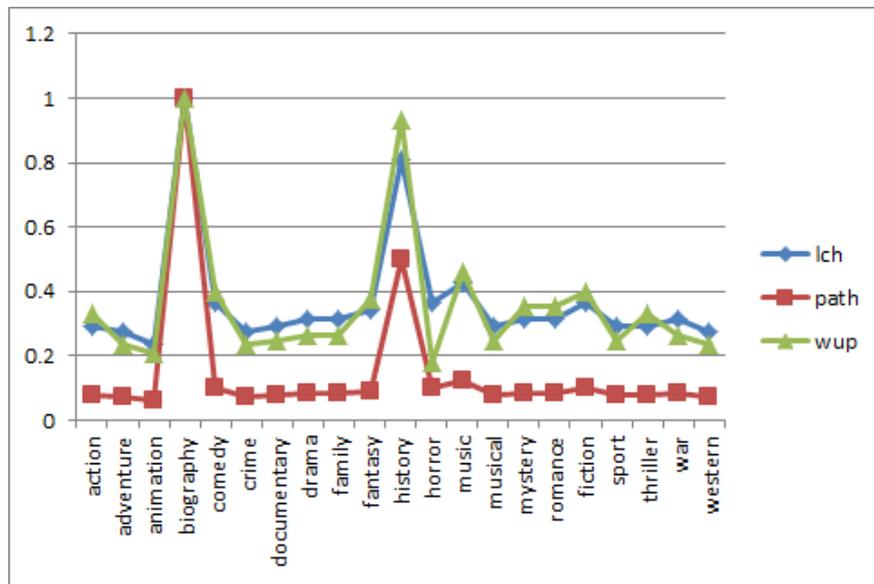


Figura 5.6: Comparativo de similitudes de **biography** contra los demás géneros

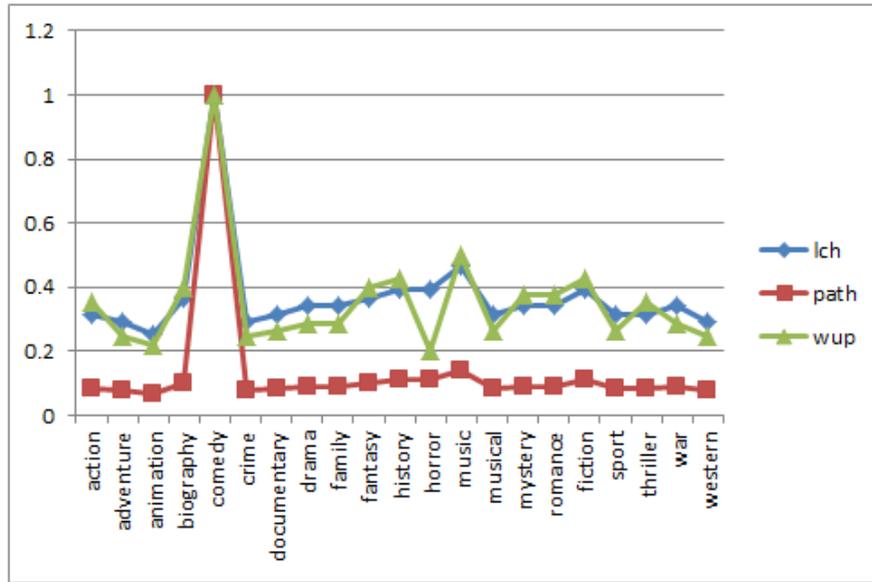


Figura 5.7: Comparativo de similitudes de **comedy** contra los demás géneros

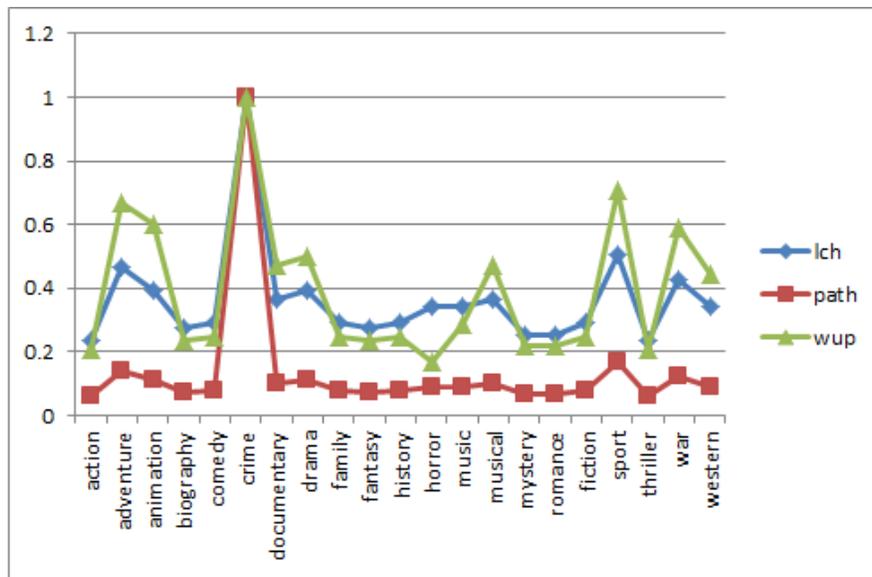


Figura 5.8: Comparativo de similitudes de **crime** contra los demás géneros

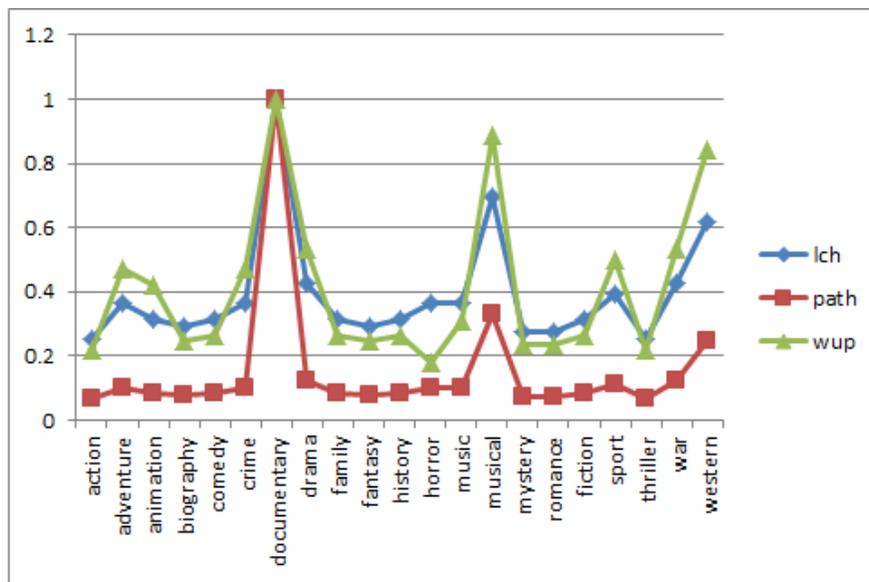


Figura 5.9: Comparativo de similitudes de **documentary** contra los demás géneros

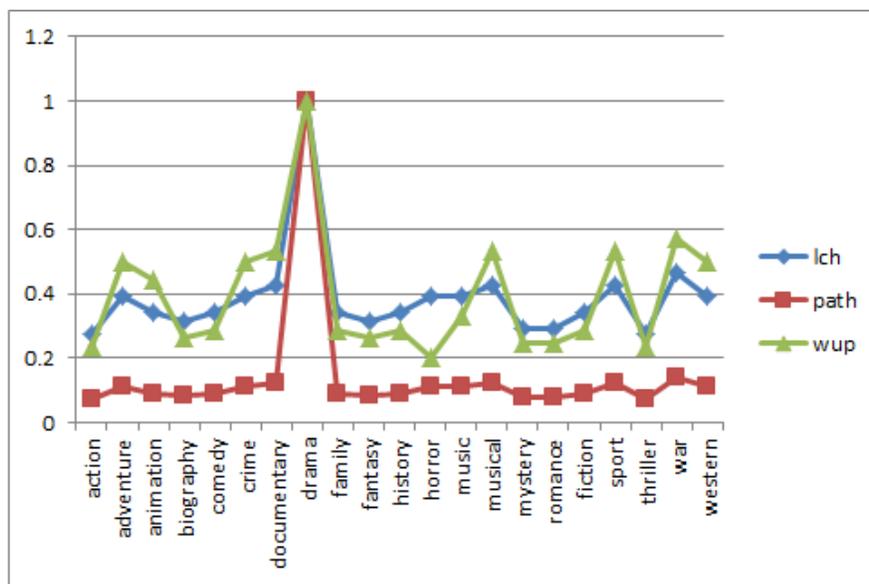


Figura 5.10: Comparativo de similitudes de **drama** contra los demás géneros

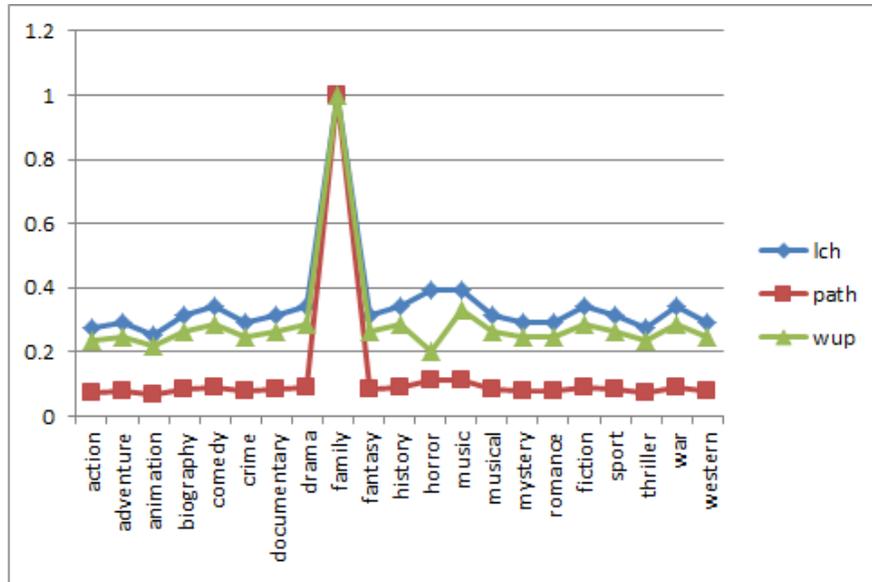


Figura 5.11: Comparativo de similitudes de **family** contra los demás géneros

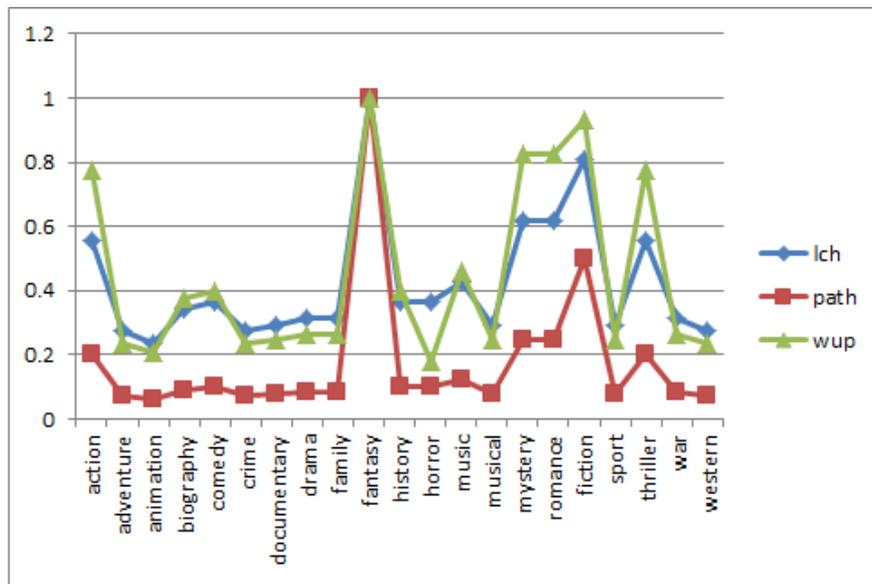


Figura 5.12: Comparativo de similitudes de **fantasy** contra los demás géneros

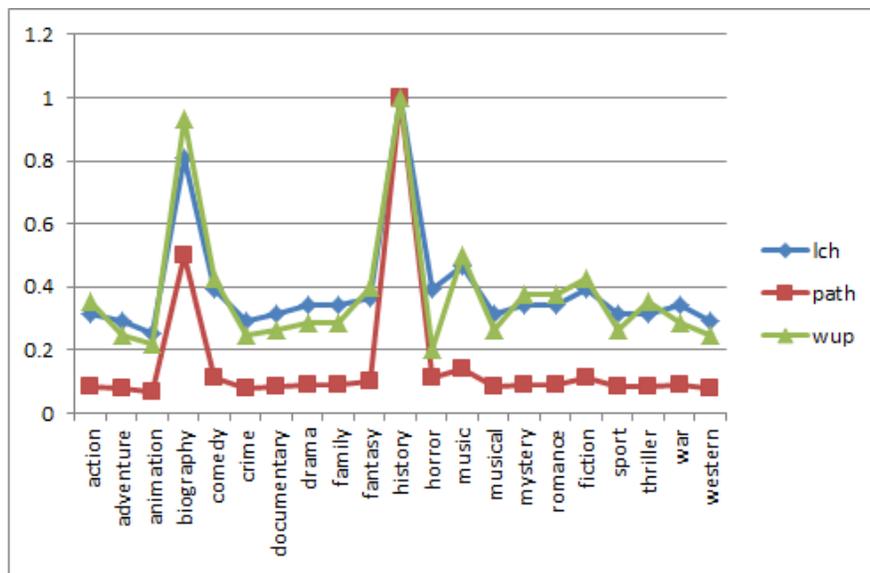


Figura 5.13: Comparativo de similitudes de **history** contra los demás géneros

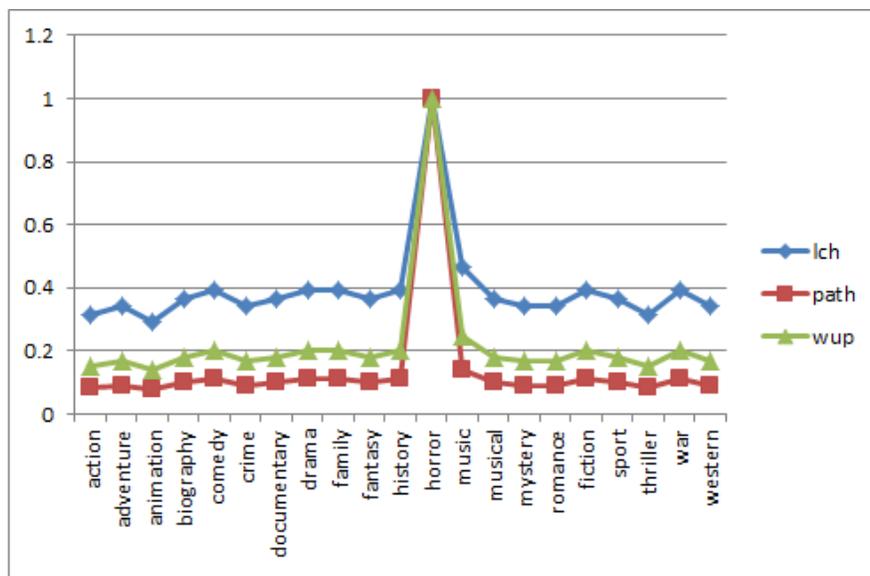


Figura 5.14: Comparativo de similitudes de **horror** contra los demás géneros

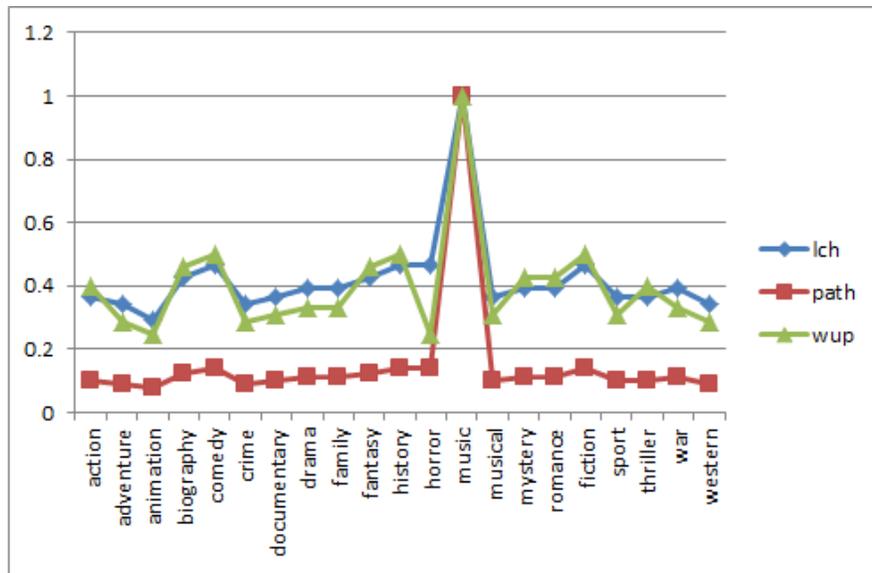


Figura 5.15: Comparativo de similitudes de **music** contra los demás géneros

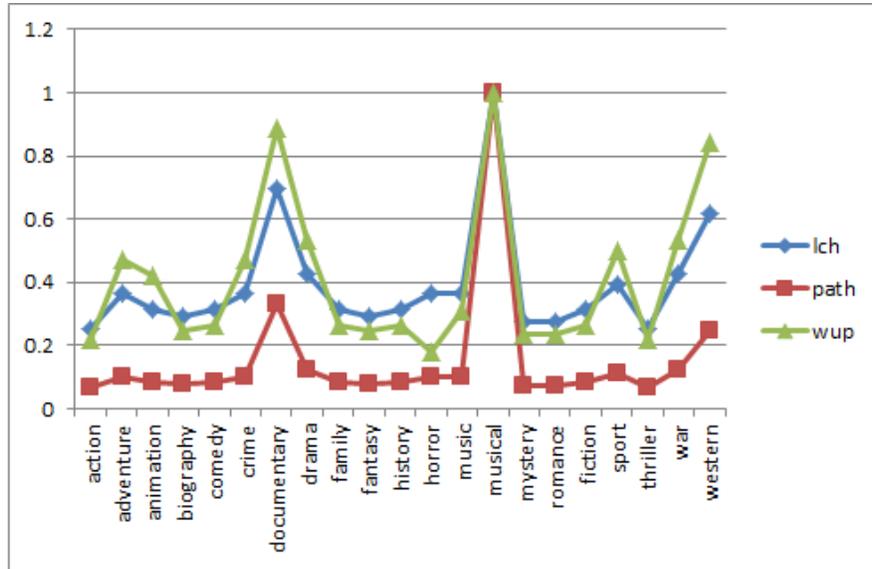


Figura 5.16: Comparativo de similitudes de **musical** contra los demás géneros

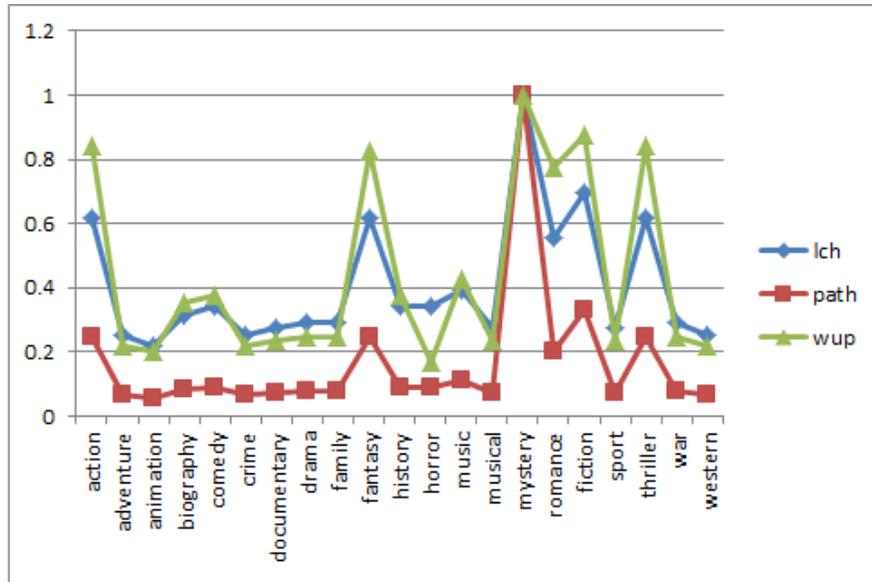


Figura 5.17: Comparativo de similitudes de **mystery** contra los demás géneros

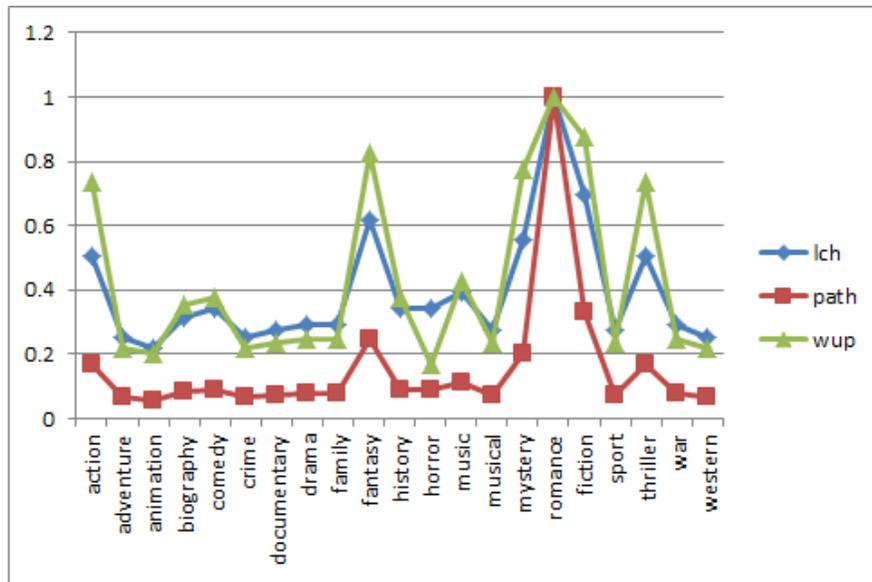


Figura 5.18: Comparativo de similitudes de **romance** contra los demás géneros

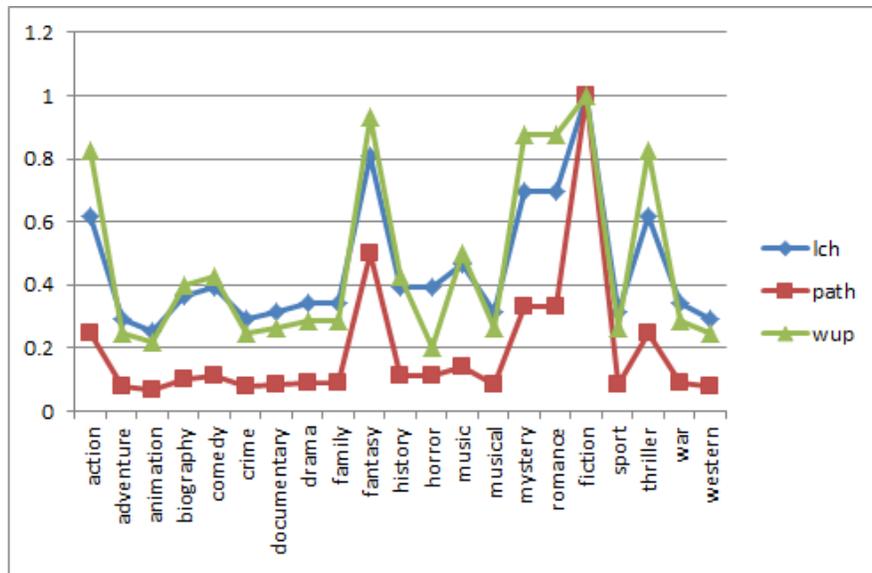


Figura 5.19: Comparativo de similitudes de **fiction (sci-fi)** contra los demás géneros

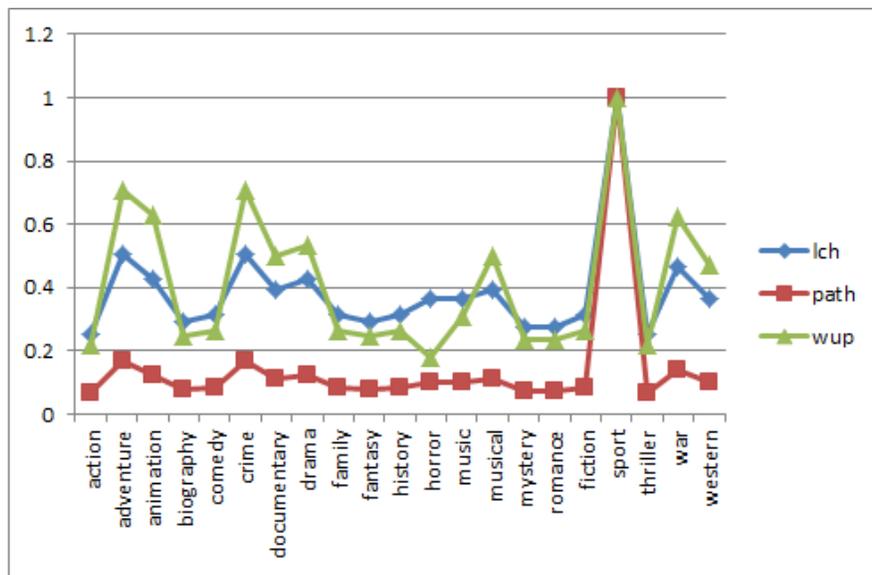


Figura 5.20: Comparativo de similitudes de **sport** contra los demás géneros

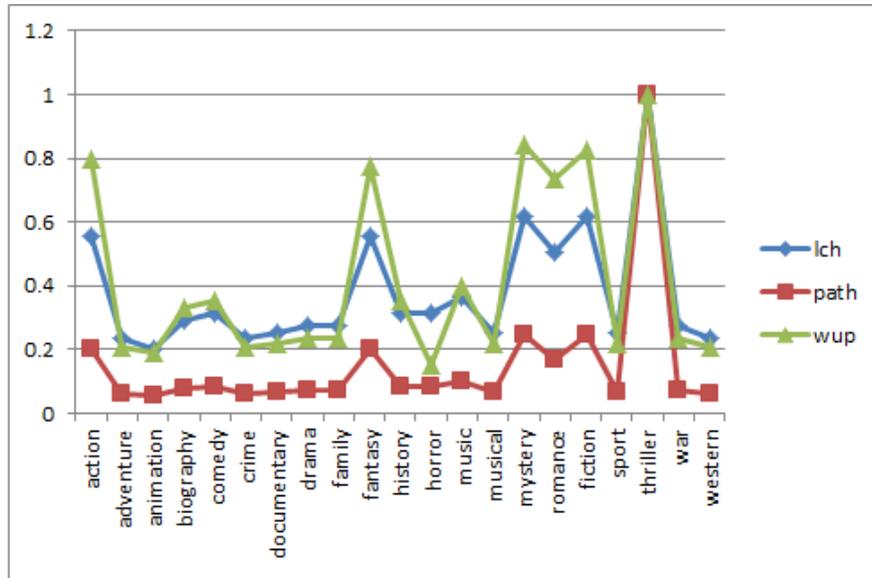


Figura 5.21: Comparativo de similitudes de **thriller** contra los demás géneros

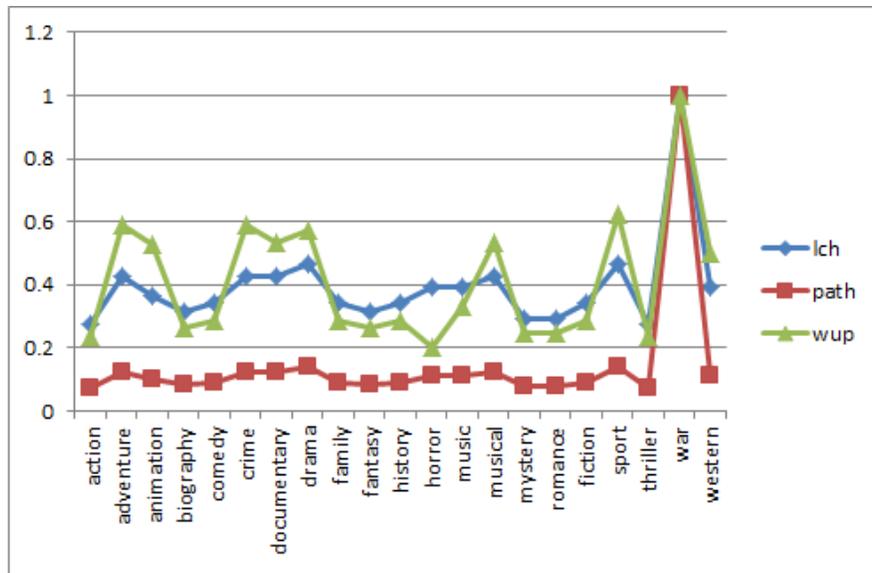


Figura 5.22: Comparativo de similitudes de **war** contra los demás géneros

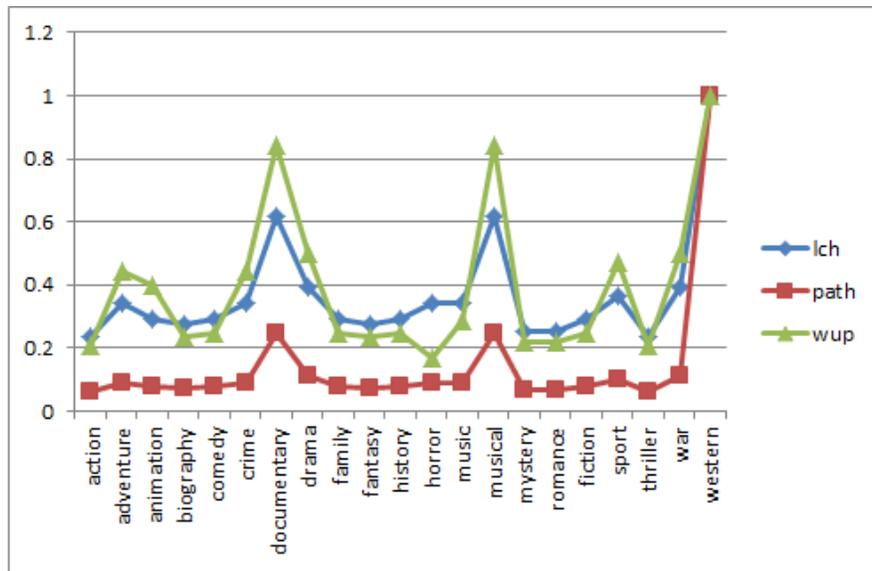


Figura 5.23: Comparativo de similitudes de **western** contra los demás géneros

Como se aprecia en las imágenes se puede por el movimiento de las gráficas que tanto *lch* como *wup* muestran mejores resultados ya que sus medidas de similitud son más precisas, ya que en *path* no se alcanza a percibir la diferencia entre muchos de los puntos, pareciendo estos ser iguales, mientras que en *wup* a pesar de ser la similitud entre 0 y 1 al igual que *path* una variación mas pronunciada entre los resultados, así a su vez podemos ver que en *lch* al ser su rango de similitud entre 0 y 3.63758616 los resultados se muestran con mayor detalle y las medidas de similitud pueden ser más precisas.

5.2.2. Filtrado

Una vez que se tiene el perfil del usuario con las selecciones de los gustos de los géneros de las películas (Cliente) y se cuenta con las tablas relacionales para la obtención del filtrado (Servidor), las cuales se pueden observar en la Figura 5.24, realizando una consulta SQL se pueden obtener las similitudes, para obtener los datos del perfil de usuario y usarlo dentro de la consulta se pueden recuperar en Android de la siguiente manera:

```
String cadena = "";
boolean comprobar;
comprobar = preferences.getBoolean("action", false);
if(comprobar && primero)
{
    primero = false;
    cadena += "id_genero1 = 1 ";
}else if(comprobar){
    cadena += "or id_genero1 = 1 ";
}
}
```

Así, continuando para cada uno de los géneros, y teniendo la cadena ésta se envía por medio de un servicio web, el cual se recibe por PHP utilizando el framework de Laravel; su función es devolver una respuesta de una petición POST mediante un arreglo, en donde esta respuesta contiene la recuperación de las películas con mayor similitud al perfil de usuario. A continuación se muestra el código del método que se encarga de recibir y dar respuesta al servicio web.

```
public function postIndex(Request $request)
{
    $generos = $request->all();
    $cadena = $generos['cadena'];

    $cines = DB::select("select titulo,
        clasificacion, genero, similitud from
            peliculas, generos,
            genero_pelicula, similitudes where
            genero_pelicula.pelicula_id =
            peliculas.id and
            genero_pelicula.genero_id = generos.id
            and ($cadena) and id_genero2 =
            generos.id order by similitud desc limit
            10");

    return response()->json($cines,200);
}
```

```
}
```

Asimismo, es posible enviar la solicitud del servicio web desde Android de la siguiente manera y con la cadena que anteriormente se generó; el código es el siguiente:

```
public String POST(String url, String cadena) {
    InputStream inputStream = null;
    String result = "";
    try {
        //1 crea el HTTPClient
        HttpClient httpClient = new DefaultHttpClient();
        //2 hace el post a la url dada
        HttpPost httpPost = new HttpPost(url);
        String json = "";
        //3 construir el objeto JSON
        JSONObject jsonObject = new JSONObject();
        jsonObject.accumulate("cadena", cadena);
        //4 convertir JSONObject a JSON string
        json = jsonObject.toString();
        //5 set Json a StringEntity (datos crudos)
        StringEntity se = new StringEntity(json);
        //6 set httpst Entity
        httpPost.setEntity(se);
        //7 set algunos encabezados para informar al
            servidor acerca del tipo del contenido
        httpPost.setHeader("Accept", "application/json");
        httpPost.setHeader("Content-type",
            "application/json");
        //8 Ejecutar la solicitud del POST a la url dada
        HttpResponse httpResponse =
            httpClient.execute(httpPost);
        //recibir la respuesta como un inputStream
        inputStream = httpResponse.getEntity().getContent();
        //convertir el inputStream a string
        if (inputStream != null) {
```

```
result = convertInputStreamToString(inputStream);
} else {
result = "no funciona";
}
} catch (Exception e) {
Log.d("InputStream ", e.getLocalizedMessage());
}
resultado = result;
//11 regresar result
return result;
}
```

La respuesta del servicio web es un arreglo que dependiendo de los gustos que fueron seleccionados en su perfil de usuario pueden ser representados de la siguiente manera:

```
[
{
"titulo": "Deadpool",
"clasificacion": "B15",
"genero": "action",
"similitud": "3.63758616"
},
{
"titulo": "Leyenda",
"clasificacion": "B",
"genero": "action",
"similitud": "3.63758616"
},
{
"titulo": "La Quinta Ola",
"clasificacion": "B",
"genero": "action",
"similitud": "3.63758616"
},
{
"titulo": "600 Millas",
"clasificacion": "B",
```

```
"genero": "thriller",
"similitud": "2.028148247"
},
{
"titulo": "El Lado Oscuro del Deseo",
"clasificacion": "C",
"genero": "thriller",
"similitud": "2.028148247"
},
{
"titulo": "Carol",
"clasificacion": "B15",
"genero": "thriller",
"similitud": "2.028148247"
},
{
"titulo": "Yo Vi Al Diablo",
"clasificacion": "B15",
"genero": "thriller",
"similitud": "2.028148247"
},
{
"titulo": "Frente al Mar",
"clasificacion": "B15",
"genero": "romance",
"similitud": "1.84582669"
},
{
"titulo": "Suite Francesa",
"clasificacion": "B",
"genero": "romance",
"similitud": "1.84582669"
},
{
"titulo": "Orgullo, Prejuicio y Zombies",
"clasificacion": "B",
"genero": "romance",
"similitud": "1.84582669"
```

```
}
]
```

Al ser un arreglo, a éste se le da el trato necesario para su visualización.

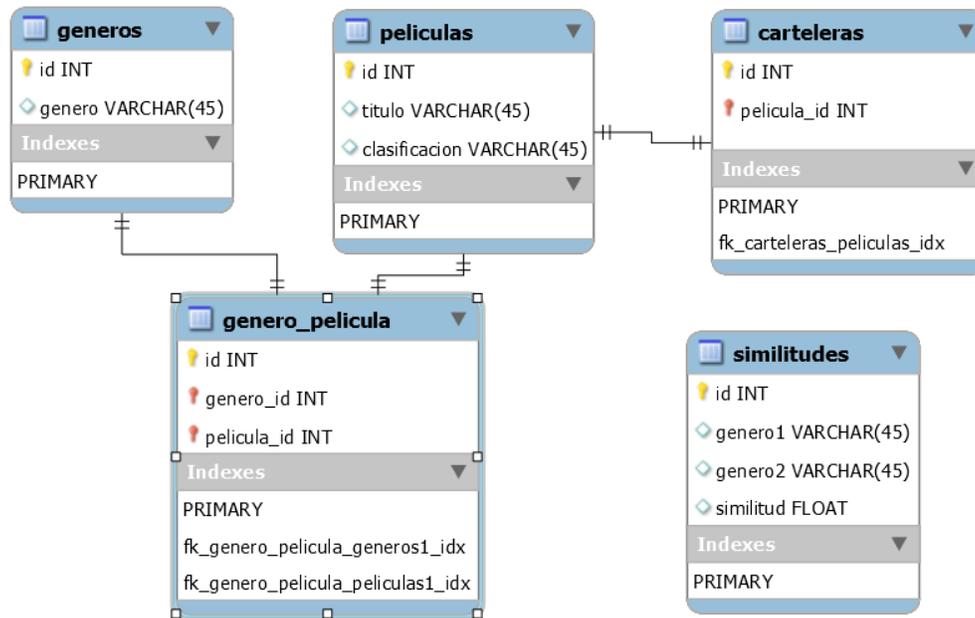


Figura 5.24: Diagrama de la base de datos de películas y géneros

5.3. Visualización

En esta etapa se utilizaron dos bases de datos: la primera es la que se obtiene por medio de OpenStreetMap www.openstreetmap.org y el resto se puede apreciar de la Figura 5.24.

Los datos de los puntos de interés se obtienen desde el sitio web de OpenStreetMap www.openstreetmap.org y con ayuda de la herramienta Osm2pgsql podemos exportarla a PostGIS PostGIS, en el caso de este trabajo como puntos.

En la tabla de “generos” se encuentran los 21 géneros más utilizados en las películas. En la tabla de “películas” los títulos que se encuentran en ese momento en la cartelera o un listado más completo. En la tabla “carteleras” se tiene la relación de las tablas que están en ese momento en cartelera, la tabla “genero_pelicula” mantiene la relación con la tabla de “generos” y “películas”, ya que es una relación *muchos-muchos*.

Lo primero que debe realizarse es recuperar por medio de un servicio web y por medio de SQL, una consulta espacial del lado del servidor. Por tanto, la consulta se estructura de la siguiente manera:

```
class PuntosController extends Controller {

public function postIndex(Request $request){

    $puntos = $request->all();
    $lat = $puntos['latitud'];
    $lon = $puntos['longitud'];
    $cines = DB::select("select name, ST_ASTEXT(way)
from planet_osm_point
where ST_DWithin(way,
ST_GeographyFromText('SRID=4326;POINT($lon $lat)'),
3000) and name like '%Cine%'");
    return response()->json($cines,200);
}
}
```

Donde:

3000 es la distancia del buffer

lat es la latitud actual

lon es la longitud actual

Con lo anterior, se recupera un arreglo de puntos que representan en este caso de estudio a todos los cines en la distancia especificada dentro del buffer, la cual es de 3 km. Para la representación se utiliza la API *Appunta*. En la Figura 3.20 se observa su funcionamiento, con la recuperación del arreglo y con ayuda de las clases *PointRenderer* y *EyeViewRenderer*, pertenecientes a esta API. Así, podemos iniciar el dibujado de los puntos recuperados, en la Figura 5.25 y la Figura 5.26 se muestra la recuperación de el o los puntos

correspondientes a las salas de cines, auxiliándose de la orientación de la cámara aparecen solo las salas de cine que se encuentran en ese rango de visión como se muestra en la Figura 5.27.

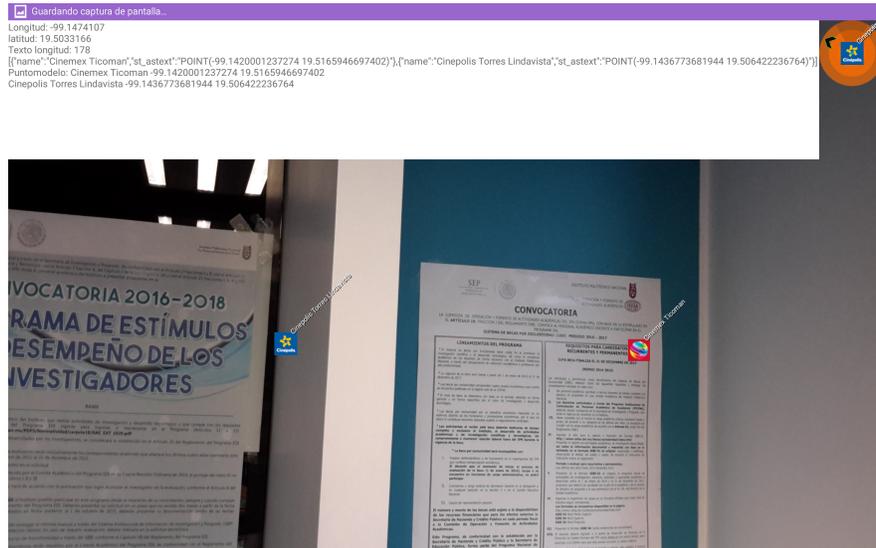


Figura 5.25: Resultado de la ubicación 1.

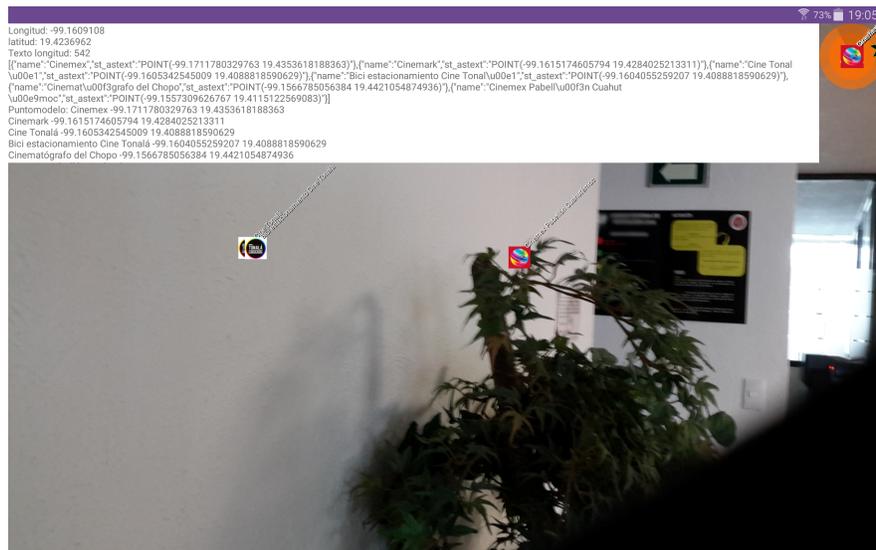


Figura 5.26: Resultado de la ubicación 2.

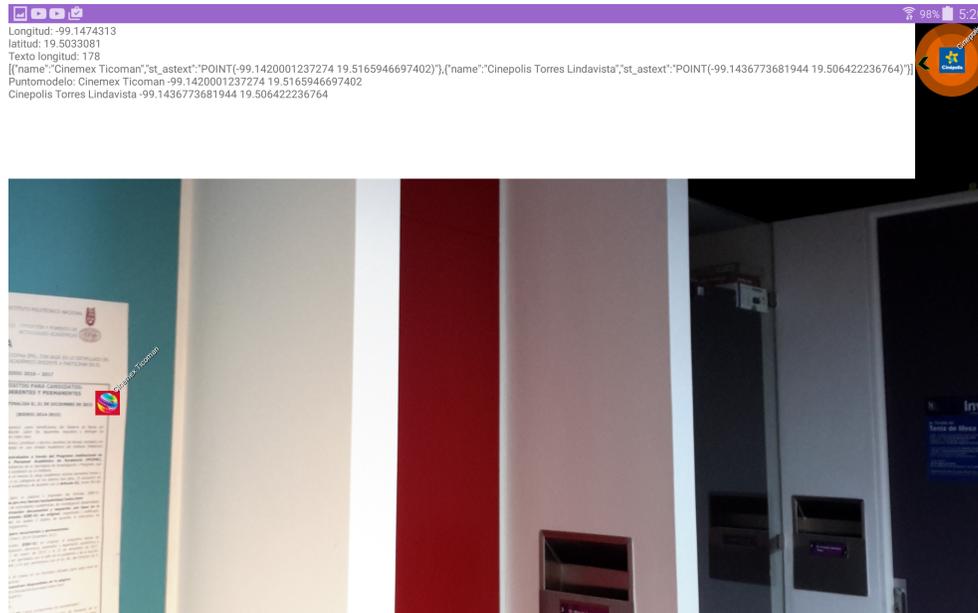


Figura 5.27: Resultado de la ubicación 1 con una rotación diferente de la cámara mostrando una sola coincidencia.

En la parte superior se muestra la ubicación actual, la cual se va actualizando conforme se va moviendo el usuario y recalculando los puntos a recuperar. También se puede observar el arreglo recuperado que muestra el punto donde se localiza la latitud y longitud de la sala de cine y a continuación el nombre del cine encontrado.

Cada icono funciona como un botón y al hacer clic despliega la cartelera, de acuerdo con los datos recuperados del servicio web, entonces se agrega a la lista para mostrarlo en pantalla. El código que se muestra a continuación es una posible forma para mostrar los resultados:

```

ArrayList<Pelicula> peliculas = new
    ArrayList<Pelicula>();
termino = true;
ListView lv = (ListView)findViewById(R.id.listview);
try {
pelicula = null;
pelicula = recuperarJSON(resPOST);

```

```
String cad = "";  
  
for (Pelicula peli : pelicula) {  
    peliculas.add(new Pelicula(peli.getTitulo(),  
        peli.getClasificacion(), peli.getGenero(),  
        peli.getSimilitud()));  
}  
  
//textView.setText(cad);  
  
MobileArrayAdapter mobileArrayAdapter = new  
    MobileArrayAdapter(CarteleraActivity.this,  
        peliculas);  
lv.setAdapter(mobileArrayAdapter);
```

Finalmente, el procedimiento anterior despliega los resultados de diez películas en cartelera con la mayor coincidencia, por medio de la similitud con respecto a las selecciones de los gustos de géneros en el perfil de usuario, tal como se muestra en la Figura 5.28.

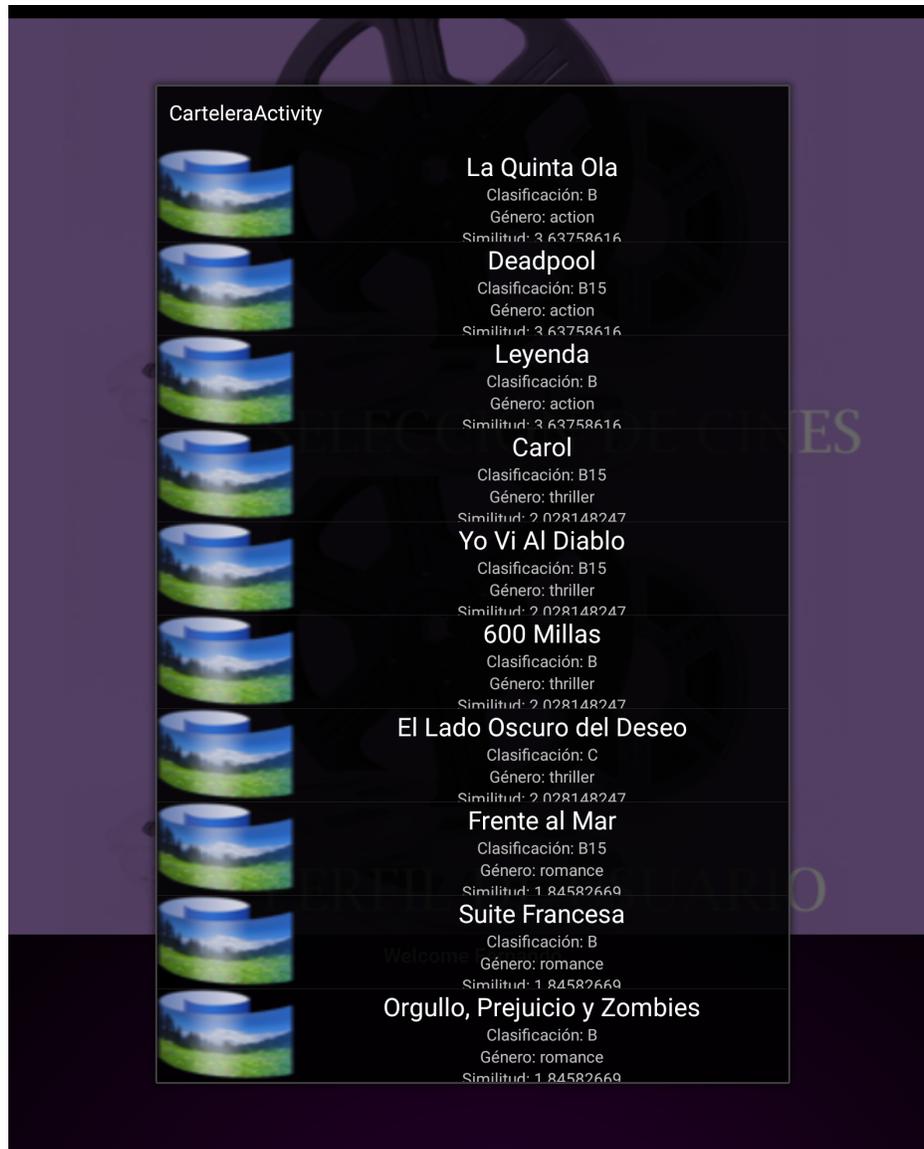


Figura 5.28: Vista de las películas con su género y similitud.

Capítulo 6

Conclusiones y Trabajos Futuros

6.1. Conclusiones

En el presente trabajo de tesis, se desarrolló una metodología enfocada a la recuperación semántica de información con base en un perfil de usuario y representada con ayuda de la realidad aumentada.

La metodología se divide en tres secciones, en la *etapa de personalización* por medio de un perfil de usuario se eligen las preferencias del usuario, las cuales son guardadas en un vector característicos, el cual es utilizado para su posterior filtrado.

La segunda etapa es el *procesamiento semántico* en el cual se utiliza la ontología de WordNet la cual es una base de datos léxica grande del idioma Ingles, esta se divide en sustantivos, verbos, adjetivos y adverbios, cada uno de estos cuenta con una definición, la cual se utiliza en este trabajo junto con un filtrado por contexto (puede ser descrito como una función), este toma datos parciales de las preferencias del usuario como parámetro de entrada y produce una lista de recomendaciones para cada usuario como salida, esta recomendación esta basada en la similitud de los géneros cinematográficos.

La tercera etapa de visualización, en esta etapa se visualizan dos cosas, la primera son los puntos de interés con ayuda de la realidad aumentada y por medio de una base de datos espacial la cual cuenta con las posiciones geográficas en el caso de este trabajo el *punto* el cual a su vez cuenta con una

latitud y *longitud*, cada punto en el caso de ser representado en dispositivo móvil renderiza el icono o imagen por medio de la realidad aumentada. La segunda visualización es la lista de resultados ya ranqueados, estos pueden ser representados tanto en un dispositivo móvil (este puede ser mostrado con ayuda de la realidad aumentada), en una página web, o guardar los datos en una base de datos o en una base de datos espacial si se requiere las ubicaciones geográficas.

La aplicación móvil para probar el funcionamiento sistema recomendador brinda al usuario la facilidad de encontrar dentro de un radio configurable a las necesidades del usuario, esta genera una consulta geoespacial la cual regresa los puntos de interés, en el caso de la aplicación las salas cinematográficas, al acceder a la imagen que nos muestra la aplicación despliega diez resultados de la cartelera cinematográfica por medio del filtrado que se genera, mostrándose en las primeras posiciones las películas de mayor interés para el usuario y en las últimas posiciones las que son de menor interés.

La metodología enfocada a la recuperación semántica de información con base en un perfil de usuario y representada con ayuda de la realidad aumentada es resultado de un arduo trabajo de investigación en análisis y transformación de datos geoespaciales, en el procesamiento semántico, en la investigación de la creación y uso de las ontologías, el funcionamiento y uso de la realidad aumentada.

6.2. Aportaciones científicas

Entre las aportaciones tenemos las siguientes:

- Evitar mostrar resultados vacíos, ya que al ser por medio de similitudes, aunque estas sean muy pequeñas siempre se obtendrá un resultado.
- Al utilizarse una ontología como WordNet la cual cuenta con una gran variedad de sustantivos, verbos, adjetivos y adverbios, se pueden brindar resultados de recomendación sin necesidad de tener datos de otros usuarios, solo con las preferencias del usuario

- Aplicación en Android la cual brinda ubicaciones reales dentro de la Ciudad de México de las salas cinematográficas y es capaz por medio de las similitudes de brindar recomendaciones de películas.
- Aplicación en Python la cual permite generar las listas de similitudes de géneros cinematográficos utilizando las medidas de similitud Path length, Leacock & Chodorow y Wu & Palmer
- Aplicación en PHP (Laravel) la cual por medio de dos webservice, el primero regresa los resultados de los puntos de interés en un radio seleccionado y el segundo la información de la película como el título, la clasificación, el genero y la similitud basándose en las preferencias del usuario.
- Clasificación de géneros de películas en WordNet

6.3. Limitaciones

- El vector característico solo es de una dimensión para denotar con un 1 si al usuario es de su preferencia y con un 0 si no es de su agrado
- No se cuenta con datos de redes sociales
- No se cuenta con crowdsourcing
- En la ontología WordNet se tienen que escoger manualmente las definiciones
- La alimentación de la cartelera es de forma manual
- La representación de los puntos de interés esta limitada por la api de Appunta

6.4. Trabajos a futuro

Entre las implementaciones que se pueden agregar, están las siguientes:

- Incrementar el área de búsqueda para que pueda ser empleado en otras zonas geográficas más allá de la Ciudad de México

- Aumentar las dimensiones del vector característico para obtener mejores resultados
- Darle una ponderación a los géneros para que el usuario elija cuanto le gusta y no solo si lo prefiere o no
- Integrarlo con redes sociales para obtener información de los puntos de interés y poderlos utilizar junto con las preferencias del usuario para una mejor recomendación
- Recuperar por medio de api's o web scrapping la información de la cartelera en tiempo real
- Crear histogramas de gustos del usuario para inferir nuevos géneros.

Referencias

- Aditya, T., & Kraak, M.-J. (2005). Reengineering the geoportal: Applying hci and geovisualization disciplines. In *Proceedings of the 11th EC-GI & GIS Workshop, Alghero, Italy*. volume 29.
- Al-Shamri, M. Y. H. (2016). User profiling approaches for demographic recommender systems. *Knowledge-Based Systems*, 100, 175–187.
- android (). <http://www.android.com/>. URL: <http://www.android.com/history>.
- Android, D. (). <https://developer.android.com>. URL: <https://developer.android.com>.
- androidcentral (). <http://www.androidcentral.com/>. URL: <http://www.androidcentral.com/android-history>.
- Apple (). Apple : <https://swift.org/about/>. URL: <https://swift.org/about/>.
- Arano, S. (2004). La ontología: una zona de interacción entre la lingüística y la documentación. *www.hipertext.net*, .
- ARToolKit (). Hitlab : <http://www.hitl.washington.edu/artoolkit>. URL: <http://www.hitl.washington.edu/artoolkit>.
- Avery, B., Sandor, C., & Thomas, B. H. (2009). Improving spatial perception for augmented reality x-ray vision. In *2009 IEEE Virtual Reality Conference* (pp. 79–82). IEEE.
- Azuma, R. T. (2013). A survey of augmented reality presence: Teleoperators and virtual environment.

- Baizal, Z., Iskandar, A., & Nasution, E. (2016). Ontology-based recommendation involving consumer product reviews. In *Information and Communication Technology (ICoICT), 2016 4th International Conference on* (pp. 1–6). IEEE.
- Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., & Su, Z. (2007). Optimizing web search using social annotations. In *Proceedings of the 16th international conference on World Wide Web* (pp. 501–510). ACM.
- Bischoff, K., Firan, C. S., Nejdil, W., & Paiu, R. (2008). Can all tags be used for search? In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 193–202). ACM.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. URL: <http://www.sciencedirect.com/science/article/pii/S0950705113001044>. doi:10.1016/j.knosys.2013.03.012.
- Bogdanović, M., Stanimirović, A., & Stoimenov, L. (2015). Methodology for geospatial data source discovery in ontology-driven geoinformation integration architectures. *Web Semantics: Science, Services and Agents on the World Wide Web*, 32, 1 – 15. URL: <http://www.sciencedirect.com/science/article/pii/S1570826815000037>. doi:<http://dx.doi.org/10.1016/j.websem.2015.01.002>.
- Buccella, A., Cechich, A., & Fillottrani, P. (2009). Ontology-driven geographic information integration: A survey of current approaches. *Computers & Geosciences*, 35, 710–723.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12, 331–370.
- Cai, Y., & Li, Q. (2010). Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 969–978). ACM.
- Calvanese, D., & De Giacomo, G. (2005). Data integration: A logic-based perspective. *AI magazine*, 26, 59.

- Carrer-Neto, W., Hernández-Alcaraz, M. L., Valencia-García, R., & García-Sánchez, F. (2012). Social knowledge-based recommender system. application to the movies domain. *Expert Systems with Applications*, 39, 10990 – 11000. URL: <http://www.sciencedirect.com/science/article/pii/S0957417412004952>. doi:<http://dx.doi.org/10.1016/j.eswa.2012.03.025>.
- Chirita, P.-A., Firan, C. S., & Nejdl, W. (2007). Personalized query expansion for the web. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 7–14). ACM.
- Chirita, P. A., Nejdl, W., Paiu, R., & Kohlschütter, C. (2005). Using odp metadata to personalize search. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 178–185). ACM.
- Christakou, C., & Stafylopatis, A. (2005). A hybrid movie recommender system based on neural networks. In *ISDA* (pp. 500–505). IEEE Computer Society. URL: <http://dblp.uni-trier.de/db/conf/isda/isda2005.html#ChristakouS05>.
- Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2006). Ontological engineering: principles, methods, tools and languages. In *Ontologies for software engineering and software technology* (pp. 1–48). Springer.
- Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Construcción de ontologías legales con la metodología methontology y la herramienta webode, .
- De Courcy, R. (1992). Les systèmes d'information en réadaptation. *Québec, Réseau International CIDIH et facteurs environnementaux*, 1, 7–10.
- Delicious (). <https://delicious.com/>. URL: <https://delicious.com/>.
- dmoz (). <http://www.dmoz.org/>. URL: <http://www.dmoz.org/>.
- lexical database for English, W. A. (). <https://wordnet.princeton.edu/>. URL: <https://wordnet.princeton.edu/>.

- Fatemi, M., & Tokarchuk, L. (2013). A community based social recommender system for individuals & groups. In *Social Computing (SocialCom), 2013 International Conference on* (pp. 351–356). IEEE.
- Fellbaum, C. (Ed.) (1998). *Wordnet, an Electronic Lexical Database*. MIT Press.
- Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering, .
- Fombona Cadavieco, J., Sevillano, P., Ángeles, M., & Madeira Ferreira Amador, M. F. (2012). Realidad aumentada, una evolución de las aplicaciones de los dispositivos móviles, .
- Furht, B. (2011). *Handbook of augmented reality*. Springer Science & Business Media.
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of information science*, *32*, 198–208.
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2006). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, *5*, 199–220.
- Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies, .
- Hakimpour, F. (2003). *Using ontologies to resolve semantic heterogeneity for integrating spatial database schemata*.
- Hamynen, K., Korhonen, P., Kahari, M., Sorvari, A., Huhtala, Y., Murphy, D., & Paalasmaa, J. (2010). Displaying network objects in mobile devices based on geolocation. US Patent 7,720,436.
- Kalkofen, D., Mendez, E., & Schmalstieg, D. (2009). Comprehensible visualization for augmented reality. *IEEE transactions on visualization and computer graphics*, *15*, 193–204.

- Lara, L. H., & Villarreal, J. L. (2004). La realidad aumentada: una tecnología en espera de usuarios. *Revista Digital Universitaria*, 10.
- Laravel (). <https://laravel.com/>. URL: <https://laravel.com/>.
- Last.fm (). <http://www.last.fm/>. URL: <http://www.last.fm>.
- Leaflet (). Leaflet : <http://leafletjs.com/reference.html>. URL: <http://Leafletjs.com/reference.html>.
- Ledur, C., Griebler, D., Manssour, I., & Fernandes, L. G. (2015). Towards a domain-specific language for geospatial data visualization maps with big data sets. In *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1–8). IEEE.
- Liarokapis, F., Greatbatch, I., Mountain, D., Gunesh, A., Brujic-Okretic, V., & Raper, J. (2005). Mobile augmented reality techniques for geovisualisation. In *Ninth International Conference on Information Visualisation (IV'05)* (pp. 745–751). IEEE.
- López, M. F., Gómez-Pérez, A., Sierra, J. P., & Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment. *IEEE intelligent Systems*, (pp. 37–46).
- Maedche, A., Motik, B., Stojanovic, L., Studer, R., & Volz, R. (2003). Ontologies for enterprise knowledge management. *IEEE Intelligent Systems*, (pp. 26–33).
- Martínez, S. (2012). Appunta framework <http://appunta.com>. URL: <http://appunta.com>.
- Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38, 39–41. URL: <http://wordnet.princeton.edu>. doi:10.1145/219717.219748.
- NLTK (). Natural language toolkit <http://www.nltk.org/>. URL: <http://www.nltk.org/>.
- Noll, M. G., & Meinel, C. (2007). *Web search personalization via social bookmarking and tagging*. Springer.

- Noy, N. F., Fergerson, R. W., & Musen, M. A. (2000). The knowledge model of protege-2000: Combining interoperability and flexibility. In *Knowledge Engineering and Knowledge Management Methods, Models, and Tools* (pp. 17–32). Springer.
- OpenLayers (). Openlayers : <http://openlayers.org/en/v3.6.0/doc/tutorials/introduction.html>.
URL: <http://openlayers.org/en/v3.6.0/doc/tutorials/introduction.html>.
- Osm2pgsql (). Osm2pgsql <http://wiki.openstreetmap.org/wiki/osm2pgsql>.
URL: <http://wiki.openstreetmap.org/wiki/Osm2pgsql>.
- php (). <http://www.php.net/>. URL: <http://www.php.net>.
- PinkFroot (). Ship ahoy! : <http://my.pinkfroot.com/>. URL: <http://my.pinkfroot.com/>.
- Poli, R. (2010). *Theory and applications of ontology*. Dordrecht London: Springer.
- PostGIS (). <http://postgis.net/>. URL: <http://postgis.net/>.
- PostgreSQL (). <http://www.postgresql.org.es/>. URL: <http://www.postgresql.org.es/>.
- Pretschner, A., & Gauch, S. (1999). Ontology based personalized search. In *Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on* (pp. 391–398). IEEE.
- pythonprogramming.net (). <http://pythonprogramming.net/>. URL: <https://pythonprogramming.net/wordnet-nltk-tutorial/>.
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Commun. ACM*, 40, 56–58. URL: <http://doi.acm.org/10.1145/245108.245121>. doi:10.1145/245108.245121.
- Ricci, F. (2011). *Recommender systems handbook*. New York London: Springer.
- Sayar, A. (2009). *High performance, federated, service-oriented geographic information systems*. Ph.D. thesis Citeseer.

- Staab, S., Studer, R., Schnurr, H.-P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent systems*, (pp. 26–34).
- Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., & Wenke, D. (2002). Ontoedit: Collaborative ontology engineering for the semantic web. In *Proceedings of the First International Semantic Web Conference (ISWC'02)*.
- Svennerberg, G. (2010). *Beginning Google Maps API 3*. Apress.
- Tait, M. G. (2005). Implementing geoportals: applications of distributed {GIS}. *Computers, Environment and Urban Systems*, 29, 33 – 47. URL: <http://www.sciencedirect.com/science/article/pii/S019897150400047X>. doi:<http://dx.doi.org/10.1016/j.compenvurbsys.2004.05.011>. Geoportals.
- Tellez-Arenas, A. (2009). *Best Practice Report on Geoportals*. Technical Report ECP-2007-GEO-317001, OneGeology-Europe.
- Uschold, M., & King, M. (1995). *Towards a methodology for building ontologies*. Citeseer.
- Ward, M. O., Grinstein, G., & Keim, D. (2010). *Interactive data visualization: foundations, techniques, and applications*. CRC Press.
- Wellner, P. (). The digitaldesk calculator: Tactile manipulation on a desktop display.(1992) in proceedings of uist'92, the a cm symposium on user interface software and technology.
- Wilson, J. R., & D´ Cruz, M. (2006). Virtual and interactive environments for work of the future. *International Journal of Human-Computer Studies*, 64, 158–169.
- www.openstreetmap.org, O. (). <http://www.openstreetmap.org/>. URL: <http://www.openstreetmap.org/>.
- Xu, S., Bao, S., Fei, B., Su, Z., & Yu, Y. (2008). Exploring folksonomy for personalized search. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 155–162). ACM.

- Yu, Y., Kim, J., Shin, K., & Jo, G. S. (2009). Recommendation system using location-based ontology on wireless internet: An example of collective intelligence by using 'mashup' applications. *Expert systems with applications*, *36*, 11675–11681.

Anexos

Tabla de similitudes Path

Tabla 6.1: Tabla de similitudes path

ID	GENERO 1	GENERO 2	SIMILITUD
1	action	action	1
2	action	adventure	0.0625
3	action	animation	0.055555556
4	action	biography	0.076923077
5	action	comedy	0.083333333
6	action	crime	0.0625
7	action	documentary	0.066666667
8	action	drama	0.071428571
9	action	family	0.071428571
10	action	fantasy	0.2
11	action	history	0.083333333
12	action	horror	0.083333333
13	action	music	0.1
14	action	musical	0.066666667
15	action	mystery	0.25
16	action	romance	0.166666667
17	action	fiction	0.25
18	action	sport	0.066666667
19	action	thriller	0.2
20	action	war	0.071428571
21	action	western	0.0625
22	adventure	action	0.0625

23	adventure	adventure	1
24	adventure	animation	0.111111111
25	adventure	biography	0.071428571
26	adventure	comedy	0.076923077
27	adventure	crime	0.142857143
28	adventure	documentary	0.1
29	adventure	drama	0.111111111
30	adventure	family	0.076923077
31	adventure	fantasy	0.071428571
32	adventure	history	0.076923077
33	adventure	horror	0.090909091
34	adventure	music	0.090909091
35	adventure	musical	0.1
36	adventure	mystery	0.066666667
37	adventure	romance	0.066666667
38	adventure	fiction	0.076923077
39	adventure	sport	0.166666667
40	adventure	thriller	0.0625
41	adventure	war	0.125
42	adventure	western	0.090909091
43	animation	action	0.055555556
44	animation	adventure	0.111111111
45	animation	animation	1
46	animation	biography	0.0625
47	animation	comedy	0.066666667
48	animation	crime	0.111111111
49	animation	documentary	0.083333333
50	animation	drama	0.090909091
51	animation	family	0.066666667
52	animation	fantasy	0.0625
53	animation	history	0.066666667
54	animation	horror	0.076923077
55	animation	music	0.076923077
56	animation	musical	0.083333333
57	animation	mystery	0.058823529
58	animation	romance	0.058823529

59	animation	fiction	0.066666667
60	animation	sport	0.125
61	animation	thriller	0.055555556
62	animation	war	0.1
63	animation	western	0.076923077
64	biography	action	0.076923077
65	biography	adventure	0.071428571
66	biography	animation	0.0625
67	biography	biography	1
68	biography	comedy	0.1
69	biography	crime	0.071428571
70	biography	documentary	0.076923077
71	biography	drama	0.083333333
72	biography	family	0.083333333
73	biography	fantasy	0.090909091
74	biography	history	0.5
75	biography	horror	0.1
76	biography	music	0.125
77	biography	musical	0.076923077
78	biography	mystery	0.083333333
79	biography	romance	0.083333333
80	biography	fiction	0.1
81	biography	sport	0.076923077
82	biography	thriller	0.076923077
83	biography	war	0.083333333
84	biography	western	0.071428571
85	comedy	action	0.083333333
86	comedy	adventure	0.076923077
87	comedy	animation	0.066666667
88	comedy	biography	0.1
89	comedy	comedy	1
90	comedy	crime	0.076923077
91	comedy	documentary	0.083333333
92	comedy	drama	0.090909091
93	comedy	family	0.090909091
94	comedy	fantasy	0.1

95	comedy	history	0.111111111
96	comedy	horror	0.111111111
97	comedy	music	0.142857143
98	comedy	musical	0.083333333
99	comedy	mystery	0.090909091
100	comedy	romance	0.090909091
101	comedy	fiction	0.111111111
102	comedy	sport	0.083333333
103	comedy	thriller	0.083333333
104	comedy	war	0.090909091
105	comedy	western	0.076923077
106	crime	action	0.0625
107	crime	adventure	0.142857143
108	crime	animation	0.111111111
109	crime	biography	0.071428571
110	crime	comedy	0.076923077
111	crime	crime	1
112	crime	documentary	0.1
113	crime	drama	0.111111111
114	crime	family	0.076923077
115	crime	fantasy	0.071428571
116	crime	history	0.076923077
117	crime	horror	0.090909091
118	crime	music	0.090909091
119	crime	musical	0.1
120	crime	mystery	0.066666667
121	crime	romance	0.066666667
122	crime	fiction	0.076923077
123	crime	sport	0.166666667
124	crime	thriller	0.0625
125	crime	war	0.125
126	crime	western	0.090909091
127	documentary	action	0.066666667
128	documentary	adventure	0.1
129	documentary	animation	0.083333333
130	documentary	biography	0.076923077

131	documentary	comedy	0.083333333
132	documentary	crime	0.1
133	documentary	documentary	1
134	documentary	drama	0.125
135	documentary	family	0.083333333
136	documentary	fantasy	0.076923077
137	documentary	history	0.083333333
138	documentary	horror	0.1
139	documentary	music	0.1
140	documentary	musical	0.333333333
141	documentary	mystery	0.071428571
142	documentary	romance	0.071428571
143	documentary	fiction	0.083333333
144	documentary	sport	0.111111111
145	documentary	thriller	0.066666667
146	documentary	war	0.125
147	documentary	western	0.25
148	drama	action	0.071428571
149	drama	adventure	0.111111111
150	drama	animation	0.090909091
151	drama	biography	0.083333333
152	drama	comedy	0.090909091
153	drama	crime	0.111111111
154	drama	documentary	0.125
155	drama	drama	1
156	drama	family	0.090909091
157	drama	fantasy	0.083333333
158	drama	history	0.090909091
159	drama	horror	0.111111111
160	drama	music	0.111111111
161	drama	musical	0.125
162	drama	mystery	0.076923077
163	drama	romance	0.076923077
164	drama	fiction	0.090909091
165	drama	sport	0.125
166	drama	thriller	0.071428571

167	drama	war	0.142857143
168	drama	western	0.111111111
169	family	action	0.071428571
170	family	adventure	0.076923077
171	family	animation	0.066666667
172	family	biography	0.083333333
173	family	comedy	0.090909091
174	family	crime	0.076923077
175	family	documentary	0.083333333
176	family	drama	0.090909091
177	family	family	1
178	family	fantasy	0.083333333
179	family	history	0.090909091
180	family	horror	0.111111111
181	family	music	0.111111111
182	family	musical	0.083333333
183	family	mystery	0.076923077
184	family	romance	0.076923077
185	family	fiction	0.090909091
186	family	sport	0.083333333
187	family	thriller	0.071428571
188	family	war	0.090909091
189	family	western	0.076923077
190	fantasy	action	0.2
191	fantasy	adventure	0.071428571
192	fantasy	animation	0.0625
193	fantasy	biography	0.090909091
194	fantasy	comedy	0.1
195	fantasy	crime	0.071428571
196	fantasy	documentary	0.076923077
197	fantasy	drama	0.083333333
198	fantasy	family	0.083333333
199	fantasy	fantasy	1
200	fantasy	history	0.1
201	fantasy	horror	0.1
202	fantasy	music	0.125

203	fantasy	musical	0.076923077
204	fantasy	mystery	0.25
205	fantasy	romance	0.25
206	fantasy	fiction	0.5
207	fantasy	sport	0.076923077
208	fantasy	thriller	0.2
209	fantasy	war	0.083333333
210	fantasy	western	0.071428571
211	history	action	0.083333333
212	history	adventure	0.076923077
213	history	animation	0.066666667
214	history	biography	0.5
215	history	comedy	0.111111111
216	history	crime	0.076923077
217	history	documentary	0.083333333
218	history	drama	0.090909091
219	history	family	0.090909091
220	history	fantasy	0.1
221	history	history	1
222	history	horror	0.111111111
223	history	music	0.142857143
224	history	musical	0.083333333
225	history	mystery	0.090909091
226	history	romance	0.090909091
227	history	fiction	0.111111111
228	history	sport	0.083333333
229	history	thriller	0.083333333
230	history	war	0.090909091
231	history	western	0.076923077
232	horror	action	0.083333333
233	horror	adventure	0.090909091
234	horror	animation	0.076923077
235	horror	biography	0.1
236	horror	comedy	0.111111111
237	horror	crime	0.090909091
238	horror	documentary	0.1

239	horror	drama	0.111111111
240	horror	family	0.111111111
241	horror	fantasy	0.1
242	horror	history	0.111111111
243	horror	horror	1
244	horror	music	0.142857143
245	horror	musical	0.1
246	horror	mystery	0.090909091
247	horror	romance	0.090909091
248	horror	fiction	0.111111111
249	horror	sport	0.1
250	horror	thriller	0.083333333
251	horror	war	0.111111111
252	horror	western	0.090909091
253	music	action	0.1
254	music	adventure	0.090909091
255	music	animation	0.076923077
256	music	biography	0.125
257	music	comedy	0.142857143
258	music	crime	0.090909091
259	music	documentary	0.1
260	music	drama	0.111111111
261	music	family	0.111111111
262	music	fantasy	0.125
263	music	history	0.142857143
264	music	horror	0.142857143
265	music	music	1
266	music	musical	0.1
267	music	mystery	0.111111111
268	music	romance	0.111111111
269	music	fiction	0.142857143
270	music	sport	0.1
271	music	thriller	0.1
272	music	war	0.111111111
273	music	western	0.090909091
274	musical	action	0.066666667

275	musical	adventure	0.1
276	musical	animation	0.083333333
277	musical	biography	0.076923077
278	musical	comedy	0.083333333
279	musical	crime	0.1
280	musical	documentary	0.333333333
281	musical	drama	0.125
282	musical	family	0.083333333
283	musical	fantasy	0.076923077
284	musical	history	0.083333333
285	musical	horror	0.1
286	musical	music	0.1
287	musical	musical	1
288	musical	mystery	0.071428571
289	musical	romance	0.071428571
290	musical	fiction	0.083333333
291	musical	sport	0.111111111
292	musical	thriller	0.066666667
293	musical	war	0.125
294	musical	western	0.25
295	mystery	action	0.25
296	mystery	adventure	0.066666667
297	mystery	animation	0.058823529
298	mystery	biography	0.083333333
299	mystery	comedy	0.090909091
300	mystery	crime	0.066666667
301	mystery	documentary	0.071428571
302	mystery	drama	0.076923077
303	mystery	family	0.076923077
304	mystery	fantasy	0.25
305	mystery	history	0.090909091
306	mystery	horror	0.090909091
307	mystery	music	0.111111111
308	mystery	musical	0.071428571
309	mystery	mystery	1
310	mystery	romance	0.2

311	mystery	fiction	0.333333333
312	mystery	sport	0.071428571
313	mystery	thriller	0.25
314	mystery	war	0.076923077
315	mystery	western	0.066666667
316	romance	action	0.166666667
317	romance	adventure	0.066666667
318	romance	animation	0.058823529
319	romance	biography	0.083333333
320	romance	comedy	0.090909091
321	romance	crime	0.066666667
322	romance	documentary	0.071428571
323	romance	drama	0.076923077
324	romance	family	0.076923077
325	romance	fantasy	0.25
326	romance	history	0.090909091
327	romance	horror	0.090909091
328	romance	music	0.111111111
329	romance	musical	0.071428571
330	romance	mystery	0.2
331	romance	romance	1
332	romance	fiction	0.333333333
333	romance	sport	0.071428571
334	romance	thriller	0.166666667
335	romance	war	0.076923077
336	romance	western	0.066666667
337	fiction	action	0.25
338	fiction	adventure	0.076923077
339	fiction	animation	0.066666667
340	fiction	biography	0.1
341	fiction	comedy	0.111111111
342	fiction	crime	0.076923077
343	fiction	documentary	0.083333333
344	fiction	drama	0.090909091
345	fiction	family	0.090909091
346	fiction	fantasy	0.5

347	fiction	history	0.111111111
348	fiction	horror	0.111111111
349	fiction	music	0.142857143
350	fiction	musical	0.083333333
351	fiction	mystery	0.333333333
352	fiction	romance	0.333333333
353	fiction	fiction	1
354	fiction	sport	0.083333333
355	fiction	thriller	0.25
356	fiction	war	0.090909091
357	fiction	western	0.076923077
358	sport	action	0.066666667
359	sport	adventure	0.166666667
360	sport	animation	0.125
361	sport	biography	0.076923077
362	sport	comedy	0.083333333
363	sport	crime	0.166666667
364	sport	documentary	0.111111111
365	sport	drama	0.125
366	sport	family	0.083333333
367	sport	fantasy	0.076923077
368	sport	history	0.083333333
369	sport	horror	0.1
370	sport	music	0.1
371	sport	musical	0.111111111
372	sport	mystery	0.071428571
373	sport	romance	0.071428571
374	sport	fiction	0.083333333
375	sport	sport	1
376	sport	thriller	0.066666667
377	sport	war	0.142857143
378	sport	western	0.1
379	thriller	action	0.2
380	thriller	adventure	0.0625
381	thriller	animation	0.055555556
382	thriller	biography	0.076923077

383	thriller	comedy	0.083333333
384	thriller	crime	0.0625
385	thriller	documentary	0.066666667
386	thriller	drama	0.071428571
387	thriller	family	0.071428571
388	thriller	fantasy	0.2
389	thriller	history	0.083333333
390	thriller	horror	0.083333333
391	thriller	music	0.1
392	thriller	musical	0.066666667
393	thriller	mystery	0.25
394	thriller	romance	0.166666667
395	thriller	fiction	0.25
396	thriller	sport	0.066666667
397	thriller	thriller	1
398	thriller	war	0.071428571
399	thriller	western	0.0625
400	war	action	0.071428571
401	war	adventure	0.125
402	war	animation	0.1
403	war	biography	0.083333333
404	war	comedy	0.090909091
405	war	crime	0.125
406	war	documentary	0.125
407	war	drama	0.142857143
408	war	family	0.090909091
409	war	fantasy	0.083333333
410	war	history	0.090909091
411	war	horror	0.111111111
412	war	music	0.111111111
413	war	musical	0.125
414	war	mystery	0.076923077
415	war	romance	0.076923077
416	war	fiction	0.090909091
417	war	sport	0.142857143
418	war	thriller	0.071428571

419	war	war	1
420	war	western	0.111111111
421	western	action	0.0625
422	western	adventure	0.090909091
423	western	animation	0.076923077
424	western	biography	0.071428571
425	western	comedy	0.076923077
426	western	crime	0.090909091
427	western	documentary	0.25
428	western	drama	0.111111111
429	western	family	0.076923077
430	western	fantasy	0.071428571
431	western	history	0.076923077
432	western	horror	0.090909091
433	western	music	0.090909091
434	western	musical	0.25
435	western	mystery	0.066666667
436	western	romance	0.066666667
437	western	fiction	0.076923077
438	western	sport	0.1
439	western	thriller	0.0625
440	western	war	0.111111111
441	western	western	1

Tabla de similitudes Wup

Tabla 6.2: Tabla de similitudes Wup

ID	GENERO 1	GENERO 2	SIMILITUD
1	action	action	1
2	action	adventure	0.210526316
3	action	animation	0.19047619
4	action	biography	0.333333333
5	action	comedy	0.352941176
6	action	crime	0.210526316
7	action	documentary	0.222222222
8	action	drama	0.235294118
9	action	family	0.235294118
10	action	fantasy	0.777777778
11	action	history	0.352941176
12	action	horror	0.153846154
13	action	music	0.4
14	action	musical	0.222222222
15	action	mystery	0.842105263
16	action	romance	0.736842105
17	action	fiction	0.823529412
18	action	sport	0.222222222
19	action	thriller	0.8
20	action	war	0.235294118
21	action	western	0.210526316
22	adventure	action	0.210526316
23	adventure	adventure	1
24	adventure	animation	0.6
25	adventure	biography	0.235294118
26	adventure	comedy	0.25
27	adventure	crime	0.666666667
28	adventure	documentary	0.470588235
29	adventure	drama	0.5
30	adventure	family	0.25
31	adventure	fantasy	0.235294118

32	adventure	history	0.25
33	adventure	horror	0.166666667
34	adventure	music	0.285714286
35	adventure	musical	0.470588235
36	adventure	mystery	0.222222222
37	adventure	romance	0.222222222
38	adventure	fiction	0.25
39	adventure	sport	0.705882353
40	adventure	thriller	0.210526316
41	adventure	war	0.588235294
42	adventure	western	0.444444444
43	animation	action	0.19047619
44	animation	adventure	0.6
45	animation	animation	1
46	animation	biography	0.210526316
47	animation	comedy	0.222222222
48	animation	crime	0.6
49	animation	documentary	0.421052632
50	animation	drama	0.444444444
51	animation	family	0.222222222
52	animation	fantasy	0.210526316
53	animation	history	0.222222222
54	animation	horror	0.142857143
55	animation	music	0.25
56	animation	musical	0.421052632
57	animation	mystery	0.2
58	animation	romance	0.2
59	animation	fiction	0.222222222
60	animation	sport	0.631578947
61	animation	thriller	0.19047619
62	animation	war	0.526315789
63	animation	western	0.4
64	biography	action	0.333333333
65	biography	adventure	0.235294118
66	biography	animation	0.210526316
67	biography	biography	1

68	biography	comedy	0.4
69	biography	crime	0.235294118
70	biography	documentary	0.25
71	biography	drama	0.266666667
72	biography	family	0.266666667
73	biography	fantasy	0.375
74	biography	history	0.933333333
75	biography	horror	0.181818182
76	biography	music	0.461538462
77	biography	musical	0.25
78	biography	mystery	0.352941176
79	biography	romance	0.352941176
80	biography	fiction	0.4
81	biography	sport	0.25
82	biography	thriller	0.333333333
83	biography	war	0.266666667
84	biography	western	0.235294118
85	comedy	action	0.352941176
86	comedy	adventure	0.25
87	comedy	animation	0.222222222
88	comedy	biography	0.4
89	comedy	comedy	1
90	comedy	crime	0.25
91	comedy	documentary	0.266666667
92	comedy	drama	0.285714286
93	comedy	family	0.285714286
94	comedy	fantasy	0.4
95	comedy	history	0.428571429
96	comedy	horror	0.2
97	comedy	music	0.5
98	comedy	musical	0.266666667
99	comedy	mystery	0.375
100	comedy	romance	0.375
101	comedy	fiction	0.428571429
102	comedy	sport	0.266666667
103	comedy	thriller	0.352941176

104	comedy	war	0.285714286
105	comedy	western	0.25
106	crime	action	0.210526316
107	crime	adventure	0.666666667
108	crime	animation	0.6
109	crime	biography	0.235294118
110	crime	comedy	0.25
111	crime	crime	1
112	crime	documentary	0.470588235
113	crime	drama	0.5
114	crime	family	0.25
115	crime	fantasy	0.235294118
116	crime	history	0.25
117	crime	horror	0.166666667
118	crime	music	0.285714286
119	crime	musical	0.470588235
120	crime	mystery	0.222222222
121	crime	romance	0.222222222
122	crime	fiction	0.25
123	crime	sport	0.705882353
124	crime	thriller	0.210526316
125	crime	war	0.588235294
126	crime	western	0.444444444
127	documentary	action	0.222222222
128	documentary	adventure	0.470588235
129	documentary	animation	0.421052632
130	documentary	biography	0.25
131	documentary	comedy	0.266666667
132	documentary	crime	0.470588235
133	documentary	documentary	1
134	documentary	drama	0.533333333
135	documentary	family	0.266666667
136	documentary	fantasy	0.25
137	documentary	history	0.266666667
138	documentary	horror	0.181818182
139	documentary	music	0.307692308

140	documentary	musical	0.888888889
141	documentary	mystery	0.235294118
142	documentary	romance	0.235294118
143	documentary	fiction	0.266666667
144	documentary	sport	0.5
145	documentary	thriller	0.222222222
146	documentary	war	0.533333333
147	documentary	western	0.842105263
148	drama	action	0.235294118
149	drama	adventure	0.5
150	drama	animation	0.444444444
151	drama	biography	0.266666667
152	drama	comedy	0.285714286
153	drama	crime	0.5
154	drama	documentary	0.533333333
155	drama	drama	1
156	drama	family	0.285714286
157	drama	fantasy	0.266666667
158	drama	history	0.285714286
159	drama	horror	0.2
160	drama	music	0.333333333
161	drama	musical	0.533333333
162	drama	mystery	0.25
163	drama	romance	0.25
164	drama	fiction	0.285714286
165	drama	sport	0.533333333
166	drama	thriller	0.235294118
167	drama	war	0.571428571
168	drama	western	0.5
169	family	action	0.235294118
170	family	adventure	0.25
171	family	animation	0.222222222
172	family	biography	0.266666667
173	family	comedy	0.285714286
174	family	crime	0.25
175	family	documentary	0.266666667

176	family	drama	0.285714286
177	family	family	1
178	family	fantasy	0.266666667
179	family	history	0.285714286
180	family	horror	0.2
181	family	music	0.333333333
182	family	musical	0.266666667
183	family	mystery	0.25
184	family	romance	0.25
185	family	fiction	0.285714286
186	family	sport	0.266666667
187	family	thriller	0.235294118
188	family	war	0.285714286
189	family	western	0.25
190	fantasy	action	0.777777778
191	fantasy	adventure	0.235294118
192	fantasy	animation	0.210526316
193	fantasy	biography	0.375
194	fantasy	comedy	0.4
195	fantasy	crime	0.235294118
196	fantasy	documentary	0.25
197	fantasy	drama	0.266666667
198	fantasy	family	0.266666667
199	fantasy	fantasy	1
200	fantasy	history	0.4
201	fantasy	horror	0.181818182
202	fantasy	music	0.461538462
203	fantasy	musical	0.25
204	fantasy	mystery	0.823529412
205	fantasy	romance	0.823529412
206	fantasy	fiction	0.933333333
207	fantasy	sport	0.25
208	fantasy	thriller	0.777777778
209	fantasy	war	0.266666667
210	fantasy	western	0.235294118
211	history	action	0.352941176

212	history	adventure	0.25
213	history	animation	0.222222222
214	history	biography	0.933333333
215	history	comedy	0.428571429
216	history	crime	0.25
217	history	documentary	0.266666667
218	history	drama	0.285714286
219	history	family	0.285714286
220	history	fantasy	0.4
221	history	history	1
222	history	horror	0.2
223	history	music	0.5
224	history	musical	0.266666667
225	history	mystery	0.375
226	history	romance	0.375
227	history	fiction	0.428571429
228	history	sport	0.266666667
229	history	thriller	0.352941176
230	history	war	0.285714286
231	history	western	0.25
232	horror	action	0.153846154
233	horror	adventure	0.166666667
234	horror	animation	0.142857143
235	horror	biography	0.181818182
236	horror	comedy	0.2
237	horror	crime	0.166666667
238	horror	documentary	0.181818182
239	horror	drama	0.2
240	horror	family	0.2
241	horror	fantasy	0.181818182
242	horror	history	0.2
243	horror	horror	1
244	horror	music	0.25
245	horror	musical	0.181818182
246	horror	mystery	0.166666667
247	horror	romance	0.166666667

248	horror	fiction	0.2
249	horror	sport	0.181818182
250	horror	thriller	0.153846154
251	horror	war	0.2
252	horror	western	0.166666667
253	music	action	0.4
254	music	adventure	0.285714286
255	music	animation	0.25
256	music	biography	0.461538462
257	music	comedy	0.5
258	music	crime	0.285714286
259	music	documentary	0.307692308
260	music	drama	0.333333333
261	music	family	0.333333333
262	music	fantasy	0.461538462
263	music	history	0.5
264	music	horror	0.25
265	music	music	1
266	music	musical	0.307692308
267	music	mystery	0.428571429
268	music	romance	0.428571429
269	music	fiction	0.5
270	music	sport	0.307692308
271	music	thriller	0.4
272	music	war	0.333333333
273	music	western	0.285714286
274	musical	action	0.222222222
275	musical	adventure	0.470588235
276	musical	animation	0.421052632
277	musical	biography	0.25
278	musical	comedy	0.266666667
279	musical	crime	0.470588235
280	musical	documentary	0.888888889
281	musical	drama	0.533333333
282	musical	family	0.266666667
283	musical	fantasy	0.25

284	musical	history	0.266666667
285	musical	horror	0.181818182
286	musical	music	0.307692308
287	musical	musical	1
288	musical	mystery	0.235294118
289	musical	romance	0.235294118
290	musical	fiction	0.266666667
291	musical	sport	0.5
292	musical	thriller	0.222222222
293	musical	war	0.533333333
294	musical	western	0.842105263
295	mystery	action	0.842105263
296	mystery	adventure	0.222222222
297	mystery	animation	0.2
298	mystery	biography	0.352941176
299	mystery	comedy	0.375
300	mystery	crime	0.222222222
301	mystery	documentary	0.235294118
302	mystery	drama	0.25
303	mystery	family	0.25
304	mystery	fantasy	0.823529412
305	mystery	history	0.375
306	mystery	horror	0.166666667
307	mystery	music	0.428571429
308	mystery	musical	0.235294118
309	mystery	mystery	1
310	mystery	romance	0.777777778
311	mystery	fiction	0.875
312	mystery	sport	0.235294118
313	mystery	thriller	0.842105263
314	mystery	war	0.25
315	mystery	western	0.222222222
316	romance	action	0.736842105
317	romance	adventure	0.222222222
318	romance	animation	0.2
319	romance	biography	0.352941176

320	romance	comedy	0.375
321	romance	crime	0.222222222
322	romance	documentary	0.235294118
323	romance	drama	0.25
324	romance	family	0.25
325	romance	fantasy	0.823529412
326	romance	history	0.375
327	romance	horror	0.166666667
328	romance	music	0.428571429
329	romance	musical	0.235294118
330	romance	mystery	0.777777778
331	romance	romance	1
332	romance	fiction	0.875
333	romance	sport	0.235294118
334	romance	thriller	0.736842105
335	romance	war	0.25
336	romance	western	0.222222222
337	fiction	action	0.823529412
338	fiction	adventure	0.25
339	fiction	animation	0.222222222
340	fiction	biography	0.4
341	fiction	comedy	0.428571429
342	fiction	crime	0.25
343	fiction	documentary	0.266666667
344	fiction	drama	0.285714286
345	fiction	family	0.285714286
346	fiction	fantasy	0.933333333
347	fiction	history	0.428571429
348	fiction	horror	0.2
349	fiction	music	0.5
350	fiction	musical	0.266666667
351	fiction	mystery	0.875
352	fiction	romance	0.875
353	fiction	fiction	1
354	fiction	sport	0.266666667
355	fiction	thriller	0.823529412

356	fiction	war	0.285714286
357	fiction	western	0.25
358	sport	action	0.222222222
359	sport	adventure	0.705882353
360	sport	animation	0.631578947
361	sport	biography	0.25
362	sport	comedy	0.266666667
363	sport	crime	0.705882353
364	sport	documentary	0.5
365	sport	drama	0.533333333
366	sport	family	0.266666667
367	sport	fantasy	0.25
368	sport	history	0.266666667
369	sport	horror	0.181818182
370	sport	music	0.307692308
371	sport	musical	0.5
372	sport	mystery	0.235294118
373	sport	romance	0.235294118
374	sport	fiction	0.266666667
375	sport	sport	1
376	sport	thriller	0.222222222
377	sport	war	0.625
378	sport	western	0.470588235
379	thriller	action	0.8
380	thriller	adventure	0.210526316
381	thriller	animation	0.19047619
382	thriller	biography	0.333333333
383	thriller	comedy	0.352941176
384	thriller	crime	0.210526316
385	thriller	documentary	0.222222222
386	thriller	drama	0.235294118
387	thriller	family	0.235294118
388	thriller	fantasy	0.777777778
389	thriller	history	0.352941176
390	thriller	horror	0.153846154
391	thriller	music	0.4

392	thriller	musical	0.222222222
393	thriller	mystery	0.842105263
394	thriller	romance	0.736842105
395	thriller	fiction	0.823529412
396	thriller	sport	0.222222222
397	thriller	thriller	1
398	thriller	war	0.235294118
399	thriller	western	0.210526316
400	war	action	0.235294118
401	war	adventure	0.588235294
402	war	animation	0.526315789
403	war	biography	0.266666667
404	war	comedy	0.285714286
405	war	crime	0.588235294
406	war	documentary	0.533333333
407	war	drama	0.571428571
408	war	family	0.285714286
409	war	fantasy	0.266666667
410	war	history	0.285714286
411	war	horror	0.2
412	war	music	0.333333333
413	war	musical	0.533333333
414	war	mystery	0.25
415	war	romance	0.25
416	war	fiction	0.285714286
417	war	sport	0.625
418	war	thriller	0.235294118
419	war	war	1
420	war	western	0.5
421	western	action	0.210526316
422	western	adventure	0.444444444
423	western	animation	0.4
424	western	biography	0.235294118
425	western	comedy	0.25
426	western	crime	0.444444444
427	western	documentary	0.842105263

428	western	drama	0.5
429	western	family	0.25
430	western	fantasy	0.235294118
431	western	history	0.25
432	western	horror	0.166666667
433	western	music	0.285714286
434	western	musical	0.842105263
435	western	mystery	0.222222222
436	western	romance	0.222222222
437	western	fiction	0.25
438	western	sport	0.470588235
439	western	thriller	0.210526316
440	western	war	0.5
441	western	western	1