



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

**ANÁLISIS DE PERFILES DE USUARIO EN REDES  
SOCIALES PARA SU CARACTERIZACIÓN EN EL  
CONTEXTO GEOESPACIAL**

**T E S I S**  
**QUE PARA OBTENER EL GRADO DE MAESTRO EN  
CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA**

**ING. ISAAC HUSÁI ARROYO MEJÍA**

**DIRECTOR DE TESIS:**  
**DR JOSÉ GIOVANNI GUZMÁN LUGO**

**México D.F. Noviembre 2012**







# INSTITUTO POLITÉCNICO NACIONAL

## SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

### ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 10:00 horas del día 16 del mes de noviembre de 2012 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:

**"Análisis de perfiles de usuario en redes sociales para su caracterización en el contexto geoespacial"**

Presentada por el alumno:

**ARROYO**

Apellido paterno

**MEJÍA**

Apellido materno

**ISAAC HUSAI**

Nombre(a)

Con registro:

|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| B | 1 | 0 | 1 | 6 | 3 | 2 |
|---|---|---|---|---|---|---|

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

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

### LA COMISIÓN REVISORA

Director de Tesis

Dr. José Giovanni Guzmán Lugo

Dr. Marco Antonio Moreno Ibarra

Dr. Rolando Quintero Téllez

Dr. Miguel Jesús Torres Ruiz

M. en C. Alejandro Botello Castillo

Dr. Rolando Menchaca Méndez

PRESIDENTE DEL COLEGIO DE PROFESORES

INSTITUTO POLITÉCNICO NACIONAL  
CENTRO DE INVESTIGACIÓN  
EN COMPUTACIÓN  
DIRECCIÓN

Dr. Luis Alfonso Villa Vargas





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

***CARTA CESIÓN DE DERECHOS***

En la Ciudad de México D.F. el día 20 del mes noviembre del año 2012, el que suscribe **Isaac Husaí Arroyo Mejía** alumno del Programa de Maestría en Ciencias de la Computación con número de registro **B101632**, adscrito a Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección del **Dr. José Giovanni Guzmán Lugo** y cede los derechos del trabajo intitulado **“Análisis de perfiles de usuario en redes sociales, para su caracterización en el contexto geoespacial”** 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 **isc.ismejia@gmail.com**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

---

**Isaac Husaí Arroyo Mejía**



---

# RESUMEN

---

La presente tesis propone una metodología para analizar la similitud entre usuarios de una red social, para lo cual hace uso de sus intereses. Se propone encontrar un enfoque para la clasificación de usuarios mediante la semántica de sus intereses.

El caso de estudio se plantea sobre un interés en particular: *la música*, para lo cual se opera sobre dos redes sociales: Facebook y LastFM, la primera engloba a una gran cantidad de usuarios y sus intereses, sin embargo se plantea el uso de LastFM debido a que es una red social especializada en música, la cual sirve como auxiliar al definir una estructura formal que represente las preferencias musicales de un usuario.

Para definir la estructura que represente las preferencias musicales de un usuario se analiza la taxonomía propuesta por Javier Sorella [1], pero también se hace uso de la indexación social propuesta en la folcsonomía de LastFM, esto debido a la naturaleza del trabajo.

Las preferencias musicales del usuario son representadas mediante un vector de diecinueve posiciones, donde cada una corresponde a un género determinado

Para encontrar la similitud entre los usuarios se hace uso del clasificador K-mean, las variables predictorias utilizadas en el algoritmo son los géneros que han sido seleccionados para representar las preferencias musicales del usuario.

La metodología propuesta resulta ser un enfoque novedoso para sistemas de recomendación, debido a que rompe con el paradigma convencional de analizar las conexiones de los nodos de una red social, haciendo uso de principios como la transitividad y hace énfasis en los intereses expuestos por los usuarios.

**Palabras clave:** redes sociales, sistemas de recomendación, facebook.

---

# ABSTRACT

---

The following thesis proposes a methodology to analyze the similarity between members of a social network using their common interests. The proposed approach is to classify users through the semantic of their interests.

The case study is set on a particular interest: the music, two social networks are used for this purpose: Facebook and LastFM, the first involves a large number of users and their interests, however the use of LastFM is considered because it is a social network specializing in music, which serves as an auxiliary to define a formal structure that represents the musical preferences of a user.

To define the structure that represents the musical preferences of a user, the taxonomy proposed by Javier Sorella is analyzed [1], but the social indexing proposal in the folksonomy of LastFM is also used, this due to the nature of work.

The user's musical preferences are represented by a vector of nineteen positions where each corresponds to a specific genre.

To find the similarity between users, a K-mean classifier its used, the predictor variables used in the algorithm are the genres that have been selected to represent the user's musical preferences.

The proposed methodology proves to be a novel approach to recommendation systems, because it breaks with the conventional paradigm of analyzing the connections of nodes in a social network, using principles such as transitivity and emphasizes the interests expressed by users.

**Keywords:** social networks, recommendation system, facebook.

---

# AGRADECIMIENTOS

---

Al Instituto Politécnico Nacional por esta década que me ha regalado invaluables enseñanzas, innumerables momentos de alegría y que ha marcado mi ser con sus principios.

Al Centro de Investigación en Computación, por darme la oportunidad de cruzar aquellas grandes puertas de cristal que miraba desde mi primer día de universidad.

Al laboratorio de Procesamiento Inteligente de Información Geoespacial por albergar este trabajo

Al CONACyT por todo el apoyo que brinda a la educación en nuestro país, por no olvidar ese noble pilar, por brindar los recursos para materializar los sueños de tantos de nosotros.

Al Dr. José Giovanni Guzmán Lugo por su infinita paciencia en el desarrollo de este trabajo, por creer en mí a pesar de los tropiezos, por aceptar dirigir este trabajo, por su guía y sus consejos.

Al Dr. Marco Antonio Moreno Ibarra por compartir esa pasión que significa la música, por abrirme las puertas al laboratorio, por todos y cada uno de los artículos compartidos, por sus aportes a mi investigación, por ser ese soporte de mis ideas, por creer en esta propuesta y por su inagotable buen humor.

Al Dr. Rolando Quintero Téllez por brindar la luz necesaria para terminar de dar forma a este trabajo.

Al Dr. Miguel Jesús Torres Ruiz por liberar el combustible necesario para terminar este trabajo.

Al M. en C. Alejandro Botello Castillo por sus puntuales observaciones, sus acertados cuestionamientos y el tiempo otorgado a este trabajo.

Al Dr. Rolando Menchaca Méndez porque me ha mostrado que existen caminos que terminan por converger.

Al M. en C. Germán Téllez Castillo por las pláticas, consejos y conocimiento compartido en esta trayectoria.

A mi madre por ser mi mejor ejemplo con su dedicación, amor, entrega, principios e inquebrantable esperanza, por ser el ser humano inspirador que es.

A mi padre por su buen humor, por su manera de ser que aporta equilibrio a la mía, por interesarse en las cosas

A mi hermano César Augusto Arroyo Mejía por su rebeldía, su energía, sus risas, sus abrazos, por transitar sin miedo esos caminos que me resultan ajenos, por su seguridad y su alegría.

A Janeth por toda la paciencia, amor, bondad y compañía durante este complicado periodo, por recordarme que el corazón se encuentra del lado izquierdo, por querer hacer de este un mundo mejor.

A mis amigos y compañeros de laboratorio Mario, Ángel y Fernando por compartir esta travesía, por hacer menos pesado el camino, por tanto esfuerzo en conjunto y todos los buenos momentos.

A mis amigos y compañeros de generación Urbelino, Raúl, Teresa por ser ese noble reencuentro con la escuela, por su buena vibra y su esfuerzo constante.

A Sergio Márquez por sus consejos técnicos en momentos complicados del desarrollo de este proyecto.

A Iliac Huerta por los sabios consejos y grandes charlas durante estos años.

A Salvador Liñan, Antonio Escobar y Osiris Méndez por las facilidades brindadas para continuar este proyecto y llevarlo a buen puerto.

A Nora Álvarez, Othon Chicon, Javier Guzmán y Eder Suárez por sus risas, bromas y buenos momentos que me llenaban de energía para seguir adelante.

A Rodrigo Martínez y Eugenio Vargas por su amistad y apoyo incondicional.

A Israel Toledo por su amistad y compartir el desorden y la rebeldía que se requiere para generar nuevas ideas.

A Arturo Cerón por haberme incitado al mundo de la computación.

Al infaltable quinteto Gabriel, Oscar, Victor, Chucho, Israel, por un infinito número de cosas

A mi familia por mantenerse siempre unida

A Dios por tantas dichas y bendiciones.

*Isaac Husai Arroyo Mejía*

# ÍNDICE GENERAL

|                                                              |            |
|--------------------------------------------------------------|------------|
| <b>RESUMEN</b>                                               | <b>I</b>   |
| <b>ABSTRACT</b>                                              | <b>II</b>  |
| <b>AGRADECIMIENTOS</b>                                       | <b>III</b> |
| <b>ÍNDICE DE FIGURAS</b>                                     | <b>VII</b> |
| <b>ÍNDICE DE TABLAS</b>                                      | <b>IX</b>  |
| <b>ÍNDICE DE ECUACIONES</b>                                  | <b>X</b>   |
| <b>CAPÍTULO 1 INTRODUCCIÓN</b>                               | <b>2</b>   |
| 1.1 FUNDAMENTOS DE LA TESIS                                  | 2          |
| 1.2 MOTIVACIÓN                                               | 3          |
| 1.3 JUSTIFICACIÓN                                            | 4          |
| 1.4 OBJETIVOS                                                | 4          |
| 1.4.1 OBJETIVO GENERAL                                       | 4          |
| 1.4.2 OBJETIVOS ESPECÍFICOS                                  | 4          |
| 1.5 ALCANCE DE LA TESIS                                      | 5          |
| 1.6 ORGANIZACIÓN DE LA TESIS                                 | 5          |
| <b>CAPÍTULO 2 ESTADO DEL ARTE</b>                            | <b>8</b>   |
| 2.1 INTRODUCCIÓN                                             | 8          |
| 2.2 MARCO TEÓRICO                                            | 8          |
| 2.2.1 REDES SOCIALES                                         | 8          |
| 2.2.2 ANÁLISIS EN REDES SOCIALES                             | 13         |
| 2.2.3 CLASIFICACIÓN                                          | 14         |
| 2.2.4 APRENDIZAJE                                            | 15         |
| 2.3 TRABAJOS RELACIONADOS                                    | 21         |
| 2.3.1 GROUP RECOMMENDATION SYSTEM FOR FACEBOOK               | 21         |
| 2.3.2 USING SOCIAL NETWORKS TO ORGANIZE RESEARCHER COMMUNITY | 22         |
| 2.4 CONCLUSIONES                                             | 24         |

|                                                    |           |
|----------------------------------------------------|-----------|
| <b>CAPÍTULO 3 METODOLOGÍA PROPUESTA</b>            | <b>26</b> |
| 3.1 INTRODUCCIÓN                                   | 26        |
| 3.2 ARQUITECTURA DEL SISTEMA                       | 26        |
| 3.2.1 RECOLECCIÓN DE INFORMACIÓN                   | 28        |
| 3.2.2 EXTRACCIÓN DE CARACTERÍSTICAS                | 29        |
| 3.2.3 CLASIFICACIÓN DE USUARIOS                    | 44        |
| 3.3 DIAGRAMAS DE CASO DE USO                       | 49        |
| 3.3.1 DIAGRAMA DE CASO DE USO GENERAL              | 49        |
| 3.3.2 DIAGRAMA CASO DE USO RECOPIRAR INFORMACIÓN.  | 50        |
| 3.3.3 DIAGRAMA CASO DE USO EXTRAER CARACTERÍSTICAS | 51        |
| 3.3.4 DIAGRAMA CASO DE USO CLASIFICAR              | 52        |
| 3.3.5 DIAGRAMA CASO DE USO CLASIFICAR POR GÉNERO   | 53        |
| 3.4 DIAGRAMAS DE SECUENCIA GENERAL                 | 54        |
| 3.5 CONCLUSIONES                                   | 54        |
| <b>CAPÍTULO 4 PRUEBAS Y RESULTADOS</b>             | <b>56</b> |
| 4.1 INTRODUCCIÓN                                   | 56        |
| 4.2 PRUEBAS DE CASOS DE USO                        | 62        |
| 4.2.1 PRUEBA PARA RECOPIRAR INFORMACIÓN            | 62        |
| 4.2.2 PRUEBA PARA EXTRAER CARACTERÍSTICAS          | 64        |
| 4.2.3 PRUEBA PARA CLASIFICAR USUARIOS              | 66        |
| 4.2.4 PRUEBA PARA CLASIFICAR USUARIOS POR GÉNERO   | 72        |
| <b>CAPÍTULO 5 CONCLUSIONES</b>                     | <b>76</b> |
| 5.1 INTRODUCCIÓN                                   | 76        |
| 5.2 TRABAJOS FUTUROS                               | 76        |
| <b>REFERENCIAS</b>                                 | <b>79</b> |

# ÍNDICE DE FIGURAS

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Figura 2.1 Llamada por ID mediante OpenGraph                              | 9  |
| Figura 2.2 Llamada por usuario mediante OpenGraph                         | 10 |
| Figura 2.3 Jerarquía de aprendizaje                                       | 15 |
| Figura 2.4 Notación algoritmo K-NN                                        | 16 |
| Figura 2.5 Pseudocódigo algoritmo K-NN                                    | 17 |
| Figura 2.6 Muestra gráfica de algoritmo K-NN                              | 17 |
| Figura 2.7 K-NN con distancia media                                       | 19 |
| Figura 2.8 K-NN con ponderación de casos seleccionados                    | 20 |
| Figura 2.9 Muestra gráfica de etapas K-mean                               | 21 |
| Figura 2.10 Group Recommendation System for Facebook Architecture         | 22 |
| Figura 3.1 Arquitectura del sistema                                       | 27 |
| Figura 3.2 Taxonomía de la música por Sorella                             | 29 |
| Figura 3.3 Principales etiquetas utilizadas en LastFM (folcsonomia)       | 30 |
| Figura 3.4 Salida en formato JSON etiquetas de LastFM                     | 31 |
| Figura 3.5 Salida género alternativo                                      | 34 |
| Figura 3.6 Salida género banda                                            | 34 |
| Figura 3.7 Salida género britpop                                          | 35 |
| Figura 3.8 Salida género clásico                                          | 35 |
| Figura 3.9 Salida género country                                          | 36 |
| Figura 3.10 Salida género grunge                                          | 36 |
| Figura 3.11 Salida género Hip-Hop                                         | 37 |
| Figura 3.12 Salida género indie                                           | 37 |
| Figura 3.13 Salida género jazz                                            | 38 |
| Figura 3.14 Salida género metal                                           | 38 |
| Figura 3.15 Salida género pop                                             | 39 |
| Figura 3.16 Salida género punk                                            | 39 |
| Figura 3.17 Salida género rap                                             | 40 |
| Figura 3.18 Salida género reggae                                          | 40 |
| Figura 3.19 Salida género rock                                            | 41 |
| Figura 3.20 Salida género blues                                           | 41 |
| Figura 3.21 Salida género electrónica                                     | 42 |
| Figura 3.22 Arquitectura I                                                | 47 |
| Figura 3.23 Arquitectura II                                               | 48 |
| Figura 3.24 Diagrama caso de uso general                                  | 49 |
| Figura 3.25 Diagrama caso de uso recopilar información                    | 50 |
| Figura 3.26 Diagrama caso de uso extraer características                  | 51 |
| Figura 3.27 Diagrama caso de uso clasificar                               | 52 |
| Figura 3.28 Diagrama caso de uso clasificar por género                    | 53 |
| Figura 3.29 Diagrama de secuencia                                         | 54 |
| Figura 4.1 Vectores característicos normalizados ordenados por distancia  | 58 |
| Figura 4.2 Vectores más cercanos y más lejanos según distancia euclidiana | 59 |
| Figura 4.3 Total de géneros en gráfica                                    | 59 |

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| <i>Figura 4.4 Recopilación implícita</i>                                          | 62 |
| <i>Figura 4.5 Recopilación explícita</i>                                          | 63 |
| <i>Figura 4.6 Recopilación explícita II</i>                                       | 63 |
| <i>Figura 4.7 Recopilación explícita III</i>                                      | 64 |
| <i>Figura 4.8 Extracción de prueba</i>                                            | 64 |
| <i>Figura 4.9 Porcentaje de géneros de usuario</i>                                | 65 |
| <i>Figura 4.10 Vectores característicos de usuario</i>                            | 65 |
| <i>Figura 4.11 Prueba de clasificación</i>                                        | 66 |
| <i>Figura 4.12 Usuarios con menor distancia Euclidiana</i>                        | 67 |
| <i>Figura 4.13 Usuarios con mayor distancia Euclidiana</i>                        | 67 |
| <i>Figura 4.14 Ejemplo de clasificación con grupos</i>                            | 68 |
| <i>Figura 4.15 Superficie de grupo 1</i>                                          | 69 |
| <i>Figura 4.16 Superficie de grupo 2</i>                                          | 69 |
| <i>Figura 4.17 Superficie de preferencias por municipio</i>                       | 70 |
| <i>Figura 4.18 Miembros de grupos visualizados en mapa</i>                        | 71 |
| <i>Figura 4.19 Gráfica jazz para comparar U1 y U4</i>                             | 72 |
| <i>Figura 4.20 Comparativa de salida reciprocidad contra distancia euclidiana</i> | 73 |

# ÍNDICE DE TABLAS

|                                                                          |    |
|--------------------------------------------------------------------------|----|
| <i>Tabla 2.1 Conexiones permitidas por OpenGraph</i>                     | 10 |
| <i>Tabla 2.2 Información personal a recopilar</i>                        | 24 |
| <i>Tabla 3.1 Ejemplos de representantes de género</i>                    | 32 |
| <i>Tabla 3.2 Estadísticas de géneros en LastFM</i>                       | 33 |
| <i>Tabla 3.3 Etiquetas de prueba</i>                                     | 43 |
| <i>Tabla 3.4 Estructura para almacenar vector característico</i>         | 44 |
| <i>Tabla 3.5 Ejemplo de entrada para clasificación</i>                   | 46 |
| <i>Tabla 3.6 Ejemplo de entrada para clasificación</i>                   | 46 |
| <i>Tabla 3.7 Descripción caso de uso recopilar información</i>           | 50 |
| <i>Tabla 3.8 Descripción caso de uso extraer características</i>         | 51 |
| <i>Tabla 3.9 Descripción caso de uso clasificar usuarios</i>             | 52 |
| <i>Tabla 3.10 Descripción caso de uso clasificar usuarios por género</i> | 53 |
| <i>Tabla 4.1 Usuarios por etapa</i>                                      | 56 |
| <i>Tabla 4.2 Datos de usuarios de prueba</i>                             | 57 |
| <i>Tabla 4.3 Vectores característicos de prueba</i>                      | 60 |
| <i>Tabla 4.4 Vectores característicos normalizados</i>                   | 61 |

# ÍNDICE DE ECUACIONES

|                                                            |    |
|------------------------------------------------------------|----|
| <i>Ecuación 2.1 Entrada K-NN</i>                           | 16 |
| <i>Ecuación 3.1 Información implícita</i>                  | 28 |
| <i>Ecuación 3.2 Información explícita</i>                  | 28 |
| <i>Ecuación 3.3 Respuesta de conjunto de etiquetas</i>     | 30 |
| <i>Ecuación 3.4 Vector característico de la extracción</i> | 32 |
| <i>Ecuación 3.5 Vector normalizado</i>                     | 45 |
| <i>Ecuación 3.6 Distancia Euclidiana</i>                   | 45 |

# CAPÍTULO I

## INTRODUCCIÓN

***"Si fuéramos capaces de unirnos,  
que hermoso y cercano sería el futuro"***  
Ernesto Rafael Guevara de la Serna

# Capítulo 1 Introducción

## 1.1 Fundamentos de la tesis

En la actualidad existe un gran número de usuarios que hacen uso de Internet en sus labores cotidianas, lo que ha dado pauta a la generación de estructuras sociales compuestas por grupos de personas, las cuales están conectadas por uno o varios tipos de relaciones, tales como amistad, parentesco, intereses comunes, ubicación geográfica, entre otras; dichas estructuras son conocidas como redes sociales en Internet (RSI) las cuales a su vez otorgan la posibilidad de que dos partes inicien comunicación sin que existan valores en común entre ambas [2], dichas interacciones se pueden dar entre cualquier par de usuarios en el mundo, por otro lado permiten que la diversidad y la colectividad se conjuguen y en esto radica gran parte de toda la energía que le da vida a los grupos humanos que las conforman, por lo tanto las bondades que brindan las RSI produce que los sitios de este tipo tales como MySpace, LastFM, Facebook y Twitter sean más populares y diversos cada día [3]. No obstante al existir tantos grupos y usuarios en estos portales se hace necesario generar un mecanismo que ayude a identificar de manera objetiva y puntual similitudes entre un individuo y el resto de los miembros de la red, teniendo como fin la formación de grupos cuyos usuarios posean características semejantes.

En años recientes, un creciente número de personas han comenzado a utilizar dispositivos con GPS para registrar sus movimientos al aire libre mediante sus trayectorias [4], lo cual ha generado el interés de los sitios de redes sociales por brindar la posibilidad e incentivar a los usuarios a compartir su ubicación geográfica con otros miembros de la misma, propiciando que se definan puntos de interés geográficos clasificados por categorías para que el usuario pueda indicar los lugares en los que ha estado sin necesidad de recurrir a algoritmos que encuentren los puntos de estancia a través de las trayectorias completas registradas mediante su GPS.

En este trabajo se propone hallar usuarios semejantes en una red social analizando la similitud entre los intereses expuestos por el mismo, específicamente nos enfocaremos a los intereses musicales con el objetivo de realizar una clasificación a nivel semántico mediante clases definidas (géneros musicales) de cada una de las instancias definidas (grupos o solistas). Para llevar a cabo este trabajo se plantea el uso de clasificadores, los cuales operen sobre las preferencias musicales del usuario, lo que desemboca en la necesidad de hallar un mecanismo que nos permita generar una relación entre instancia y clase para posteriormente diseñar una estructura formal que represente toda esta información, por otro lado cabe mencionar que para resolver este problema

es necesario la integración de datos de dos fuentes de datos diferentes (redes sociales).

## **1.2 Motivación**

Con la creciente popularización de los servicios basados en internet, la construcción de redes sociales en Internet resulta un proceso innato del ser humano ya que el hombre no es un ser solitario, necesita de los demás, es un ser social por naturaleza [5].

En los últimos años una novedosa y potencial fuente de información geoespacial ha llegado a ser disponible en forma de contenido web por los usuarios [6], con lo cual surge la oportunidad de aprovechar los datos que los usuarios de manera activa están construyendo día con día en los diversos portales de Internet.

Uno de los objetivos de las redes sociales es el construir comunidades de usuarios que tengan intereses similares; sin embargo, esta tarea se complica en función de la cantidad de perfiles que existe en las mismas, por lo cual se vuelve necesario contar con un mecanismo que nos permita identificar grupos de personas afines de manera automatizada mediante el análisis de la información que cada día proporciona cada usuario.

De manera tradicional, los esfuerzos que se han realizado para resolver este problema han sido orientados al análisis de la transitividad o el análisis de las métricas de los grafos que representan las redes sociales, tales como la intermediación, la cual refleja el número de personas que una persona conecta indirectamente a través de sus vínculos directos [7], sin embargo nuestra motivación radica en darle una mayor importancia a los intereses para poder generar grupos objetivos cuya construcción puede resultar en beneficios mercantiles o sociales tales como la detección de mercados meta [8].

Por otro lado el principal interés en este trabajo radica en generar un aporte para mantener el sentido de comunidad en una época donde la segregación es un problema que se ha incrementado derivado de conceptos tales como el individualismo, la competitividad y la mercantilización de la vida cotidiana.

### **1.3 Justificación**

Mientras que actualmente existen sistemas de recomendación en las redes sociales, estos trabajan analizando los contactos que comparten dos individuos dejando de lado los intereses en común y los puntos de interés que visitan los usuarios.

Este trabajo busca explotar los conceptos asociados a la información que los usuarios comparten con el fin de que otros individuos puedan identificar sus características y la similitud que tienen con ellos, con lo cual la existencia de una herramienta que nos permita entender y utilizar esta información contribuiría al cumplimiento de los objetivos de una red social.

Al generar una metodología que nos indique cómo proceder en la integración de grupos de usuarios en una red social nos permitiría en primera instancia generar clasificaciones objetivas de usuarios en dicha red, sin embargo se podrían considerar la integración de datos de múltiples redes sociales especializadas tales como LastFm (música), IMDb (cine), Flickr (fotografía), entre otras.

### **1.4 Objetivos**

#### **1.4.1 Objetivo general**

Diseñar e implementar una metodología para la construcción de grupos de usuarios en redes sociales con base en los intereses de los usuarios.

#### **1.4.2 Objetivos específicos**

- Analizar la similitud entre usuarios de redes sociales mediante la semántica de sus intereses.
- Dotar de significado semántico a las instancias de intereses musicales proporcionadas en Facebook a partir de una red social más especializada (LastFM).
- Construir una estructura que represente de manera ponderada los intereses musicales de los usuarios en una red social.
- Construir un prototipo de sistema de clasificación de usuarios.
- Realizar pruebas al sistema con un universo acotado del cual se conozca su similitud musical en el mundo real.

## 1.5 Alcance de la tesis

El alcance de esta tesis consiste en generar una metodología que permita construir grupos de usuarios mediante el uso de un clasificador que utilice como entrada vectores ponderados, los cuales representan los intereses musicales de un individuo, para esto es necesario seleccionar las clases o géneros más generales que puedan representar a la mayor parte de una comunidad.

## 1.6 Organización de la tesis

Como se mostró durante este capítulo se hace referencia a los fundamentos del estudio, la motivación, justificación, objetivos y alcances de la tesis. De este modo, se proporciona el marco base del trabajo desarrollado.

El resto de la tesis está organizada mediante los siguientes capítulos:

**Capítulo 2. Estado del arte y fundamentos teórico.** Se sintetiza la recopilación de antecedentes y se describe el estado del arte de los temas relacionados con la problemática definida, los temas involucrados son: sistemas de recomendación de usuarios, clasificadores, web semántica.

**Capítulo 3. Metodología Propuesta.** Se presenta el diseño e implementación de la metodología propuesta con base en las características mencionadas en capítulos anteriores; a su vez se presentan algunos detalles sobre las tecnologías que componen la solución propuesta.

**Capítulo 4. Pruebas y resultados.** Se presentan las pruebas realizadas al sistema tomando en cuenta un universo delimitado para poder analizar los resultados obtenidos basándose en el caso de estudio.

**Capítulo 5. Conclusiones y trabajos futuros.** Se presentan las conclusiones, haciendo énfasis en las diversas aplicaciones del sistema, finalmente se menciona una propuesta de trabajos futuros.



# CAPÍTULO II

## ESTADO DEL ARTE

***"Somos lo que hacemos  
para cambiar lo que somos"***  
Eduardo Galeano

## **Capítulo 2 Estado del Arte**

### **2.1 Introducción**

El presente capítulo tiene por objetivo presentar el marco teórico que sirve como base para llevar a cabo esta investigación, teniendo como elementos: las redes sociales de manera conceptual, el análisis en redes sociales como trabajo interdisciplinario, la web semántica y los sistemas clasificadores. Posteriormente se destacan algunos de los trabajos relacionados.

### **2.2 Marco teórico**

#### **2.2.1 Redes Sociales**

Los encuentros personales, las llamadas por teléfono, los correos electrónicos y las comunicaciones por video son medios comunes que permiten la interacción entre amigos, familiares y conocidos, sin embargo estos medios tradicionales son utilizados cuando dos partes han encontrado valores en común: intereses, regiones geográficas, parentesco, confianza o conocimiento mutuo [9], mientras que en las redes sociales dos partes pueden iniciar comunicación sin haber encontrado valores en común al compartir libremente su información personal con otros [2].

Una red social generalmente consiste en un gran número de nodos y aristas, donde cada nodo representa a un individuo y cada arista representa algún tipo de relación entre ambos nodos [10].

Una red social puede ser vista como un sitio web que permite a sus usuarios construir un perfil público dentro de un sistema específico y organizar una lista de contactos con el fin de establecer conexiones para mostrar a los demás y visualizar también las de estos [11].

##### **2.2.1.1 Facebook**

Es una red social fundada en 2004 por Mark Zuckerberg para uso exclusivo de estudiantes de Harvard. A partir del 2006, Facebook ha permitido a cualquier individuo mayor a 13 años crear una cuenta personal en su sitio de manera gratuita. A su vez permite al titular de la cuenta el gestionar un perfil el cual contiene una fotografía que lo represente entre otros datos personales tales como nombre, edad, género, religión, tendencia política, entre otros. Los usuarios de Facebook pueden también buscar en la

red a otros miembros con el fin de compartir intereses en común, opiniones y contenido multimedia [12].

Crear y mantener un perfil e incrementar el número de contactos en la red son las dos principales actividades en Facebook [13].

Los perfiles juegan un rol fundamental ya que no solamente representan una mera auto-presentación, sino también un tipo de requisito previo para autenticar un usuario ante los demás y a su vez presenta una forma de comunicarse con otros.

Esto se debe al hecho de que al crear las propias listas de contactos, cada usuario tiene que enviar una *solicitud de amistad* la cual puede ser aceptada o rechazada por el que la recibe. En otras palabras el emisor tiene que ser reconocido y autenticado por el receptor para obtener acceso a su perfil.

#### 2.2.1.1.1 Graph API

El núcleo de Facebook es el grafo social, la gente y las conexiones que ellos tienen a todo lo que les importa. La Graph API presenta una vista simple y consistente del grafo social de Facebook, de manera uniforme representa los objetos (nodos) en el grafo (personas, fotos, eventos, páginas) y las conexiones entre ellos (amistad, parentesco, contenido compartido, etiquetas de fotos, intereses).

Cada objeto en el grafo social tiene un identificador único (ID). Se puede acceder a las propiedades de un objeto al solicitar <https://graph.facebook.com/ID>. Por ejemplo, la página oficial de la plataforma de Facebook tiene como id 19292868552, por lo cual se puede recuperar las propiedades del objeto en <https://graph.facebook.com/19292868552>, como se muestra en la figura 2.1.

```
{
  "name": "Facebook Platform",
  "website": "http://developers.facebook.com",
  "username": "platform",
  "founded": "May 2007",
  "company_overview": "Facebook Platform enables anyone to build...",
  "mission": "To make the web more open and social.",
  "products": "Facebook Application Programming Interface (API)...",
  "likes": 449921,
  "id": 19292868552,
  "category": "Technology"
}
```

Figura 2.1 Llamada por ID mediante OpenGraph

De manera alternativa, las personas y páginas que tengan nombre de usuario pueden ser accedidas utilizando el nombre de usuario como ID, por ejemplo al realizar la consulta al objeto CIC IPN México cuyo nombre de usuario es cicipn.mexico, la consulta quedaría de la siguiente forma: <https://graph.facebook.com/cicipn.mexico>, el resultado puede visualizarse en la figura 2.2

```
{
  "id": "100003112584389",
  "name": "Cic Ipn Mexico",
  "first_name": "Cic Ipn",
  "middle_name": "Mexico",
  "last_name": "Mexico",
  "link": "https://www.facebook.com/cicipn.mexico",
  "username": "cicipn.mexico",
  "gender": "male",
  "locale": "es_LA"
}
```

Figura 2.2 Llamada por usuario mediante OpenGraph

Cualquier objeto puede ser accedido de la misma forma, algunos tipos de objetos son los siguientes: usuarios, páginas, eventos, grupos, aplicaciones, fotos, álbum de fotos, foto de perfil, videos, notas, checkins entre otros; es preciso aclarar que todas las respuestas son dadas en objetos JSON.

Todos los objetos en el grafo social de Facebook están conectados a otros vía relaciones, por ejemplo isaacarroyo es fan de la página del IPN o marco.morenoibarra es amigo de giovanni.guzman.lugo; estas relaciones son llamadas conexiones en esta API. Las conexiones entre objetos pueden ser examinadas mediando la siguiente estructura de URL: [https://graph.facebook.com/ID/TIPO\\_DE\\_CONEXION](https://graph.facebook.com/ID/TIPO_DE_CONEXION). Algunas conexiones soportadas para personas y páginas se muestran en la tabla 2.1

Tabla 2.1 Conexiones permitidas por OpenGraph

| Tipo de conexión | URL estructurada para perfil propio                                                                                         |
|------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Amigos           | <a href="https://graph.facebook.com/me/friends?access_token=...">https://graph.facebook.com/me/friends?access_token=...</a> |
| Muro             | <a href="https://graph.facebook.com/me/feed?access_token=...">https://graph.facebook.com/me/feed?access_token=...</a>       |
| Likes            | <a href="https://graph.facebook.com/me/likes?access_token=...">https://graph.facebook.com/me/likes?access_token=...</a>     |
| Películas        | <a href="https://graph.facebook.com/me/movies?access_token=...">https://graph.facebook.com/me/movies?access_token=...</a>   |
| Música           | <a href="https://graph.facebook.com/me/music?access_token=...">https://graph.facebook.com/me/music?access_token=...</a>     |
| Libros           | <a href="https://graph.facebook.com/me/books?access_token=...">https://graph.facebook.com/me/books?access_token=...</a>     |

|                      |                                                                                                                                     |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| Permisos             | <a href="https://graph.facebook.com/me/permissions?access_token=...">https://graph.facebook.com/me/permissions?access_token=...</a> |
| Álbumes de fotos     | <a href="https://graph.facebook.com/me/albums?access_token=...">https://graph.facebook.com/me/albums?access_token=...</a>           |
| Grupos               | <a href="https://graph.facebook.com/me/groups?access_token=...">https://graph.facebook.com/me/groups?access_token=...</a>           |
| Ubicación geográfica | <a href="https://graph.facebook.com/me/locations?access_token=...">https://graph.facebook.com/me/locations?access_token=...</a>     |
| Checkins             | <a href="https://graph.facebook.com/me/checkins?access_token=...">https://graph.facebook.com/me/checkins?access_token=...</a>       |

La API soporta diferentes tipos de conexiones para diferentes tipos de objetos, por ejemplo podemos obtener la lista de personas que asistirán a un evento determinado, supongamos el evento Facebook Developer Garage (ID #331218348435) mediante una petición que cumpla con la estructura mencionada: [https://graph.facebook.com/331218348435/attending?access\\_token=...](https://graph.facebook.com/331218348435/attending?access_token=...)

#### **2.2.1.1.1.1 Autorización**

La API permite fácilmente acceder a toda la información pública de un objeto. Por ejemplo, <https://graph.facebook.com/isaac.arroyo> (Isaac Husaí Arroyo Mejía) retorna toda la información pública sobre Isaac (nombre, apellido, foto de perfil y lo públicamente disponible)

Para obtener información adicional sobre un usuario, primero se debe contar con su permiso. En un primer nivel se necesita contar con un token de acceso para el usuario de Facebook, después de obtener el token de acceso para el usuario se puede proceder a realizar las solicitudes autorizadas incluyendo el token en la petición realizada mediante la API.

La API utiliza OAuth 2.0 para el proceso de autorización, el cual es un protocolo estándar abierto, que permite autorización segura y simple para aplicaciones de escritorio, móviles, y web. OAuth permite a los usuarios compartir su información privada (fotos, videos, listas de contactos, etcétera) almacenada en un sitio con otro sin necesidad de proporcionar sus credenciales de acceso. [14]

Cabe señalar que los tokens de acceso que se obtienen, tienen diferente nivel de acceso los cuales deben ser solicitados según el tipo de objeto al cual deseemos tener acceso [15].

#### **2.2.1.2 LastFM**

LastFM es un sitio web de música fundado en 2002 por Felix Miller, Martin Stiksel, Michael Breidenbruecker y Thomas Willomitzer los cuales provienen de Alemania y

Austria; el sitio fue fundado como una comunidad de música (red social) y una estación de radio en línea.

Esta red utiliza un sistema de recomendación de música llamado Audioscrobbler a partir del cual construye perfiles del gusto musical de cada usuario, almacenando los detalles de las canciones escuchadas por él, ya sea a través de las estaciones de radio por internet, las computadoras de los usuarios o los reproductores de música portátiles. Esta información es transferida (“scrobbled”) a la base de datos de LastFM a través de un reproductor de música que soporte de manera nativa la conexión a LastFM (Rdio, Spotify, Clementine, Amarok) o mediante un plugin instalado en el sistema operativo que sirve como puente entre los reproductores de música convencionales y LastFM. Posteriormente los datos son desplegados en las páginas del perfil de usuario y también son recopiladas para crear páginas de referencias de los artistas.

En abril del 2011 LastFM reportó más de 50 billones de scrobbles [16].

#### **2.2.1.2.1 Características**

A continuación se enlistan y explican las características principales de esta red social para las cuentas de usuario gratuitas:

- **Perfil de usuario.** Un usuario puede construir un perfil musical de cualquiera de las siguientes formas: escuchando su colección personal de música en un reproductor con el plugin de audioscrobbler o bien escuchando el servicio de radio por internet de LastFM. Todas las canciones son agregadas para calcular sus recomendaciones musicales a través de las gráficas de sus artistas o canciones favoritas.
- **Grupos.** Permite la formación de grupos entre usuarios con algo en común, los grupos pueden ser enlazados a artistas y países, cualquier usuario puede crear un grupo y agregar miembros. La mayoría de grupos están abiertos a todos los miembros de la red pero la membresía puede estar sujeta a la aprobación del líder del grupo; por otro lado los grupos también tienen un perfil similar al de los usuarios.
- **Eventos.** A partir de octubre del 2006 se agregó una funcionalidad a los eventos que permite al usuario especificar una posición geográfica para conocer conciertos o festivales que puedan ser del interés del usuario.

- **Etiquetas.** LastFm soporta el etiquetado de los artistas, discos y canciones por parte del usuario para crear una folcsonomía de la música, la cual es en contraste con la taxonomía una indexación social, es decir, la clasificación colaborativa por medio de etiquetas simples en un espacio de nombres llano, sin jerarquías ni relaciones de parentesco. Los usuarios pueden navegar mediante las etiquetas las cuales pueden ser por género, humor, característica artística o cualquier otra forma definida por el usuario.
- **Páginas de artistas.** Una vez que un artista ha sido escuchado una o varias ocasiones a través del scrobbled por al menos un usuario. LastFM crea automáticamente una página principal del artista, la cual podrá contener una lista de artistas similares, sus canciones más populares o más importante aún para nuestra investigación, las etiquetas más populares otorgadas a dicho artista las cuales serán utilizadas como parte medular para nuestras clasificaciones.

#### 2.2.1.2.2 LastFM API

La API de LastFM permite llamar a los métodos para recibir una respuesta en objetos JSON. Los métodos se encuentran divididos por las siguientes categorías: disco, artista, autenticación, gráficas, eventos, geo, grupos, etiquetas, canciones, usuarios, lugares [17].

La raíz de esta API se encuentra en <http://ws.audioscrobbler.com/2.0/> sin embargo cabe señalar que no existe ningún SDK oficial en ningún lenguaje para su implementación, sin embargo Matt Oakes escribió una clase en PHP la cual permite el fácil acceso a la API [18].

#### 2.2.2 Análisis en redes sociales

El análisis en una red social consiste en examinar las relaciones y las estructuras de los actores en una red para medir el comportamiento de un actor y su rendimiento [19].

Las metodologías utilizadas para estudiar y analizar las redes sociales pueden ser diferentes, según el propósito del estudio pero son utilizadas de manera general para medir las relaciones y los flujos entre la gente, grupos y las organizaciones [20].

### 2.2.2.1 Métricas en las conexiones de las redes sociales.

A continuación se listan las métricas utilizadas respecto a las conexiones en las redes sociales:

- **Homofilia.** Representa la medida en que los actores forman lazos con otros similares contra los diferentes. La similitud puede ser determinada por género, raza, edad, ocupación, educación, valores o cualquier otra característica sobresaliente [21].
- **Multiplicidad:** Muestra la cantidad de ocasiones que dos actores coinciden en una relación [22], por ejemplo, dos personas que son amigos y también trabajan juntos tendrían una multiplicidad de 2. La multiplicidad es asociada con la fuerza de la relación.
- **Reciprocidad.** El grado en que dos actores tienen recíprocamente amistades o alguna otra interacción [23].
- **Cierre de red.** Es una medida que asume la cerradura en la red de un individuo, es decir, que sus amigos son también amigos lo cual es llamado principio de transitividad.
- **Propincuidad.** Se refiere a la tendencia de los actores a tener más relaciones con otros geográficamente cercanos [23].

### 2.2.3 Clasificación

En la clasificación los datos son objetos caracterizados por atributos que pertenecen a diferentes clases (etiquetas discretas). La meta es inducir un modelo para poder predecir una clase dados los valores de los atributos de un nuevo objeto.

Con la definición anterior un clasificador puede ser visto como un elemento que proporciona una clase etiquetada como salida a partir de un conjunto de características tomadas como entradas; una manera de construir un clasificador es tomar un conjunto de ejemplos etiquetados y tratar de definir una regla que pueda asignar una etiqueta a cualquier otro dato de entrada.

## 2.2.4 Aprendizaje

El aprendizaje se puede definir como cambios adaptativos en un sistema, en el sentido de que le permiten realizar las mismas tareas pero de forma más eficiente y más efectiva la siguiente vez [24]. El aprendizaje ha sido representado en una jerarquía [25], en la cual se encuentra el aprendizaje supervisado el cual es el centro en la presente investigación, dicha jerarquía se muestra en la figura 2.3.

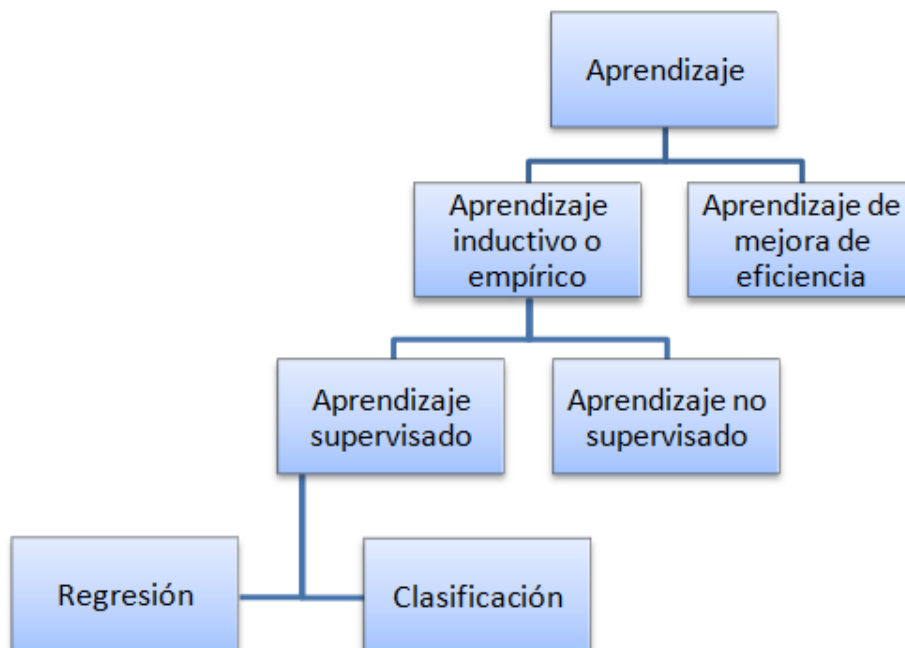


Figura 2.3 Jerarquía de aprendizaje

### 2.2.4.1 Aprendizaje empírico

Se realiza a través de razonamientos de ejemplos suministrados externamente, para producir reglas de carácter general [26].

#### 2.2.4.1.1 Aprendizaje supervisado

El algoritmo de inducción recibe un conjunto de ejemplos, denominado conjunto de entrenamiento, en el que cada instancia consiste en una lista de valores de características y una etiqueta discreta. La tarea del algoritmo es aprender modelos, en amplio sentido, que clasifiquen correctamente instancias sin etiquetar. El término "supervisado" sugiere que algún proceso ha etiquetado las instancias del conjunto de entrenamiento [24].

Para el aprendizaje supervisado existen diversos algoritmos para ser utilizados como clasificadores, sin embargo la presente tesis se centra en el uso del clasificador K-NN debido a que no buscamos predecir a cual clase pertenece el nuevo usuario sino obtener como resultado una lista ordenada según la distancia entre él y las demás personas en la red.

#### 2.2.4.1.1.1 Clasificador supervisado K-NN

El clasificador K-NN (K-Nearest Neighbour) tiene como idea fundamental que un nuevo caso se va a clasificar en la clase más frecuente a la que pertenezcan sus K vecinos más cercanos [27].

##### 2.2.4.1.1.1.1 Algoritmo K-NN básico

La notación a utilizar en este algoritmo se muestra en figura 2.4

|              |          | $X_1$       | ... | $X_j$       | ... | $X_n$       | $C$      |
|--------------|----------|-------------|-----|-------------|-----|-------------|----------|
| $(x_1, c_1)$ | 1        | $x_{11}$    | ... | $x_{1j}$    | ... | $x_{1n}$    | $c_1$    |
| $\vdots$     | $\vdots$ | $\vdots$    |     | $\vdots$    |     | $\vdots$    | $\vdots$ |
| $(x_i, c_i)$ | $i$      | $x_{i1}$    | ... | $x_{ij}$    | ... | $x_{in}$    | $c_i$    |
| $\vdots$     | $\vdots$ | $\vdots$    |     | $\vdots$    |     | $\vdots$    | $\vdots$ |
| $(x_N, c_N)$ | $N$      | $x_{N1}$    | ... | $x_{Nj}$    | ... | $x_{Nn}$    | $c_N$    |
| $x$          | $N + 1$  | $x_{N+1,1}$ | ... | $x_{N+1,j}$ | ... | $x_{N+1,n}$ | ?        |

Figura 2.4 Notación algoritmo K-NN

La figura anterior indica un archivo de  $N$  casos, cada uno de los cuales está caracterizado por  $n$  variables predictorias,  $X_1, \dots, X_n$  y una variable a predecir, la clase  $C_{N+1}$ .

Los  $N$  casos se denotan por la ecuación 2.1

$$\begin{aligned}
 (x_1, c_1), \dots, (x_N, c_N) & \quad \text{Donde} \\
 x_i = (x_{i,1} \dots x_{i,n}) & \quad \text{para todo } i = 1, \dots, N \\
 c_i \in \{c^1, \dots, c^m\} & \quad \text{para todo } i = 1, \dots, N
 \end{aligned}$$

Ecuación 2.1 Entrada K-NN

El nuevo caso que se pretende clasificar se denota por  $x_{N+1} = (x_{N+1,1}, \dots, x_{N+1,n})$ .

### 2.2.4.1.1.2 Pseudocódigo algoritmo K-NN

En la Figura 2.5 se presenta un pseudocódigo para el clasificador K-NN básico. Tal y como puede observarse en el mismo, se calculan las distancias de todos los casos ya clasificados al nuevo caso  $X_{N+1}$  que se pretende clasificar. Una vez seleccionados los  $k$  casos más cercanos al nuevo caso,  $X_{N+1}$  según las distancias  $D_x^K$ , a éste se le asignará la clase (valor de la variable C) más frecuente de entre los  $k$  objetos.

---

#### COMIENZO

Entrada:  $D = \{(x_1, c_1), \dots, (x_N, c_N)\}$

$x = (x_1, \dots, x_n)$  nuevo caso a clasificar

PARA todo objeto ya clasificado  $(x_i, c_i)$

calcular  $d_i = d(x_i, x)$

Ordenar  $d_i (i = 1, \dots, N)$  en orden ascendente

Quedarnos con los  $K$  casos  $D_x^K$  ya clasificados más cercanos a  $x$

Asignar a  $x$  la clase más frecuente en  $D_x^K$

FIN

---

Figura 2.5 Pseudocódigo algoritmo K-NN

En la figura 2.6 podemos ver un ejemplo gráfico de lo anterior.

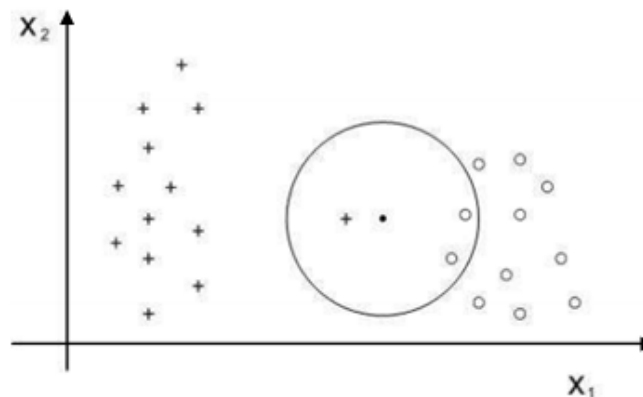


Figura 2.6 Muestra gráfica de algoritmo K-NN

En el ejemplo anterior tenemos 24 casos ya clasificados en dos posibles valores  $m = 2$ . Las variables predictorias son  $X_1$ ,  $X_2$  y se ha seleccionado  $k = 3$ . De los 3 casos ya clasificados que se encuentran más cercanos al nuevo caso a clasificar  $X_{N+1}$  (representado por  $\bullet$ ), dos de ellos pertenecen a la clase  $\circ$ , por lo tanto el clasificador

3-NN (el valor de  $k$  puede ser reemplazado por el número de vecinos seleccionados) predice la clase  $\circ$  para el nuevo caso.

Nótese que el caso más cercano a  $X_{N+1}$  pertenece a la clase  $+$ . Es decir, que si se utiliza un clasificador 1-NN, se hubiese asignado la clase  $+$ .

Conviene aclarar que el clasificador K-NN es un tanto atípico si lo comparamos con el resto de clasificadores, ya que mientras que en el resto de paradigmas, la clasificación de un nuevo caso se lleva a cabo a partir de dos tareas, como son la inducción del modelo clasificatorio y la posterior deducción (o aplicación) sobre el nuevo caso, en el paradigma K-NN al no existir modelo explícito, las dos tareas anteriores se encuentran colapsadas en lo que se acostumbra denominar transducción.

En caso de que se produzca un empate entre dos o más clases, conviene tener una regla heurística para su ruptura, por ejemplo seleccionar la clase con distancia media menor.

### **2.2.4.1.1.1.3 Variantes sobre el algoritmo básico.**

#### **2.2.4.1.1.1.3.1 K-NN con rechazo**

La idea subyacente al K-NN con rechazo es que para poder clasificar un caso debo de tener ciertas garantías. Es por ello por lo que puede ocurrir que un caso quede sin clasificar, si no existen ciertas garantías de que la clase a asignar sea la correcta.

Dos ejemplos utilizados para llevar a cabo clasificaciones con garantías son los siguientes:

- El número de votos obtenidos por la clase deberá superar un umbral prefijado. Si suponemos que trabajamos con  $k = 10$  y  $m = 2$ , dicho umbral puede establecerse en 6.
- Establecimiento de algún tipo de mayoría absoluta para la clase a asignar. Así, si suponemos que  $k = 20$  y  $m = 4$ , podemos convenir en que la asignación del nuevo caso a una clase sólo se llevaría a cabo en el caso de que la diferencia entre las frecuencias mayor y segunda mayor supere 3

### 2.2.4.1.1.3.2 K-NN con distancia media

En el K-NN con distancia media la idea es asignar un nuevo caso a la clase cuya distancia media sea menor. Así que en el ejemplo de la Figura 2.7 con  $k=7$  y  $m=2$ , a pesar de que 5 de los 7 casos más cercanos al mismo pertenecen a la clase  $\circ$ , el nuevo caso se clasifica como  $+$ , ya que la distancia media a los dos casos  $+$  es menor que la distancia media a los cinco casos  $\circ$ .

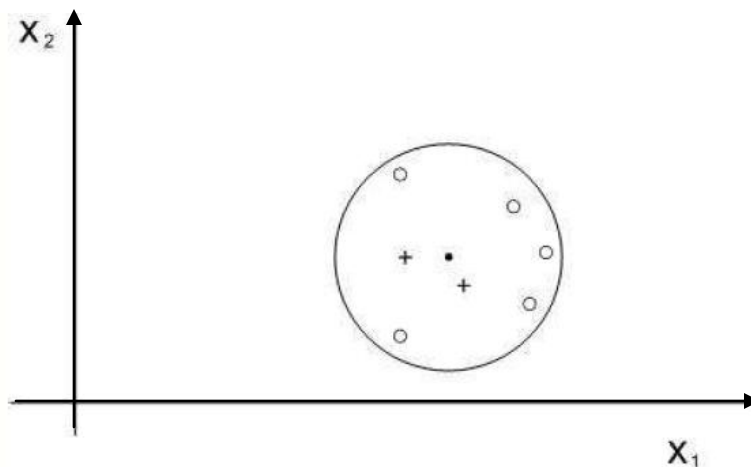


Figura 2.7 K-NN con distancia media

### 2.2.4.1.1.3.3 K-NN con distancia mínima

En el K-NN con distancia mínima se comienza seleccionando un caso por clase, normalmente el caso más cercano al baricentro de todos los elementos de dicha clase. En este paso se reduce la dimensión del archivo de casos a almacenar de  $N$  a  $m$ . A continuación se asigna el nuevo caso a la clase cuyo representante esté más cercano.

El procedimiento anterior puede verse como un 1-NN aplicado a un conjunto de  $m$  casos (uno por cada clase). El coste computacional de este procedimiento es inferior al K-NN genérico, si bien su efectividad está condicionada a la homogeneidad dentro de las clases (cuanto mayor sea dicha homogeneidad, se espera que el procedimiento sea más efectivo).

### 2.2.4.1.1.3.4 K-NN con ponderación de casos seleccionados

La idea en el K-NN con el que se efectúa una ponderación de los casos seleccionados es que los  $K$  casos seleccionados no se contabilicen de igual forma, sino que se tenga en cuenta la distancia de cada caso seleccionado al nuevo caso que pretendemos seleccionar. Como ejemplo podemos convenir en ponderar cada caso seleccionado de

manera inversamente proporcional a la distancia del mismo al nuevo caso. Como resulta que de los 6 casos seleccionados, los pesos relativos a los casos tipo  $\circ$ , suman 2, cantidad inferior al peso de los dos casos tipo  $+$  por lo que el K-NN con ponderación de casos seleccionados clasificará el nuevo caso como perteneciente a la clase  $+$ .

Los resultados de esta variante los podemos ver en la figura 2.8

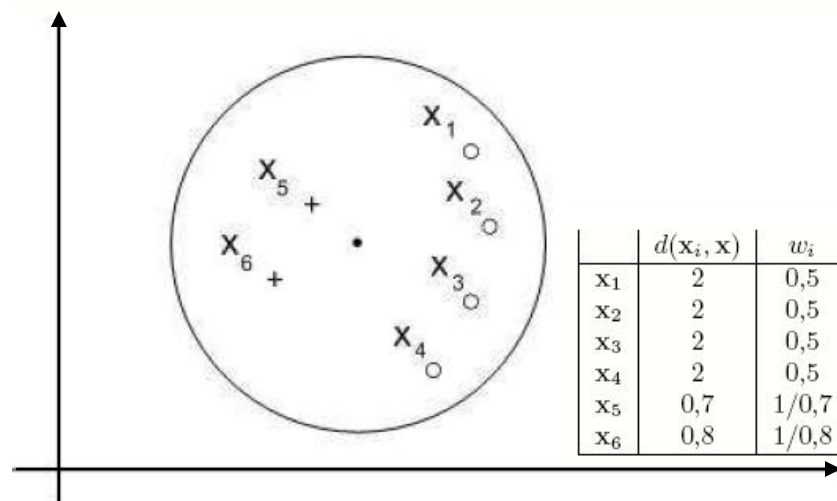


Figura 2.8 K-NN con ponderación de casos seleccionados

### 2.2.4.1.2 Aprendizaje no supervisado

Consiste en aprender a partir de patrones de entradas para los que no se especifican sus valores de salida, por lo que trata de descubrir estructuras en instancias no etiquetadas [24].

#### 2.2.4.1.2.1 Clasificador K-mean

K-mean es un método iterativo para la clasificación de grupos, el cual particiona los datos de entrada en K conjuntos disjuntos. La solución comienza con definir K centroides de manera aleatoria, para determinar el grupo de un valor de entrada consiste en calcular su distancia respecto a cada centroide para elegir aquel que tenga una distancia mínima; para las siguientes iteraciones se deben actualizar los centroides conforme al nuevo grupo de datos y asignar el nuevo punto al grupo con menor distancia, esto se realiza hasta que no haya más entradas que asignar [28].

El método de agrupamiento de K-mean básico es simple. Se comienza por determinar el número de k grupos que se desean generar y se asumen los centroides para estos

grupos. Se pueden tomar puntos aleatorios como los centroides iniciales de los primeros K grupos.

Iterar hasta que no haya cambios en la asignación de los datos de entrada

1. Asignar los puntos al centroide más cercano.
2. Mover la media (calcular el centroide) de los nuevos k grupos.

El proceso se puede visualizar en la figura 2.9

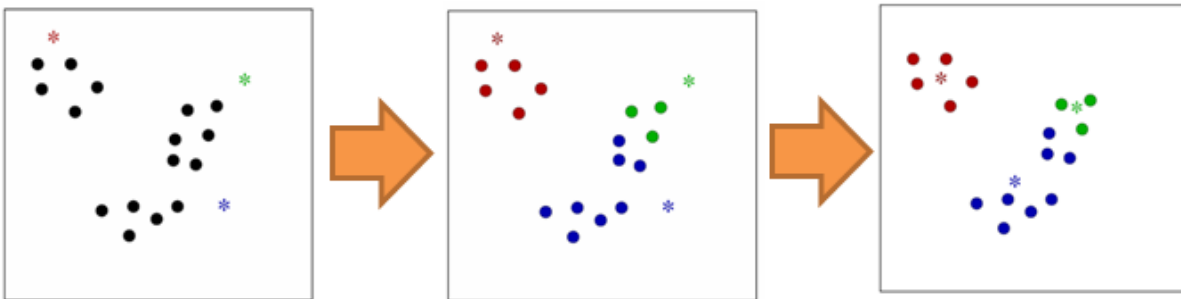


Figura 2.9 Muestra gráfica de etapas K-mean

#### 2.2.4.2 Aprendizaje de mejora de eficiencia

Tiene por objetivo mejorar la eficiencia de procesos existentes, sin embargo esto no es competencia de la presente investigación.

### 2.3 Trabajos relacionados

#### 2.3.1 Group Recommendation System for Facebook

Una de las actividades principales en las redes sociales consiste en crear grupos y formar parte de ellos, sin embargo la gran cantidad de grupos que se crean se deben a muy diversas motivaciones y creencias, por otro lado el abrumador número de grupos existentes en una misma categoría hacen que el seleccionar uno correcto sea complicado para el usuario. Para resolver este problema en la Universidad del Norte de Texas han introducido un sistema de recomendación de grupos combinando técnicas de clasificación jerárquica y árboles de decisión, esta investigación parte de la premisa de que los grupos en una red social pueden ser identificados basándose en el perfil de sus usuarios [9].

Para hacer más precisa la recomendación de grupos se hace uso de la clasificación jerárquica la cual remueve a los miembros del grupo cuyas características no sean relevantes para la mayoría de los miembros del mismo. Una vez que se ha removido el ruido en cada grupo, se construye el motor del sistema de recomendación (árbol de decisión).

En esta investigación se tiene la creencia que mientras más miembros tengan valores en común el grupo crecerá en tamaño y también estará más fuertemente relacionado.

La metodología del trabajo se muestra en la figura 2.9 y se base en una arquitectura de tres componentes: i) extracción de características, ii) sistema de clasificación, iii) recomendación final.

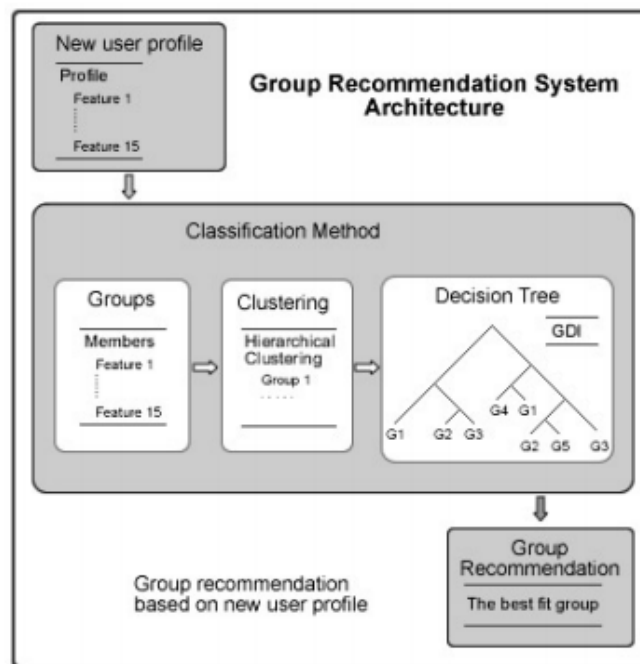


Figura 2.10 Group Recommendation System for Facebook Architecture

### 2.3.2 Using Social Networks to Organize Researcher Community

En este trabajo se busca identificar personas con intereses similares en investigación y con esto ayudar a construir alguna especie de red social. Las fuentes de información provienen principalmente de artículos de investigación que pueden ser buscados libremente en algunos portales de organizaciones tales como ACM y la IEEE. Adicionalmente tiene como motivación poder encontrar más información personal sobre

expertos en algún campo en particular, utilizando herramientas de búsqueda comunes como Google [29].

Para poder construir una red social a partir de la información con la que se cuenta actualmente se plantea seguir los siguientes tres pasos:

1. Construir relaciones entre los autores de artículos extraídos de algún portal de investigación.
2. Extraer la información de cada autor utilizando motores de búsqueda generales como Google.
3. Reorganizar la información a modo de conexiones agregando nueva relaciones a partir de la información de cada autor.

Para realizar la búsqueda y extraer la información de los artículos se plantea el uso de un principio similar a los robots o arañas a través del cual se va extendiendo el rango de artículos, los pasos se listan a continuación:

1. Enviar palabras clave a los portales de investigación.
2. Analizar la lista de artículos retornada por los portales y extraer las referencias.
3. Seguir las referencias para encontrar más artículos.
4. Repetir pasos 2 y 3 hasta que se alcance un extenso número de artículos.
5. Encontrar a los autores de estos artículos
6. Utilizar los portales de investigación para saber cuántas veces ha sido citado cada artículo.
7. Ponderar la importancia de cada autor según el número de veces que han sido citados sus artículos en conjunto.

Referente a la información personal del autor que se plantea recapitular definen la estructura mostrada en la tabla 2.2

**Tabla 2.2 Información personal a recopilar**

| Información seleccionada |                       |                                               |
|--------------------------|-----------------------|-----------------------------------------------|
| Profile information      | Identity information  | Name                                          |
|                          |                       | Age                                           |
|                          |                       | Gender                                        |
|                          |                       | Phone                                         |
|                          |                       | Email                                         |
|                          |                       | Organization / company / institute            |
|                          |                       | Title / Position                              |
|                          |                       | Address                                       |
|                          | Accessory information | Biography                                     |
|                          |                       | Personal homepage                             |
|                          |                       | Instant messaging                             |
|                          |                       | Blog                                          |
|                          |                       | Social network websites, Facebook, Hi, Orkut. |
|                          |                       | Interest                                      |
|                          |                       | Blog article                                  |
| Professional information |                       | Newsgroup, BBS, Google group                  |
|                          |                       | Papers                                        |
|                          |                       | Project                                       |

Finalmente para reorganizar la información se colocan los autores en una gráfica, la cual contendrá líneas que representen algún tipo de relación con otro ya sean del tipo coautor, graduados de la misma universidad, etcétera.

## 2.4 Conclusiones

Se han analizado algoritmos de aprendizaje supervisado y no supervisado, particularmente K-NN y K-mean respectivamente y debido a la naturaleza del trabajo se requiere utilizar un método que nos construya grupos de manera automatizada y flexible por lo cual el algoritmo de K-mean resulta ideal para implementar en la presente investigación ya que cumple con dichas características.

En la actualidad los sistemas de recomendación para Facebook hacen uso exclusivo de las conexiones existentes en la red social, sin embargo pocos son los esfuerzos en analizar la similitud entre usuarios a partir de los intereses, el mayor acercamiento consiste en contabilizar a cuántos de tus contactos en la red les interesan las mismas páginas que a ti (objetos en facebook), sin embargo a pesar de este esfuerzo aparece otra limitante, la cual consiste en que si existen múltiples objetos que quieran hacer referencia al mismo concepto pero con variantes en su nombre, entonces estamos hablando de objetos totalmente diferentes, lo que hace limitado el enfoque anterior.

## **CAPÍTULO III**

### **METODOLOGÍA PROPUESTA**

***"Ahora puedo ver ante mis ojos  
como caen todas las piezas"***  
Ian Curtis

## Capítulo 3 Metodología Propuesta

### 3.1 Introducción

El presente capítulo describe la metodología propuesta para la construcción de grupos de usuarios mediante los perfiles que exponen en una red social. El caso de estudio es mediante dos redes sociales: Facebook y LastFM, la estrategia a seguir es tomar información de la primera y auxiliándonos de la segunda para mapearla a una estructura ponderada (vector) que servirá como entrada de nuestro clasificador.

### 3.2 Arquitectura del sistema

La arquitectura del sistema se muestra en la figura 3.1 la cual se encuentra dividida en cuatro etapas secuenciales: recolección de información, extracción de características, clasificación y visualización; la primera etapa se realiza mediante el uso de Facebook, la segunda con apoyo de LastFM y las últimas dos son internas del sistema.

Las etapas se encuentran indicadas en rectángulos azules mientras que las entradas y salidas se encuentran en rectángulos con borde punteado

En la etapa de recolección de información se tiene como entrada la información del perfil de usuario de Facebook y tiene como salida el conjunto de instancias musicales de su interés (información explícita) y su última ubicación geográfica (información implícita).

En la etapa de extracción de características se toma como entrada el conjunto de instancias musicales del usuario y tiene como salida el vector característico que representa los géneros musicales ponderados.

En la etapa de clasificación se toma como entrada los vectores característicos de los usuarios los cuales son normalizados al iniciar esta etapa y se tiene como salida una lista ordenada con la distancia que existe entre el usuario de interés y el resto que han sido indexados, así como el grupo al que pertenece cada uno de acuerdo al algoritmo k-mean.

En la etapa de visualización se despliegan los resultados sobre un mapa con la ubicación geográfica obtenida en la primera etapa, también se cuenta con un mapa que permite observar cómo se distribuyen cada uno de los grupos encontrados sobre el mapa.

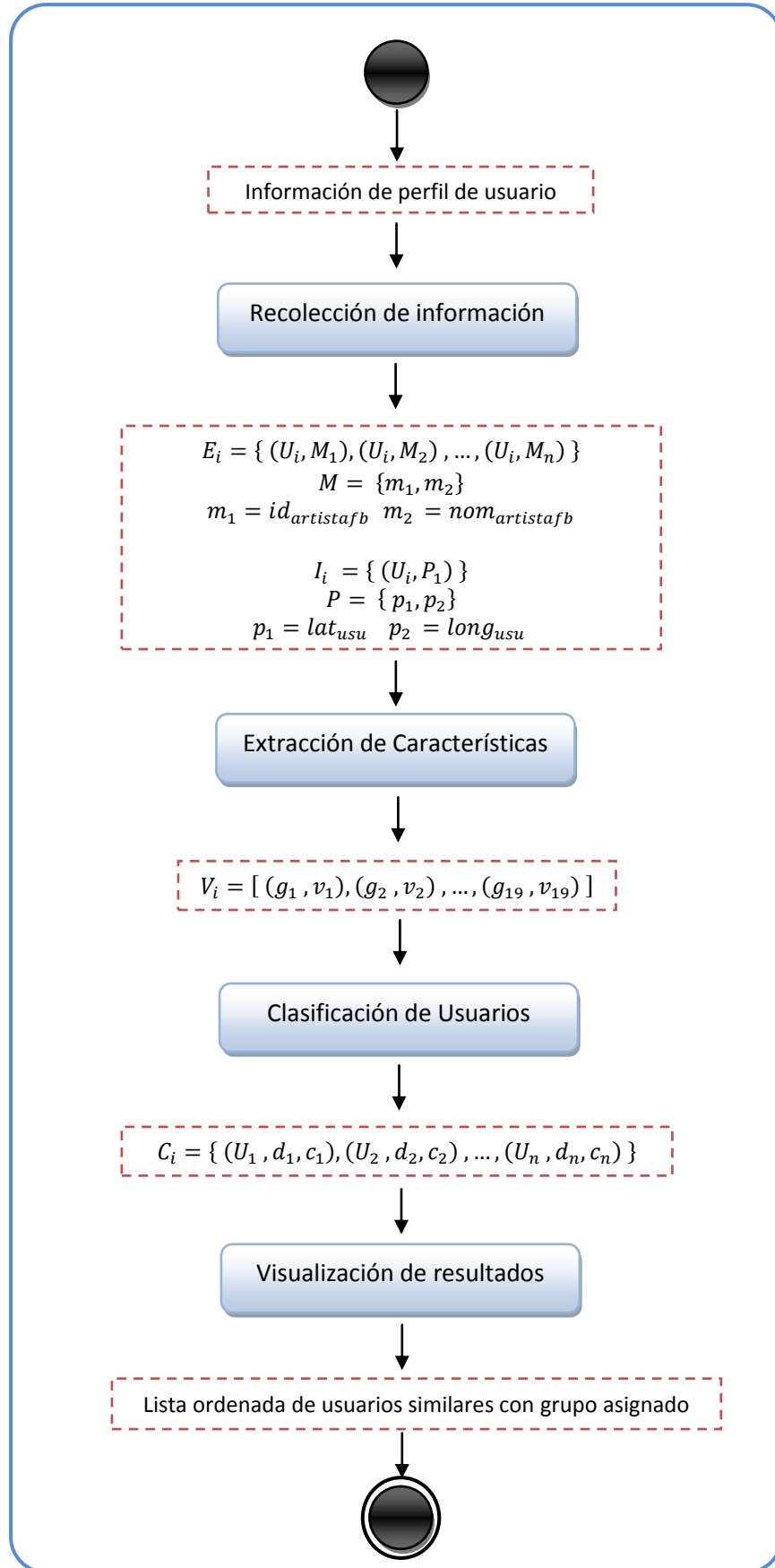


Figura 3.1 Arquitectura del sistema

### 3.2.1 Recolección de información

En la etapa de recolección de información se obtienen dos tipos de información del perfil de usuario de Facebook; en primera instancia obtenemos datos que denominamos explícitos, esto se debe a que es información que la persona ha generado con la intención de que los demás la conozcan, en particular obtenemos su nombre de usuario, su identificador y la lista de instancias musicales de su interés (artistas y grupos) la cual utilizaremos en la extracción de características, la segunda clase de información que obtendremos la llamamos implícita esto se debe a su carácter ya que a diferencia del caso anterior el usuario no la proporciona como parte de su información personal sino que se va produciendo al georeferenciar su actividad diaria en la red social ya sean objetos tales como fotos, video, publicaciones, entre otros; estos objetos tienen asociados una latitud y longitud que será utilizada como ubicación geográfica del usuario, se hará uso exclusivamente del último objeto que haya sido creado con una referencia geográfica por parte la persona.

La información implícita se encuentra definida en la ecuación 3.1

$$\begin{aligned}
 E_i &= \{ (U_i, M_1), (U_i, M_2), \dots, (U_i, M_n) \} \\
 \text{donde } M_k &= \{m_1, m_2\} \quad U_i = \{c_1, c_2\} \\
 m_1 &= id_{artista fb} \quad m_2 = nom_{artista fb} \\
 c_1 &= id_{usuario fb} \quad c_2 = nom_{usuario fb} \\
 &\text{para } 1 \leq k \leq n
 \end{aligned}$$

Ecuación 3.1 Información implícita

La información explícita se encuentra definida en la ecuación 3.2

$$\begin{aligned}
 I_i &= \{ (U_i, P_u) \} \\
 \text{donde } P_u &= \{p_1, p_2\} \quad U_i = \{c_1, c_2\} \\
 p_1 &= lat_{usu} \quad p_2 = long_{usu} \\
 c_1 &= id_{usuario fb} \quad c_2 = nom_{usuario fb}
 \end{aligned}$$

Ecuación 3.2 Información explícita

Cabe mencionar que en esta etapa la información que se almacenará en una base de datos será la lista de instancias musicales del usuario y la ubicación geográfica mediante un valor para la latitud y otro para la longitud, lo anterior se realizará en dos tablas la primera con una cardinalidad  $M \times N$  y la segunda  $1 \times 1$ .

### 3.2.2 Extracción de características

La extracción de características es realizada sobre la información implícita por lo tanto es necesario definir una estructura formal que nos ayude a representar los intereses musicales de una persona.

La idea consiste en construir una lista de géneros musicales que resulte representativa para la mayoría de personas que habitan en una red social, para este punto se debe conocer alguna clasificación de géneros musicales para poder seleccionar los que consideremos más representativos.

En primera instancia partiremos de la taxonomía de la música mostrada en la figura 3.2 [1]

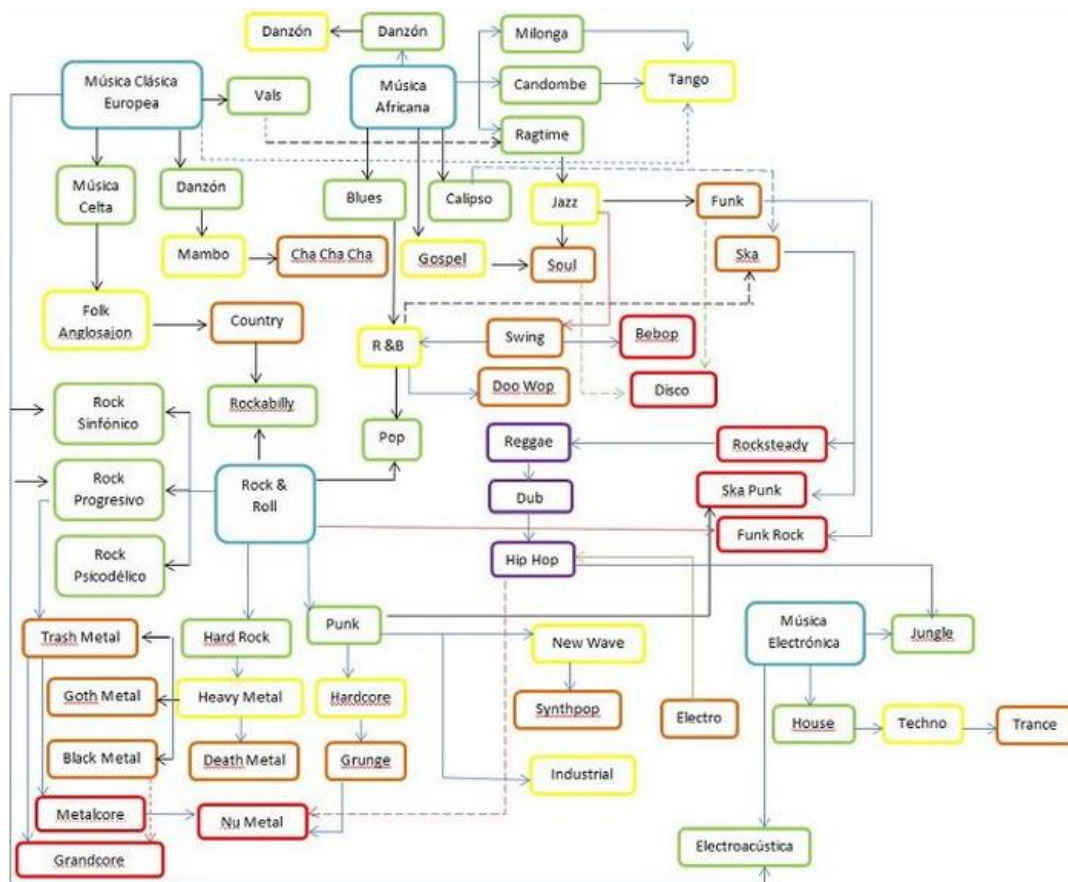


Figura 3.2 Taxonomía de la música por Sorella

En segunda instancia analizamos la folcsonomía construida por las etiquetas más populares que han sido asignadas por los usuarios de LastFM como se comentó en el capítulo 2, en la figura 3.3 se muestra un ejemplo donde el tamaño indica la popularidad de las etiquetas.

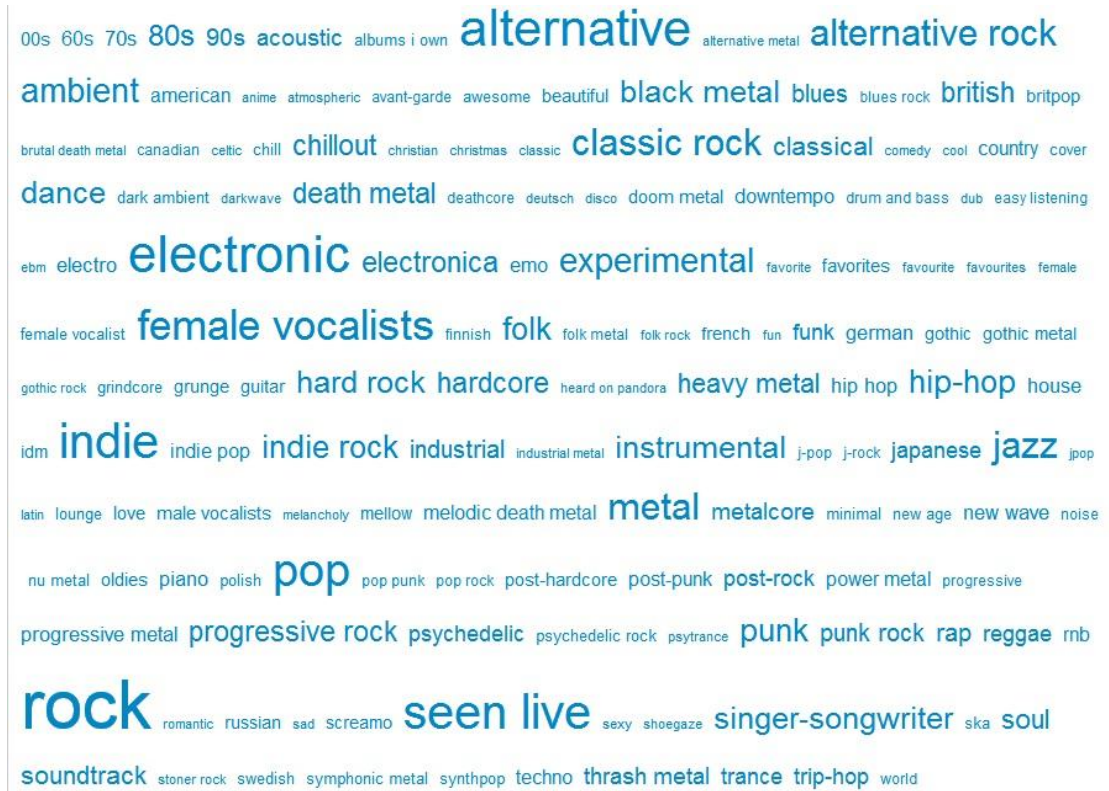


Figura 3.3 Principales etiquetas utilizadas en LastFM (folcsonomía)

La información anterior puede ser consultada realizando una solicitud mediante su API, particularmente del método *tag.gettoptags*. En la ecuación 3.3 se define el conjunto de etiquetas que se obtienen en la petición anterior.

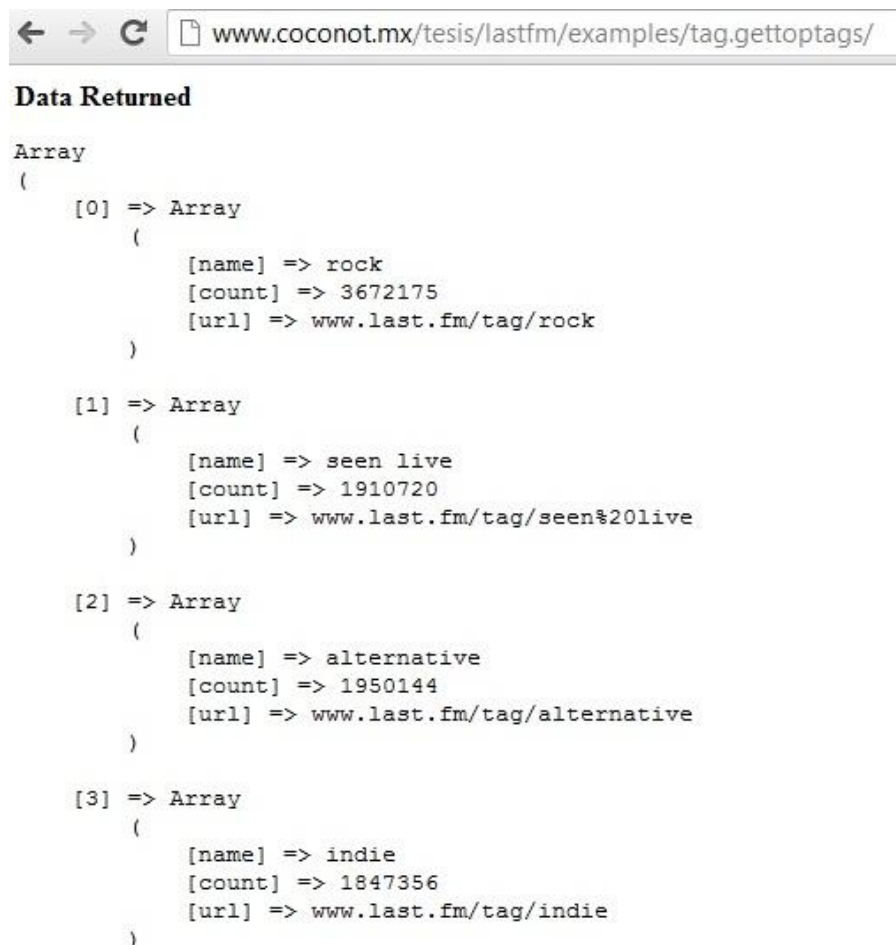
$$A = \{ E_1, E_2, E_3, \dots, E_m \}$$

$$\text{donde } E = \{ t_1, t_2, t_3 \}$$

$$\text{para } t_1 = \text{name}_{eti q} \quad t_2 = \text{count}_{eti q} \quad t_3 = \text{url}_{eti q}$$

Ecuación 3.3 Respuesta de conjunto de etiquetas

En la Figura 3.4 se muestra el resultado de dicha petición.



```

Data Returned

Array
(
    [0] => Array
        (
            [name] => rock
            [count] => 3672175
            [url] => www.last.fm/tag/rock
        )

    [1] => Array
        (
            [name] => seen live
            [count] => 1910720
            [url] => www.last.fm/tag/seen%20live
        )

    [2] => Array
        (
            [name] => alternative
            [count] => 1950144
            [url] => www.last.fm/tag/alternative
        )

    [3] => Array
        (
            [name] => indie
            [count] => 1847356
            [url] => www.last.fm/tag/indie
        )

```

**Figura 3.4 Salida en formato JSON etiquetas de LastFM**

Para seleccionar los géneros que representen el perfil musical del usuario es necesario eliminar las etiquetas que no tengan que ver con géneros musicales las cuales han sido utilizadas por la comunidad de LastFM debido a la libertad que existe para definir los propios métodos de organización de las bibliotecas musicales, un ejemplo de etiqueta que no es de interés en este trabajo es *seen live*, la cual es utilizada únicamente para generar listas personalizadas de los grupos que los usuarios de la comunidad de LastFM han visto en vivo (concierto).

Finalmente al realizar un balance entre la taxonomía que proporciona Sorella y la folcsonomía de LastFM se concluye que los géneros seleccionados para representar el perfil musical del usuario se muestran en la Tabla 3.1

**Tabla 3.1 Ejemplos de representantes de género**

| ID | Género (etiqueta) | Ejemplos de representante (LastFM)                                                    |
|----|-------------------|---------------------------------------------------------------------------------------|
| 1  | Rock clásico      | The Beatles, The Rolling Stones, Led Zepellin, Queen, Pink Floyd, The Doors, The Who  |
| 2  | Rock              | Red Hot Chili Peppers, U2, Foo Fighters, Muse, White Stripes, Smashing Pumkins        |
| 3  | Punk              | Ramones, The Clash, Green Day, Bad Religion, The Offspring, Misfits, Sex Pistols      |
| 4  | Reggae            | Bob Mayley, UB40, Manu Chao, Peter Tosh, Jimmi Cliff, Seeed                           |
| 5  | Alternativa       | Radiohead, Beck, The White Stripes, Bjork, Tori Amos, The Cure Sonic Youth            |
| 6  | BritPop           | Blur, Oasis, Coldplay, Blur, Keane, Travis, The Stone Roses, The Verve, Kalser Chiefs |
| 7  | Rap               | Beastie Boys, Jay-Z, The Game, Kanye West, DMX, Ludacris                              |
| 8  | Pop               | Madonna, Britney Spears, Michael Jackson, Lady Gaga, Katy Perry, ABBA                 |
| 9  | Grunge            | Nirvana, Pearl Jam, Alice in Chains, Stone Temple Pilots, Bush                        |
| 10 | Indie             | Belle and Sebastian, Florence + The Machine, Arctic Monkeys, Bloc Party               |
| 11 | Hip-Hop           | Cypress Hill, Emine, Beastie Boys, Kanye West, The Notorious B.I.G                    |
| 12 | Blues             | Tom Waits, Eric Clapton, B.B. King, Muddy Waters, Robert Johnson                      |
| 13 | Clásica           | Wolfgang Amadeus Mozart, Ludwig van Beethoven, Johann Sebastian Bach                  |
| 14 | Country           | Johnny Cash, Brad Praisley, Alan, Jackson, Tim McGraw                                 |
| 15 | Ska               | Reel Big Fish, The Specials, Sublime, Madness, Ska-P, The Aquabats                    |
| 16 | Jazz              | Miles Davis, Billie Holliday, Norah Jones, John Cotrane, Louis Armstrong              |
| 17 | Metal             | Metallica, Slipknot, Iron Maiden, System of a Down, Deftones, Megadeth                |
| 18 | Banda             | La Arrolladora Banda Limon, Banda El Recodo, Banda Machos, Valentin Elizalde          |
| 19 | Electrónica       | Daft Punk, The Chemical Brothers, Board of Canada, The Prodigy, The Knife             |

De la tabla anterior se puede observar que los representantes de cada género según las etiquetas asignadas por la comunidad de LastFm son muy precisas lo cual se debe a dos hechos principales, el primero consiste en que un artista puede tener varias etiquetas asignadas al mismo tiempo lo que permite que la clasificación sea más flexible, por otro lado el resultado es la suma de todas las etiquetas proporcionadas por cada miembro de la comunidad.

En la ecuación 3.4 se define el vector característico que será utilizado como salida de esta etapa y entrada del clasificador.

$$G = [g_1, g_2, \dots, g_{19}]$$

$$\text{donde } \forall g \in G \quad g \geq 0$$

**Ecuación 3.4 Vector característico de la extracción**

El vector representando en la ecuación anterior, contendrá los valores ponderados de cada uno de los géneros seleccionados para representar las preferencias musicales del usuario.

Los 19 géneros seleccionados funcionarán como variables predictorias para nuestro algoritmo de clasificación.

En la Tabla 3.2 se muestra la estadística para el número de veces que ha sido asignada la etiqueta de cada género seleccionado.

**Tabla 3.2 Estadísticas de géneros en LastFM**

| Género (etiqueta)   | Ocasiones Asignada | URL para detalles                                                                  |
|---------------------|--------------------|------------------------------------------------------------------------------------|
| <b>Rock clásico</b> | 1.038.103          | <a href="http://www.last.fm/tag/classic%20rock">www.last.fm/tag/classic%20rock</a> |
| <b>Rock</b>         | 3.672.175          | <a href="http://www.last.fm/tag/rock">www.last.fm/tag/rock</a>                     |
| <b>Punk</b>         | 792.122            | <a href="http://www.last.fm/tag/punk">www.last.fm/tag/punk</a>                     |
| <b>Reggae</b>       | 364.492            | <a href="http://www.last.fm/tag/reggae">www.last.fm/tag/reggae</a>                 |
| <b>Alternativa</b>  | 1.950.144          | <a href="http://www.last.fm/tag/alternative">www.last.fm/tag/alternative</a>       |
| <b>BritPop</b>      | 236.920            | <a href="http://www.last.fm/tag/britpop">www.last.fm/tag/britpop</a>               |
| <b>Rap</b>          | 462.210            | <a href="http://www.last.fm/tag/rap">www.last.fm/tag/rap</a>                       |
| <b>Pop</b>          | 1.716.067          | <a href="http://www.last.fm/tag/pop">www.last.fm/tag/pop</a>                       |
| <b>Grunge</b>       | 238.255            | <a href="http://www.last.fm/tag/grunge">www.last.fm/tag/grunge</a>                 |
| <b>Indie</b>        | 1.847.356          | <a href="http://www.last.fm/tag/indie">www.last.fm/tag/indie</a>                   |
| <b>Hip-Hop</b>      | 762.088            | <a href="http://www.last.fm/tag/hip-hop">www.last.fm/tag/hip-hop</a>               |
| <b>Blues</b>        | 482.480            | <a href="http://www.last.fm/tag/blues">www.last.fm/tag/blues</a>                   |
| <b>Clásica</b>      | 503.892            | <a href="http://www.last.fm/tag/classical">www.last.fm/tag/classical</a>           |
| <b>Country</b>      | 300.231            | <a href="http://www.last.fm/tag/country">www.last.fm/tag/country</a>               |
| <b>Ska</b>          | 201.219            | <a href="http://www.last.fm/tag/ska">www.last.fm/tag/ska</a>                       |
| <b>Jazz</b>         | 1.026.802          | <a href="http://www.last.fm/tag/jazz">www.last.fm/tag/jazz</a>                     |
| <b>Metal</b>        | 1.134.238          | <a href="http://www.last.fm/tag/metal">www.last.fm/tag/metal</a>                   |
| <b>Banda</b>        | 2.080              | <a href="http://www.last.fm/tag/banda">www.last.fm/tag/banda</a>                   |
| <b>Electrónica</b>  | 2.053.436          | <a href="http://www.last.fm/tag/electronic">www.last.fm/tag/electronic</a>         |

Para obtener esta información se puede realizar consultas mediante la API de LastFM sin embargo se realiza mediante la interfaz web de la aplicación para poder desplegar resultados de manera más didáctica como se muestra en las figuras 3.5 a 3.21

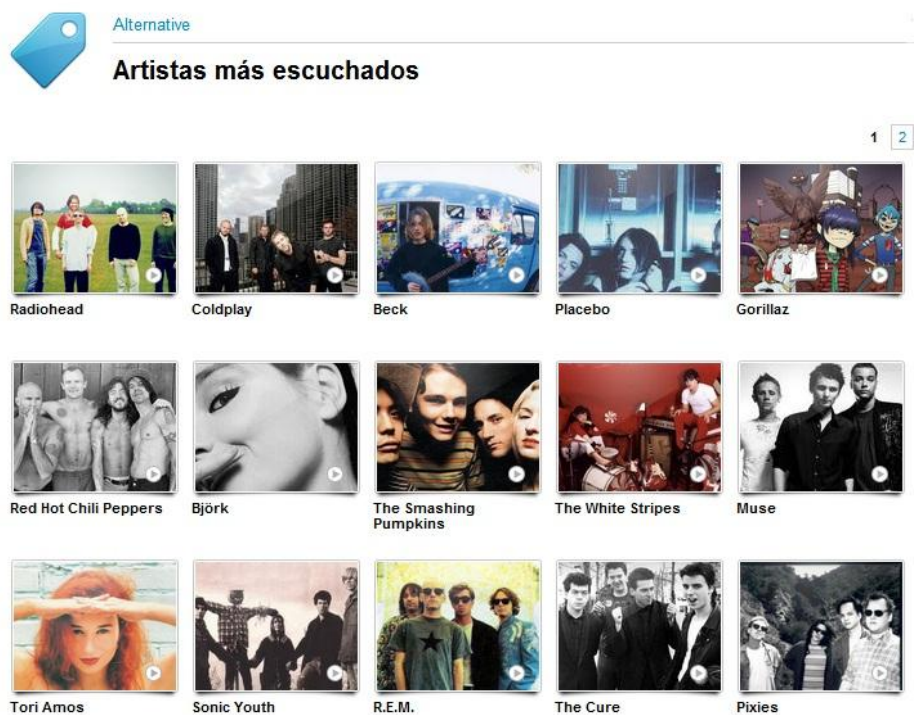


Figura 3.5 Salida género alternativo

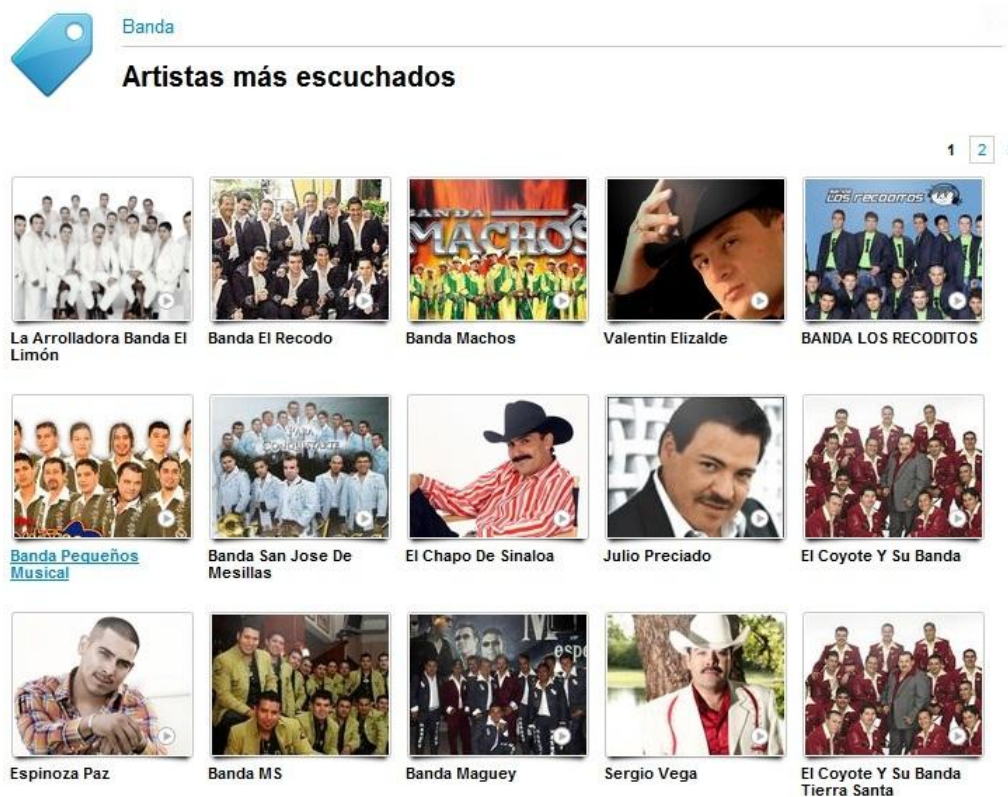


Figura 3.6 Salida género banda

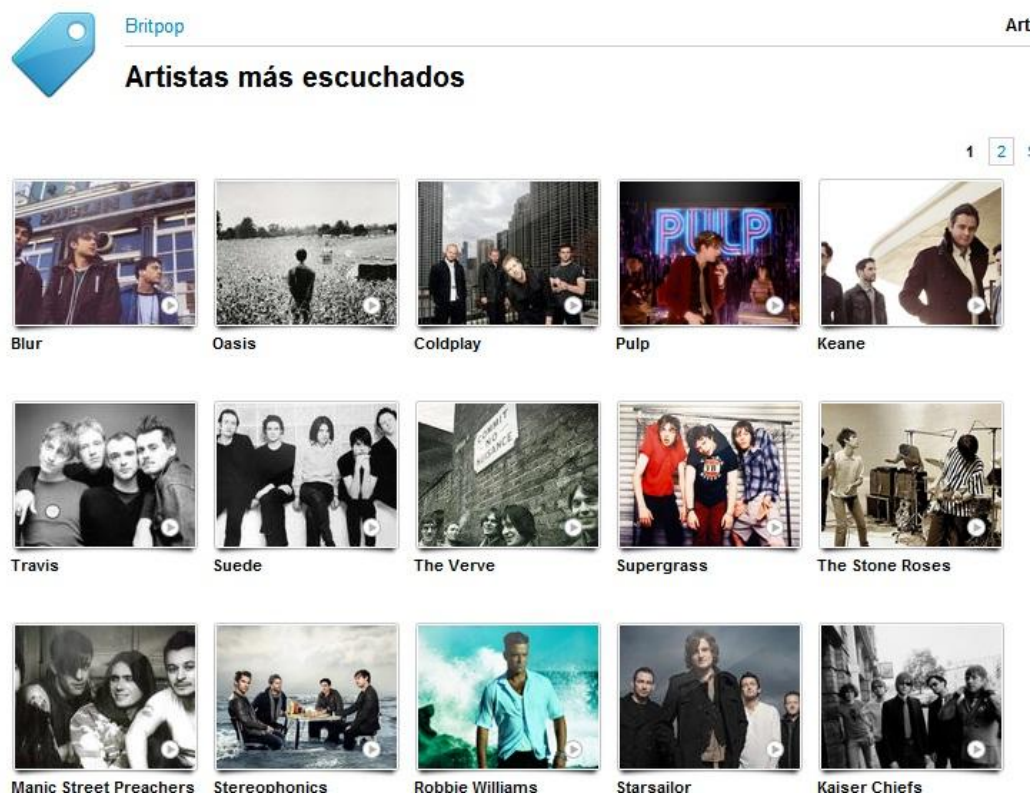


Figura 3.7 Salida género britpop



Figura 3.8 Salida género clásico

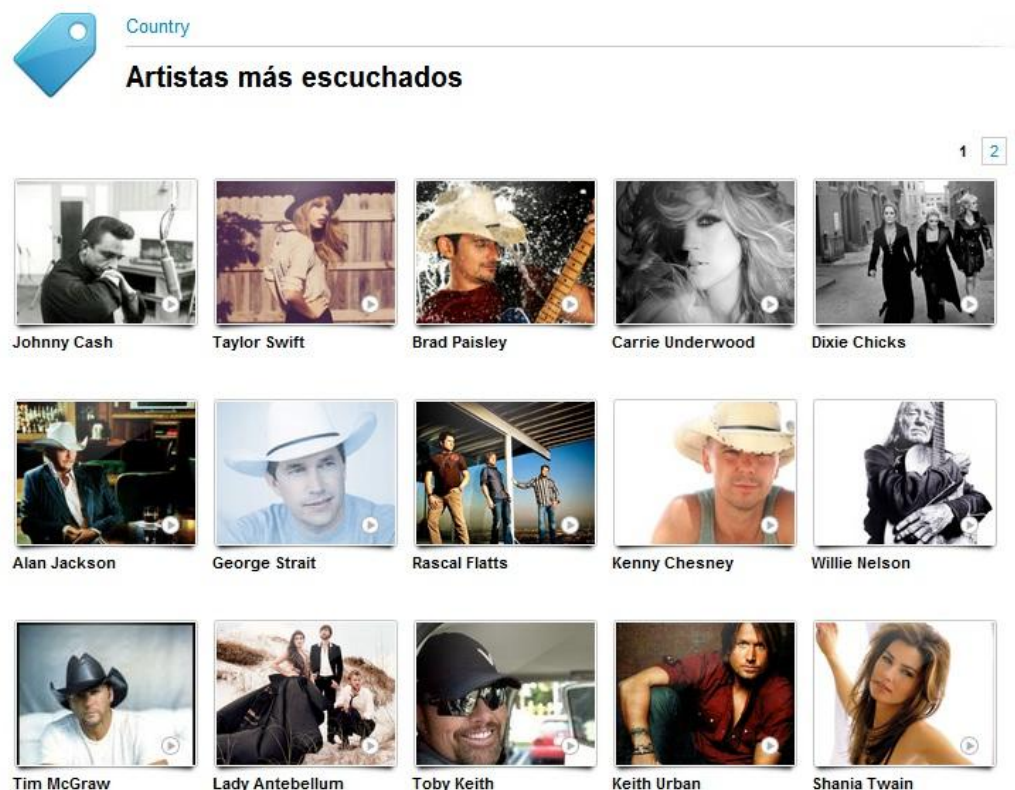


Figura 3.9 Salida género country



Figura 3.10 Salida género grunge



Figura 3.11 Salida género Hip-Hop



Figura 3.12 Salida género indie

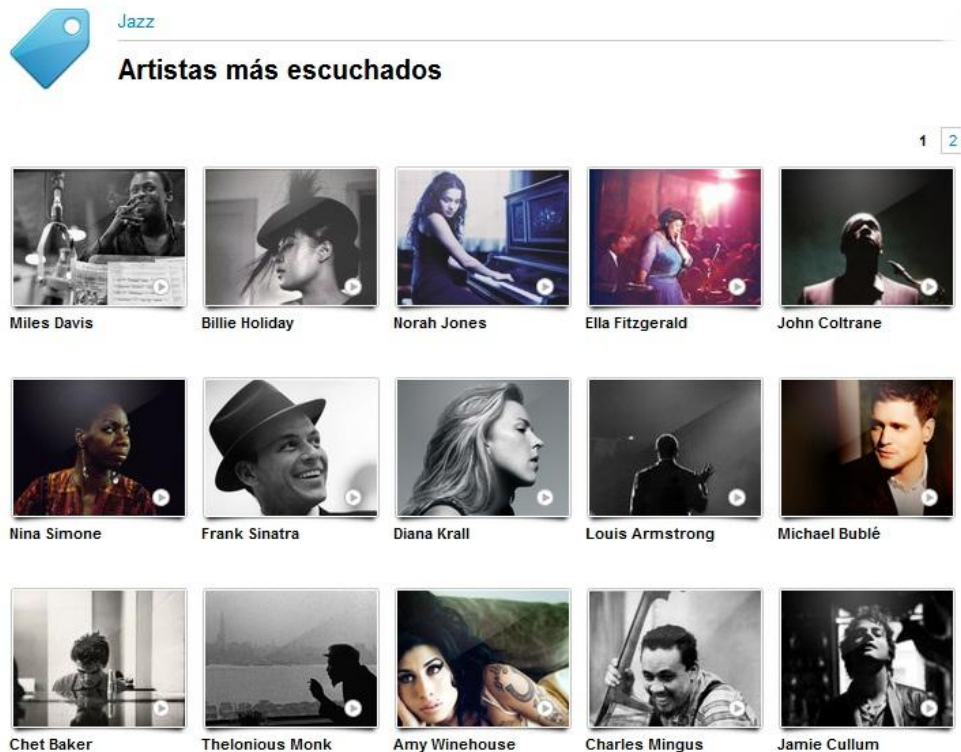


Figura 3.13 Salida género jazz



Figura 3.14 Salida género metal

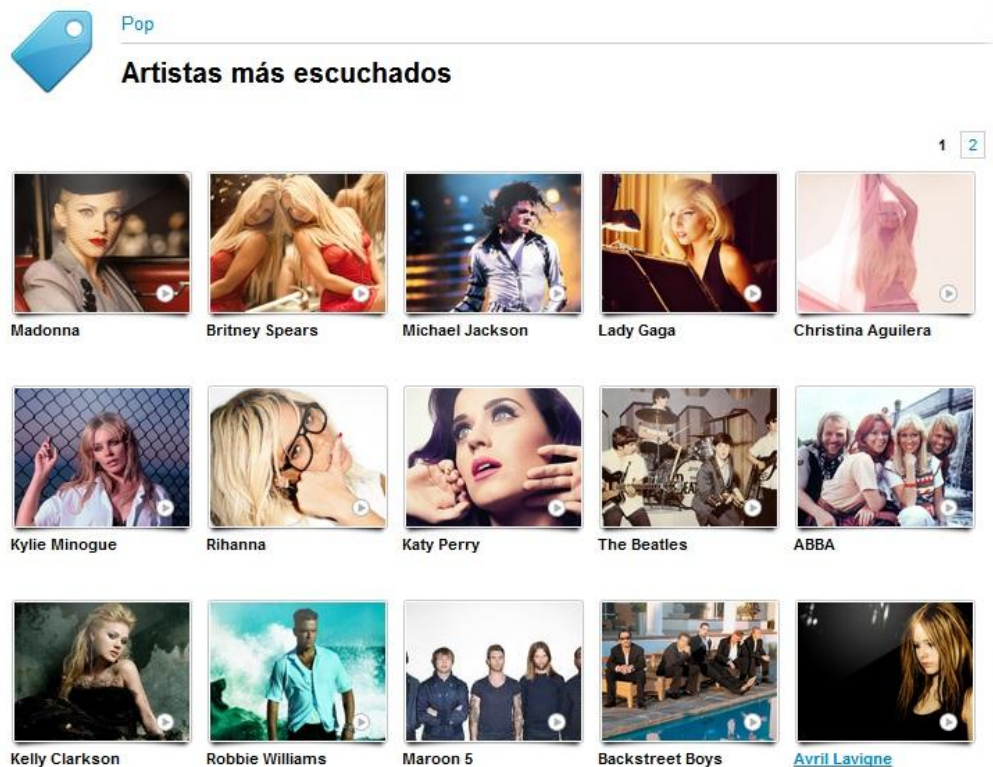


Figura 3.15 Salida género pop



Figura 3.16 Salida género punk

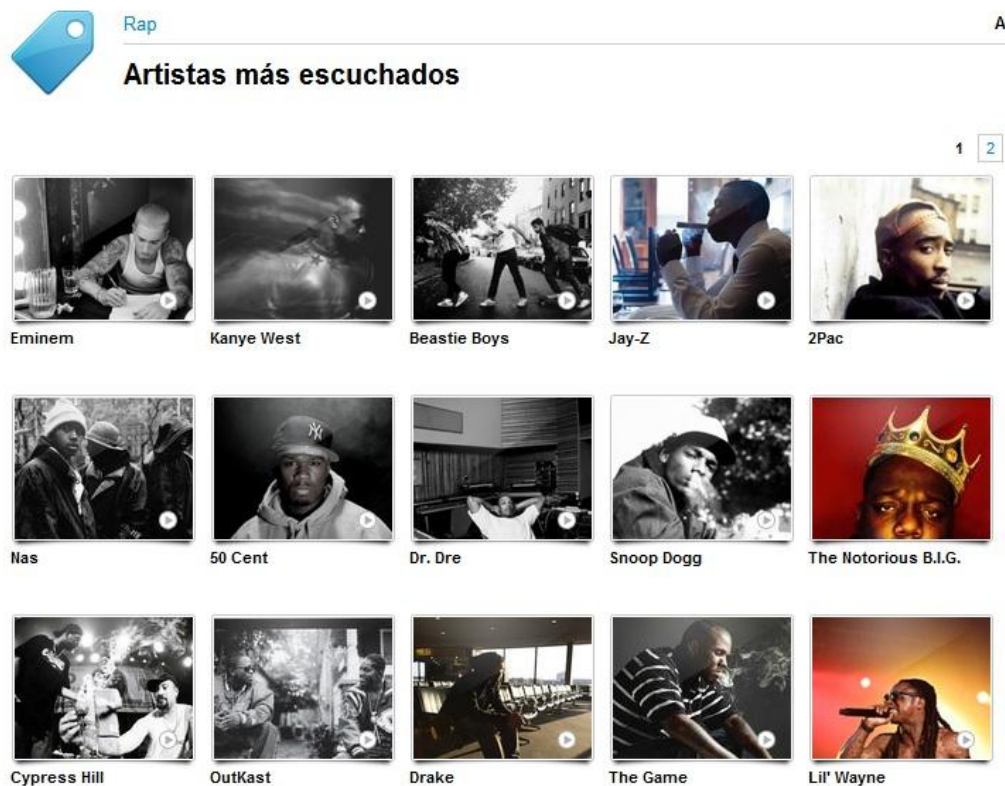


Figura 3.17 Salida género rap

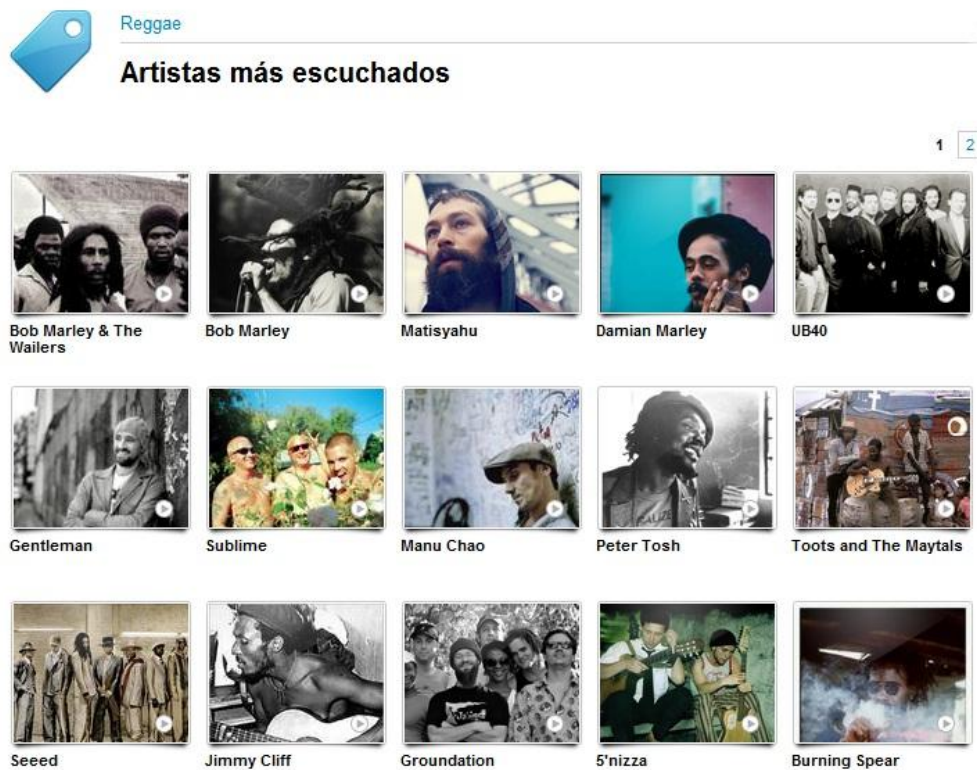


Figura 3.18 Salida género reggae

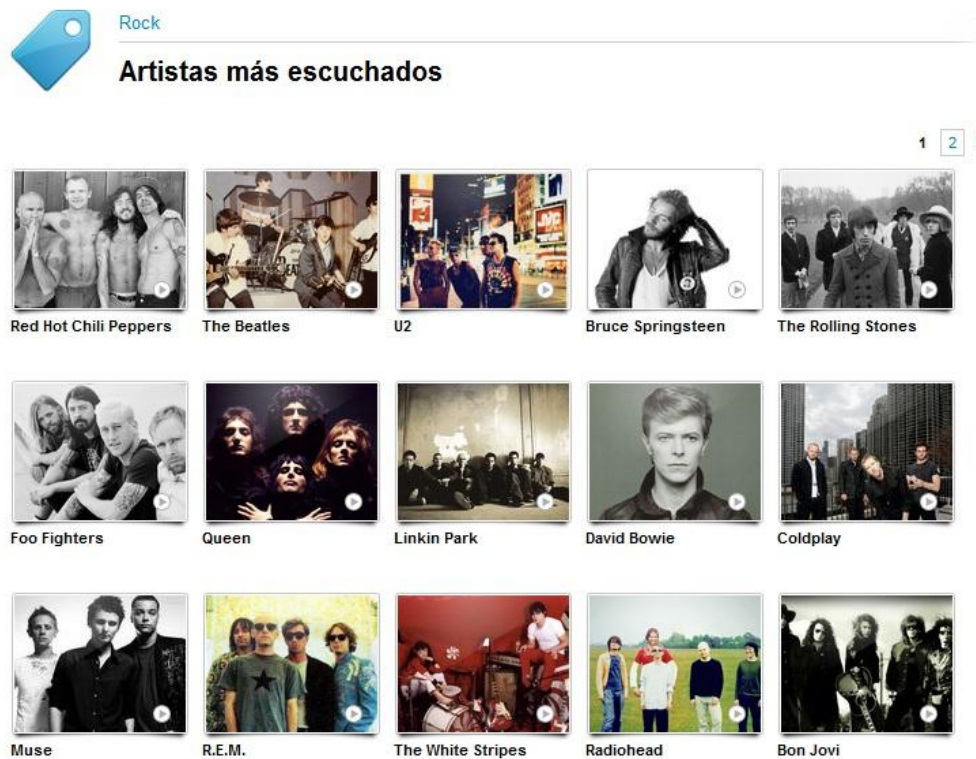


Figura 3.19 Salida género rock

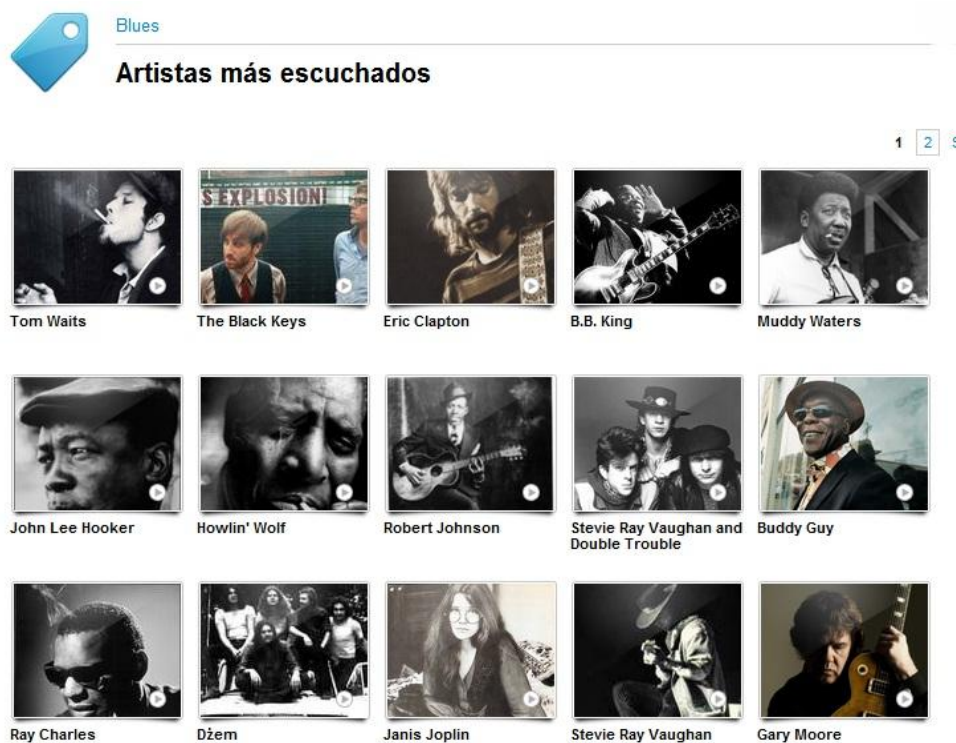


Figura 3.20 Salida género blues

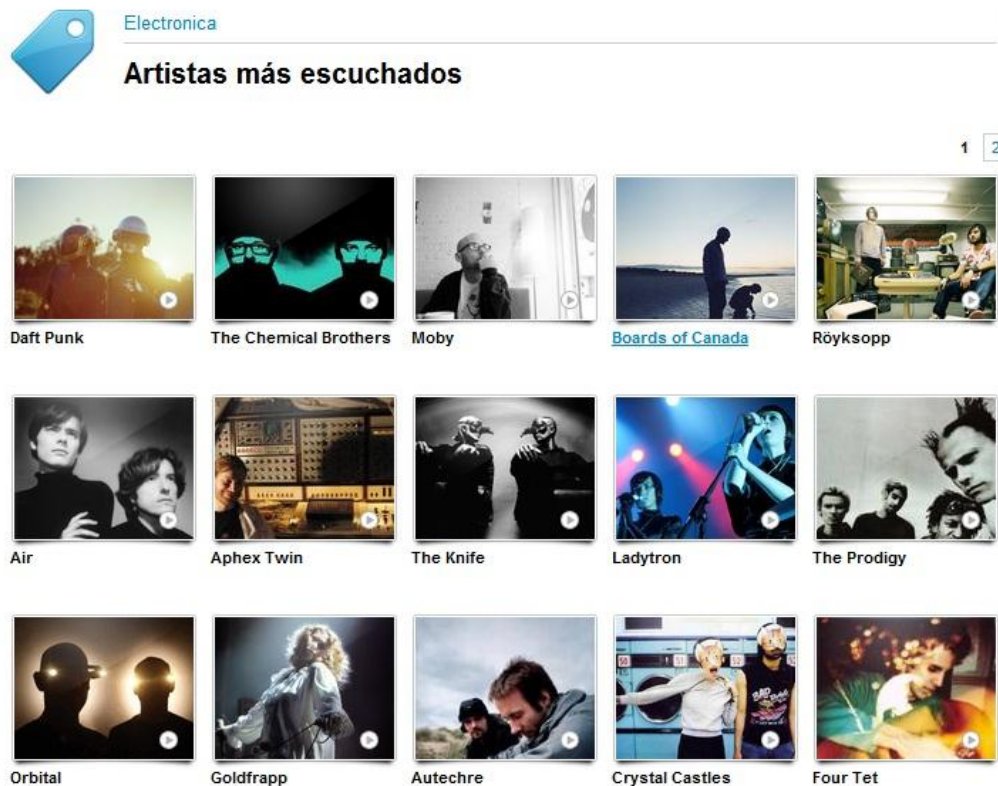


Figura 3.21 Salida género electrónica

### 3.2.2.1 Algoritmo de extracción

Con lo anterior en mente se debe proceder a ponderar el vector característico que represente el perfil musical de un usuario y lo cual se describe en el siguiente algoritmo:

1. Recorrer cada instancia musical del perfil de usuario de Facebook.
2. Para cada instancia obtener las 10 etiquetas con mayor ponderación otorgada por la comunidad de LatFM.
3. Por cada etiqueta comparar si ésta se encuentra dentro de la lista definida para representar el perfil musical del usuario, en caso positivo aumentar en uno la posición en el vector que represente dicha clase-categoría (género) y romper el ciclo para esta instancia, en caso contrario continuar con la siguiente etiqueta.
4. Integrar información implícita y explícita

A continuación se presenta un ejemplo de resultado mediante la implementación del algoritmo anterior.

Listado de instancias musicales de Usuario X:

1. Led Zepellin.
2. David Guetta.
3. Daf Punk.
4. Pulp.
5. Nirvana.
6. Manos de Topo.
7. Shakira.
8. Snoop Dog.
9. Miranda.
10. The Beatles.
11. Porter.

En la Tabla 3.3 se despliegan las etiquetas asociadas por la comunidad de LastFM a las instancias musicales del usuario.

Tabla 3.3 Etiquetas de prueba

| Instancia (artista)  | Etiquetas asociadas                                              |
|----------------------|------------------------------------------------------------------|
| <b>Led Zepellin</b>  | <b>Rock clásico</b> , Rock, Hard rock, 70's, Rock progresivo     |
| <b>David Guetta</b>  | House, Dance , <b>Electrónica</b> , Club, Techno                 |
| <b>Daf Punk</b>      | <b>Electrónica</b> , Dance, House, Techno, Club                  |
| <b>Pulp</b>          | <b>Britpop</b> , Indie, British, Alternativa, Rock               |
| <b>Nirvana</b>       | <b>Grunge</b> , Rock, Alternativa, Rock Alternativo, 90's        |
| <b>Manos de Topo</b> | Indie Pop, Spanish Indie Pop, <b>Indie</b> , Spanish, Surrealism |
| <b>Shakira</b>       | <b>Pop</b> , Latin, Female Vocalist, Spanish, Rock               |
| <b>Snoop Dog</b>     | Snoop, Snoop Dog, <b>Hip-Hop</b> , Rnb, Dog                      |
| <b>Miranda!</b>      | <b>Pop</b> , Electro Pop, Argentina, Electrónica, Dance          |
| <b>The Beatles</b>   | <b>Rock clásico</b> , Rock, British, 60's, Pop                   |
| <b>Porter</b>        | <b>Indie</b> , Experimental, Alternativa, Rock, Mexico           |

En los resultados anteriores se resalta en negrita la etiqueta seleccionada por cada instancia según nuestra lista definida.

En este punto se ha finalizado la construcción del vector característico del usuario, con la extracción de información implícita y explícita, los resultados se deben mapear a la estructura definida en la Tabla 3.4

**Tabla 3.4 Estructura para almacenar vector característico**

| ID | Género (etiqueta)     |
|----|-----------------------|
| 1  | ID_Usuario (Facebook) |
| 2  | Rock clásico          |
| 3  | Rock                  |
| 4  | Punk                  |
| 5  | Reggae                |
| 6  | Alternativa           |
| 7  | BritPop               |
| 8  | Rap                   |
| 9  | Pop                   |
| 10 | Grunge                |
| 11 | Indie                 |
| 12 | Hip-Hop               |
| 13 | Blues                 |
| 14 | Clásica               |
| 15 | Country               |
| 16 | Ska                   |
| 17 | Jazz                  |
| 18 | Metal                 |
| 19 | Banda                 |
| 20 | Electrónica           |

### 3.2.3 Clasificación de usuarios

Para la clasificación de usuarios se plantea utilizar un método de clasificación no supervisada y se ha seleccionado el algoritmo K-mean debido a que resulta natural utilizar los vectores característicos como variables predictorias en la entrada del algoritmo, por otro lado el coste del aprendizaje es nulo, es tolerante al ruido, podemos fácilmente extender o reducir el mecanismo mediante el número de variables predictorias; entre sus desventajas se menciona el hecho de que puede requerir múltiples iteraciones del algoritmo para llegar a una solución óptima ya que los grupos generados dependen de la selección de los centroides iniciales, sin embargo

consideramos que los grupos que se construyan serán objetivos respecto a las preferencias de los usuarios.

El algoritmo K-mean tendrá por entrada el vector característico normalizado que generamos en el proceso de extracción mediante la información explícita del usuario (estructura ponderada donde cada posición contiene un valor que representa el interés del usuario por dicho género),

Los vectores deben estar normalizados lo cual significa que su norma es igual a uno, lo cual se consigue dividiendo cada componente por la norma del vector de interés, como se muestra en la ecuación 3.5

$$V = \left( \frac{v_1}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}}, \frac{v_2}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}}, \dots, \frac{v_n}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \right)$$

**Ecuación 3.5 Vector normalizado**

La estructura en esta metodología para K-mean consiste en definir 19 variables predictorias que son los géneros ponderados y normalizados que se han seleccionado con N casos que representen a los usuarios que utilizan la aplicación.

Por otro lado en esta etapa se realiza el cálculo de la distancia Euclidiana, donde la distancia entre dos puntos  $P = (p_1, p_2, \dots, p_n)$  y  $Q = (q_1, q_2, \dots, q_n)$  de un espacio euclídeo n-dimensional se define como se muestra en la ecuación 3.6

$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

**Ecuación 3.6 Distancia Euclidiana**

En esta investigación se define un espacio de  $n = 19$  esto resulta de un criterio propio que surge de analizar la taxonomía y la folcsonomía de la música como se explicó previamente.

En la Tabla 3.5 se muestra un ejemplo de entrada para la clasificación según lo expuesto en el punto anterior, mientras que en la Tabla 3.6 se muestra un ejemplo de los vectores normalizados.

**Tabla 3.5 Ejemplo de entrada para clasificación**

| ClassicRock | Rock | Punk | Reggae | Electronic | Alternative | BritPop | Rap | Pop | Grunge | Indie | HipHop | Blues | Clasico | Country | Ska | Jazz | Metal | Banda |
|-------------|------|------|--------|------------|-------------|---------|-----|-----|--------|-------|--------|-------|---------|---------|-----|------|-------|-------|
| 2           | 3    | 0    | 0      | 0          | 0           | 0       | 0   | 0   | 0      | 0     | 0      | 0     | 0       | 0       | 1   | 0    | 0     | 0     |
| 2           | 4    | 1    | 3      | 1          | 0           | 0       | 3   | 4   | 1      | 0     | 2      | 1     | 0       | 0       | 0   | 1    | 1     | 0     |
| 21          | 40   | 2    | 2      | 14         | 14          | 8       | 0   | 2   | 8      | 4     | 1      | 3     | 7       | 0       | 0   | 6    | 2     | 0     |
| 2           | 2    | 0    | 1      | 1          | 0           | 0       | 0   | 0   | 1      | 0     | 0      | 0     | 8       | 0       | 1   | 2    | 0     | 0     |
| 0           | 0    | 0    | 0      | 0          | 0           | 0       | 0   | 1   | 0      | 0     | 0      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 0           | 0    | 0    | 0      | 2          | 0           | 0       | 0   | 0   | 0      | 2     | 0      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 0           | 1    | 0    | 1      | 1          | 0           | 0       | 2   | 3   | 0      | 1     | 3      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 5           | 16   | 1    | 2      | 5          | 3           | 2       | 0   | 1   | 0      | 3     | 0      | 0     | 0       | 0       | 7   | 0    | 0     | 2     |
| 0           | 7    | 0    | 0      | 2          | 3           | 0       | 0   | 0   | 0      | 5     | 1      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 0           | 5    | 0    | 0      | 9          | 5           | 1       | 0   | 0   | 0      | 2     | 0      | 1     | 0       | 0       | 0   | 1    | 0     | 0     |
| 23          | 36   | 10   | 0      | 2          | 6           | 5       | 0   | 2   | 5      | 16    | 0      | 0     | 0       | 0       | 0   | 2    | 8     | 0     |
| 2           | 7    | 0    | 0      | 12         | 3           | 3       | 0   | 9   | 0      | 12    | 1      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 0           | 0    | 0    | 0      | 0          | 0           | 0       | 0   | 1   | 0      | 0     | 0      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 0           | 0    | 0    | 0      | 1          | 0           | 0       | 0   | 0   | 0      | 0     | 0      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 0           | 1    | 0    | 0      | 0          | 0           | 0       | 0   | 2   | 0      | 0     | 0      | 0     | 0       | 0       | 1   | 0    | 0     | 0     |
| 0           | 3    | 0    | 0      | 2          | 2           | 0       | 0   | 7   | 0      | 1     | 0      | 1     | 0       | 0       | 0   | 0    | 0     | 0     |
| 6           | 31   | 4    | 2      | 9          | 8           | 3       | 0   | 13  | 2      | 32    | 2      | 2     | 0       | 0       | 9   | 14   | 0     | 2     |
| 1           | 39   | 8    | 5      | 10         | 5           | 0       | 4   | 25  | 0      | 15    | 3      | 0     | 1       | 0       | 12  | 0    | 1     | 5     |
| 0           | 0    | 0    | 0      | 0          | 0           | 0       | 0   | 0   | 0      | 0     | 0      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| 3           | 0    | 0    | 0      | 2          | 0           | 1       | 0   | 2   | 1      | 2     | 1      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |

**Tabla 3.6 Ejemplo de entrada para clasificación**

| ClassicRock | Rock   | Punk   | Reggae | Electronic | Alternative | BritPop | Rap    | Pop    | Grunge | Indie  | HipHop | Blues  | Clasico | Country | Ska    | Jazz   | Metal  | Banda  |
|-------------|--------|--------|--------|------------|-------------|---------|--------|--------|--------|--------|--------|--------|---------|---------|--------|--------|--------|--------|
| 0.0205      | 0.6971 | 0.164  | 0.1025 | 0.205      | 0.1025      | 0       | 0.082  | 0.5126 | 0      | 0.287  | 0.0615 | 0      | 0.0205  | 0       | 0.246  | 0      | 0.0205 | 0.082  |
| 0.4339      | 0.6508 | 0.1085 | 0      | 0.1085     | 0           | 0       | 0.1085 | 0.3254 | 0.1085 | 0.2169 | 0      | 0      | 0       | 0       | 0      | 0      | 0.4339 | 0      |
| 0.0877      | 0.7016 | 0.1754 | 0.0877 | 0.1754     | 0.1754      | 0       | 0      | 0.5262 | 0.0877 | 0.2631 | 0.0877 | 0      | 0       | 0       | 0      | 0.0877 | 0.1754 | 0      |
| 0.0607      | 0.8193 | 0.1214 | 0      | 0.091      | 0.1214      | 0.0607  | 0.0303 | 0      | 0.0303 | 0.4552 | 0      | 0      | 0       | 0       | 0      | 0      | 0.2731 | 0      |
| 0.3989      | 0.7789 | 0.019  | 0.057  | 0.266      | 0.266       | 0.152   | 0      | 0.038  | 0.152  | 0.076  | 0.019  | 0.057  | 0.133   | 0       | 0      | 0.114  | 0.038  | 0      |
| 0.1123      | 0.7219 | 0.0882 | 0.0481 | 0.3209     | 0.1444      | 0.0321  | 0.008  | 0.3048 | 0.016  | 0.4733 | 0.1123 | 0      | 0       | 0.016   | 0.0241 | 0      | 0.0241 | 0      |
| 0.1292      | 0.0431 | 0      | 0.0215 | 0          | 0           | 0       | 0      | 0      | 0      | 0      | 0      | 0      | 0       | 0       | 0      | 0      | 0.9905 | 0      |
| 0.1528      | 0.2547 | 0.034  | 0.0509 | 0.8999     | 0.1358      | 0       | 0      | 0.017  | 0.017  | 0.1698 | 0.034  | 0.017  | 0       | 0       | 0.017  | 0.2037 | 0.0849 | 0      |
| 0.4049      | 0.5849 | 0      | 0      | 0.225      | 0.225       | 0.045   | 0      | 0      | 0      | 0.18   | 0.045  | 0.045  | 0       | 0.045   | 0.09   | 0.5849 | 0.045  | 0      |
| 0.4376      | 0.5835 | 0      | 0      | 0.2917     | 0           | 0       | 0      | 0.1459 | 0      | 0.4376 | 0      | 0      | 0       | 0       | 0      | 0.2917 | 0.2917 | 0      |
| 0           | 0.2378 | 0      | 0.0297 | 0.0594     | 0.0594      | 0       | 0      | 0.9511 | 0      | 0.0297 | 0.0892 | 0      | 0.0594  | 0       | 0.0297 | 0.1189 | 0      | 0.0594 |
| 0.0873      | 0.3055 | 0      | 0      | 0.5237     | 0.1309      | 0.1309  | 0      | 0.3928 | 0      | 0.6547 | 0.0873 | 0      | 0       | 0       | 0      | 0      | 0      | 0      |
| 0           | 0.3354 | 0.1118 | 0      | 0.1118     | 0.2236      | 0       | 0      | 0.8944 | 0      | 0      | 0.1118 | 0      | 0       | 0       | 0      | 0      | 0      | 0      |
| 0.5781      | 0.1577 | 0      | 0      | 0.0526     | 0           | 0       | 0      | 0.0526 | 0      | 0.0526 | 0      | 0      | 0       | 0       | 0      | 0.1051 | 0.7884 | 0      |
| 0.3881      | 0.4528 | 0.1941 | 0      | 0.3234     | 0.1294      | 0.1941  | 0      | 0      | 0      | 0.5822 | 0.0647 | 0      | 0.3234  | 0       | 0      | 0      | 0      | 0      |
| 0.5604      | 0.6405 | 0      | 0      | 0          | 0.0801      | 0       | 0      | 0.0801 | 0.1601 | 0.0801 | 0      | 0      | 0       | 0       | 0      | 0      | 0.4804 | 0      |
| 0.5733      | 0.5733 | 0      | 0.0358 | 0.0717     | 0.0717      | 0       | 0      | 0      | 0.0358 | 0      | 0      | 0      | 0       | 0       | 0      | 0.0358 | 0.5733 | 0      |
| 0.2847      | 0.8068 | 0.0475 | 0.0949 | 0.2847     | 0.1898      | 0.0475  | 0.0475 | 0.0475 | 0.0475 | 0.2847 | 0.0475 | 0      | 0       | 0       | 0.0949 | 0.0475 | 0.1898 | 0      |
| 0.1753      | 0.701  | 0      | 0.1168 | 0          | 0.0584      | 0       | 0      | 0.6426 | 0      | 0.1168 | 0      | 0.0584 | 0       | 0       | 0      | 0.1753 | 0      | 0      |
| 0           | 0.4243 | 0.1414 | 0      | 0.1414     | 0.2828      | 0       | 0      | 0      | 0      | 0.7071 | 0      | 0      | 0       | 0       | 0.4243 | 0.1414 | 0      | 0      |

A la salida de esta etapa se generará una lista de las personas más similares a un usuario en particular conforme a sus distancias musicales, a su vez el algoritmo K-mean nos indicará a qué grupo pertenece cada usuario, con lo cual se deja de lado el principio de transitividad para un sistema de recomendación y se hace énfasis en la relación que existe entre los intereses de las personas.

En las figuras 3.23 y 3.24 se muestra un panorama detallado de la arquitectura completa donde todo puede verse integrado como dos fases debido a lo estrechamente relacionadas que se encuentran sus 4 etapas

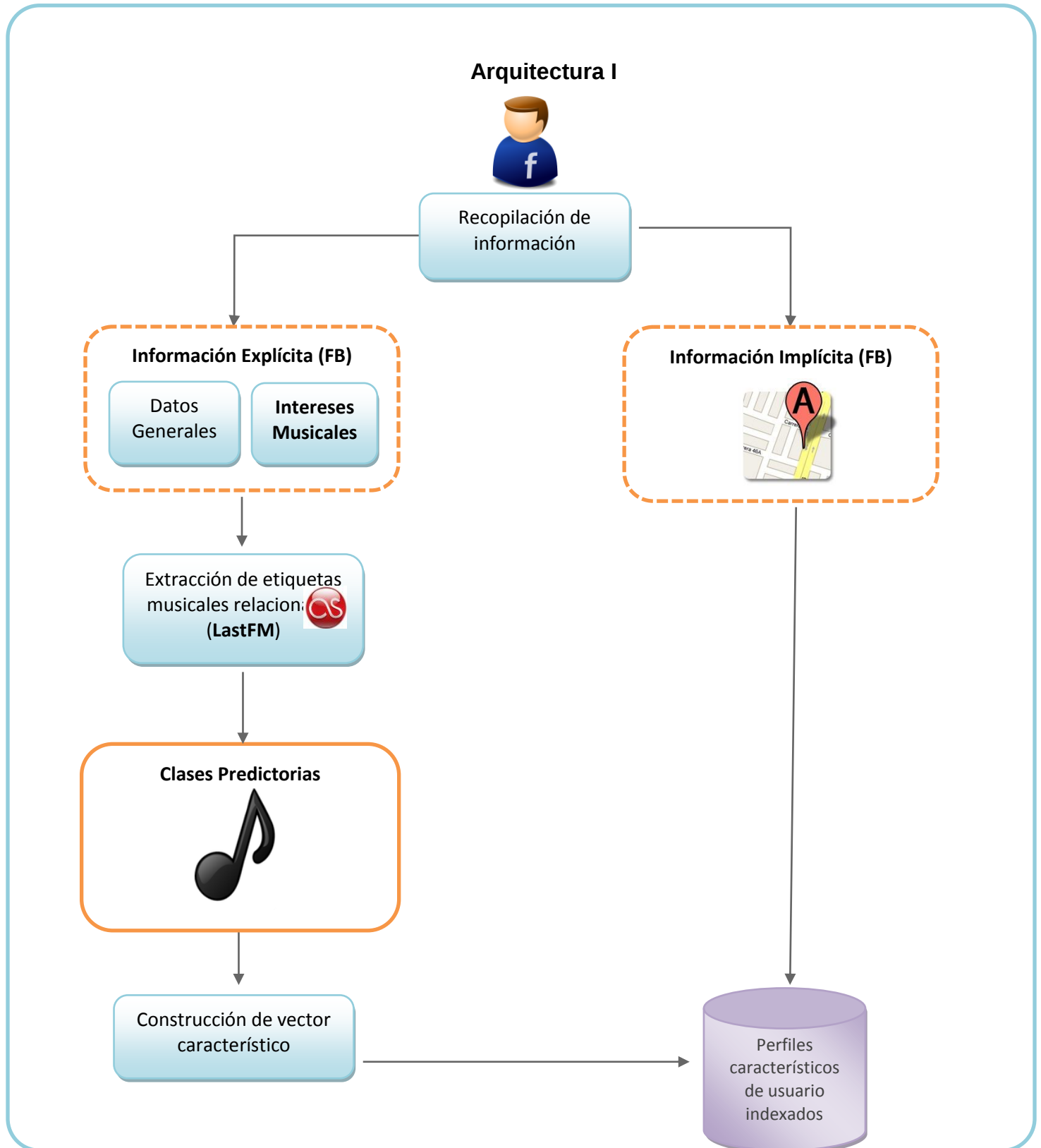


Figura 3.22 Arquitectura I

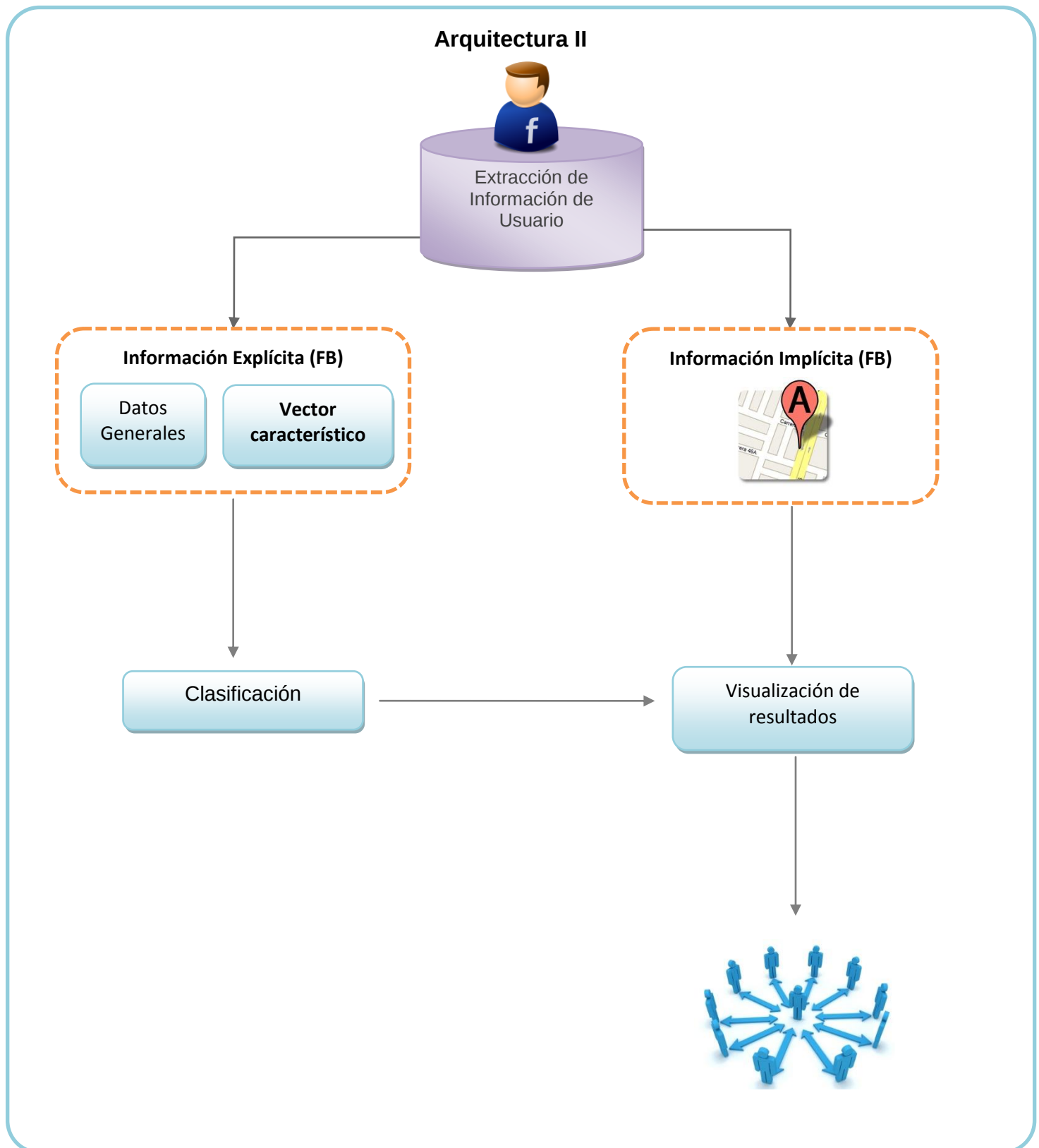


Figura 3.23 Arquitectura II

### 3.3 Diagramas de caso de uso

Un caso de uso es una secuencia de interacciones entre un sistema y alguien o algo que usa alguno de sus servicios [30].

El sistema sólo cuenta con un actor el cual es el usuario del sistema y cuenta con tres acciones: recopilar información, extraer características, clasificar.

#### 3.3.1 Diagrama de caso de uso general

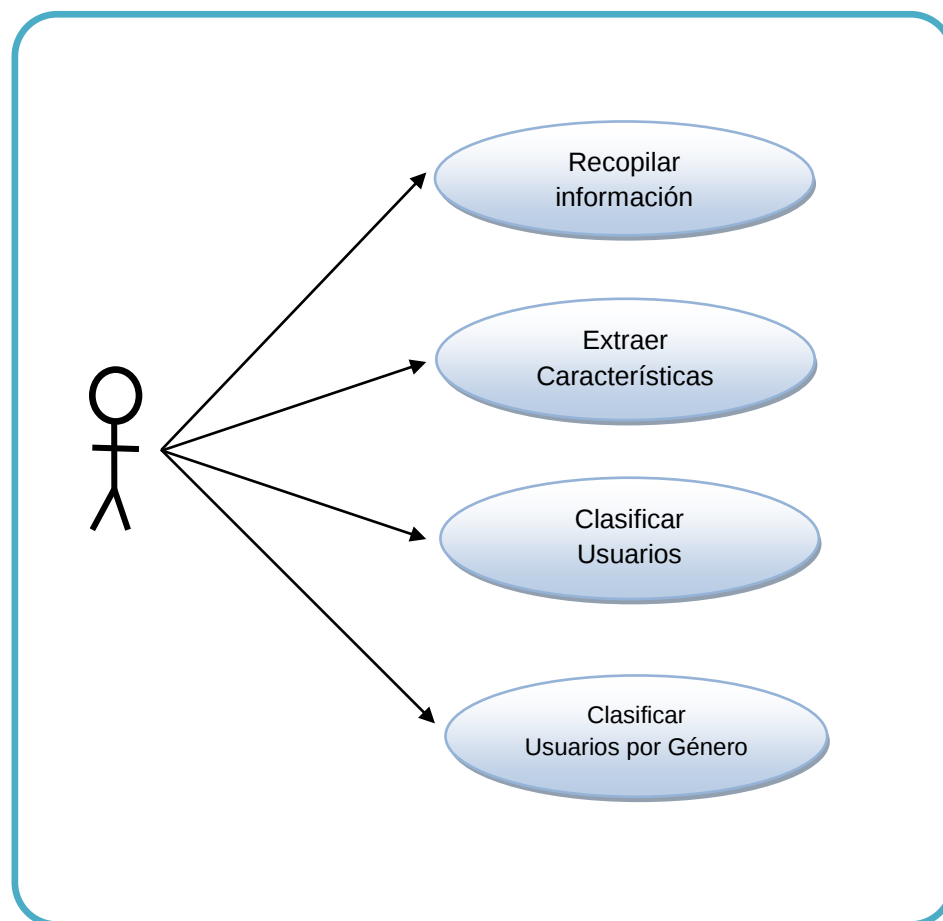


Figura 3.24 Diagrama caso de uso general

### 3.3.2 Diagrama caso de uso recopilar información.

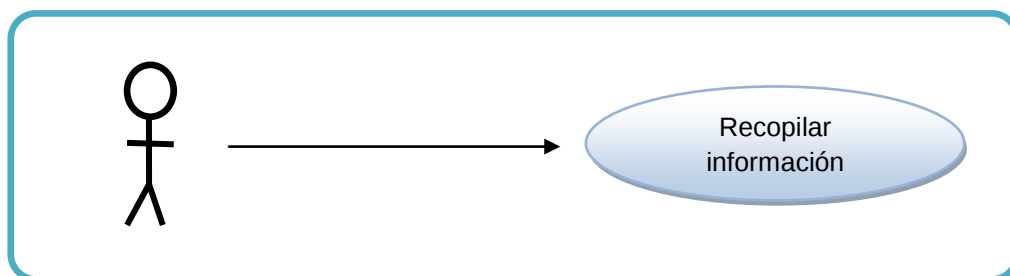


Figura 3.25 Diagrama caso de uso recopilar información

Tabla 3.7 Descripción caso de uso recopilar información

| No. Caso de Uso<br>01                      | Nombre Caso de Uso<br>Recopilar información                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ACTOR</b>                               | Usuario                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                           |
| <b>OBJETIVO</b>                            | Recopilar la información implícita y explícita del usuario de Facebook.                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                           |
| <b>PRECONDICIONES</b>                      | El sistema debe tener privilegio para acceder a la información personal del usuario.                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                           |
| <b>POSCONDICIONES</b>                      | El sistema contará con la información del usuario                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                           |
| <b>FLUJO DE EVENTOS</b>                    | Actividades del Actor                                                                                                                                                                                                                                                                                                                                                                   | <ol style="list-style-type: none"> <li>1. El usuario ingresa a la aplicación.</li> <li>2. El usuario selecciona la opción recopilar información.</li> </ol>                                                                                                                                                                                                               |
|                                            | Respuesta Del Sistema                                                                                                                                                                                                                                                                                                                                                                   | <ol style="list-style-type: none"> <li>1. El sistema verifica que su información no haya sido recopilada previamente.</li> <li>2. El sistema verifica que cuente con al menos diez instancias musicales.</li> <li>3. El sistema concede la recolección de información.</li> <li>4. El sistema almacena en BD la información implícita y explícita del usuario.</li> </ol> |
| <b>MANEJO DE SITUACIONES EXCEPCIONALES</b> | <ul style="list-style-type: none"> <li>✓ Si la información del usuario ha sido recopilada previamente se le notifica e invita a continuar con el paso de extracción. La recolección no vuelve a efectuarse.</li> <li>✓ Si el usuario no cuenta con al menos diez instancias musicales se le notifica que es un requerimiento de la aplicación. La recolección no se efectúa.</li> </ul> |                                                                                                                                                                                                                                                                                                                                                                           |

### 3.3.3 Diagrama caso de uso extraer características

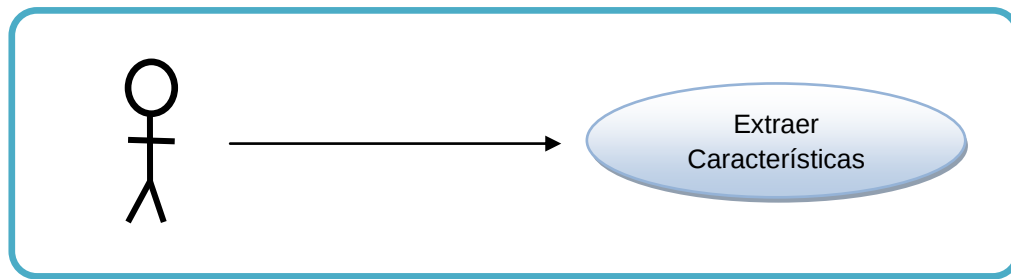


Figura 3.26 Diagrama caso de uso extraer características

Tabla 3.8 Descripción caso de uso extraer características

| No. Caso de Uso<br>02                      | Nombre Caso de Uso<br>Extraer Características                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|--------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ACTOR</b>                               | Usuario                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>OBJETIVO</b>                            | Generar el vector característico que represente el perfil musical del usuario mediante la extracción.                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>PRECONDICIONES</b>                      | El sistema debe tener privilegio para acceder a la información personal del usuario.<br>El usuario debe previamente haber recopilado la información implícita y explícita del usuario                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>POSCONDICIONES</b>                      | El sistema contará con el vector característico del usuario                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <b>FLUJO DE EVENTOS</b>                    | Actividades del Actor                                                                                                                                                                                                                                                                                                        | 1. El usuario ingresa a la aplicación<br>3. El usuario selecciona la opción extraer características.                                                                                                                                                                                                                                                                                                                                 |
|                                            | Respuesta Del Sistema                                                                                                                                                                                                                                                                                                        | 1. El sistema verifica que la información del usuario haya sido recopilada previamente.<br>2. El sistema verifica que su vector característico no haya sido generado previamente (extracción previa).<br>3. El sistema concede la extracción de características<br>4. El sistema almacena en BD el perfil característico del usuario.<br>5. El sistema despliega el vector construido y la gráfica que representa dicha información. |
| <b>MANEJO DE SITUACIONES EXCEPCIONALES</b> | ✓ Si la información del usuario no ha sido recopilada previamente se le notifica al usuario que debe hacerlo para continuar con el paso de extracción. La extracción no se efectúa.<br>✓ Si la información del usuario ha sido extraída previamente se le notifica al usuario e invita a continuar al paso de clasificación. |                                                                                                                                                                                                                                                                                                                                                                                                                                      |

### 3.3.4 Diagrama caso de uso clasificar

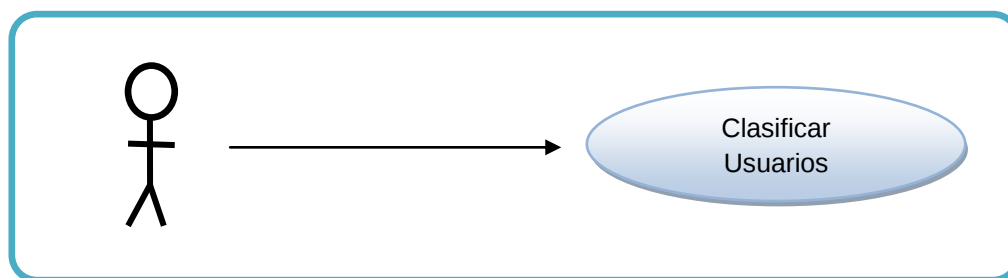


Figura 3.27 Diagrama caso de uso clasificar

Tabla 3.9 Descripción caso de uso clasificar usuarios

| No. Caso de Uso<br>03                      | Nombre Caso de Uso<br>Clasificar Usuarios                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ACTOR</b>                               | Usuario                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                             |
| <b>OBJETIVO</b>                            | Generar el listado de las personas con las cuales tenemos una mayor similitud respecto a intereses musicales                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                             |
| <b>PRECONDICIONES</b>                      | <p>El sistema debe tener privilegio para acceder a la información personal del usuario.</p> <p>El usuario debe previamente haber recopilado la información implícita y explícita del usuario.</p> <p>El sistema debe haber previamente extraído las características de la información recopilada.</p> |                                                                                                                                                                                                                                                                                                                             |
| <b>POSCONDICIONES</b>                      | El sistema contará con la relación de los usuarios más similares al usuario que realiza la petición.                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                             |
| <b>FLUJO DE EVENTOS</b>                    | Actividades del Actor                                                                                                                                                                                                                                                                                 | <ol style="list-style-type: none"> <li>1. El usuario ingresa a la aplicación</li> <li>2. El usuario selecciona la opción clasificar usuarios.</li> </ol>                                                                                                                                                                    |
|                                            | Respuesta Del Sistema                                                                                                                                                                                                                                                                                 | <ol style="list-style-type: none"> <li>1. El sistema verifica que el vector característico del usuario haya sido extraído previamente.</li> <li>2. El sistema concede la clasificación.</li> <li>3. El sistema calcula las distancias euclidianas y las ordena.</li> <li>4. El sistema despliega los resultados.</li> </ol> |
| <b>MANEJO DE SITUACIONES EXCEPCIONALES</b> | <p>✓ Si el vector característico del usuario no ha sido generado previamente se le notifica al usuario que debe hacerlo para continuar con el paso de clasificación. La clasificación no se efectúa.</p>                                                                                              |                                                                                                                                                                                                                                                                                                                             |

### 3.3.5 Diagrama caso de uso clasificar por género

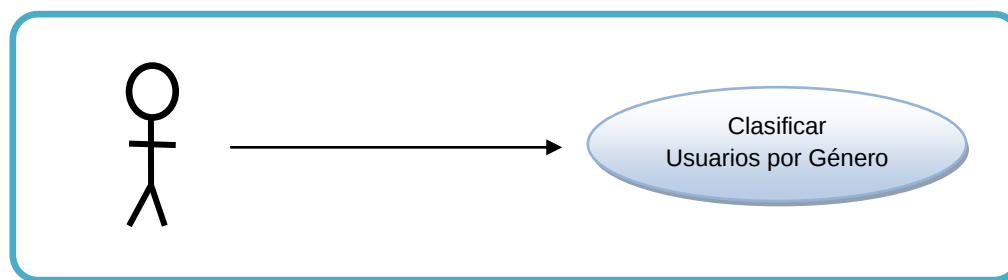


Figura 3.28 Diagrama caso de uso clasificar por género

Tabla 3.10 Descripción caso de uso clasificar usuarios por género

| No. Caso de Uso<br>04                      | Nombre Caso de Uso<br>Clasificar Usuarios por género                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ACTOR</b>                               | Usuario                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                             |
| <b>OBJETIVO</b>                            | Generar el listado de las personas con las cuales tenemos una mayor similitud respecto a un género en particular                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                             |
| <b>PRECONDICIONES</b>                      | <p>El sistema debe tener privilegio para acceder a la información personal del usuario.</p> <p>El usuario debe previamente haber recopilado la información implícita y explícita del usuario.</p> <p>El sistema debe haber previamente extraído las características de la información recopilada.</p> |                                                                                                                                                                                                                                                                                                                             |
| <b>POSCONDICIONES</b>                      | El sistema contará con la relación de los usuarios más similares al usuario que realiza la petición.                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                             |
| <b>FLUJO DE EVENTOS</b>                    | Actividades del Actor                                                                                                                                                                                                                                                                                 | <ol style="list-style-type: none"> <li>1. El usuario ingresa a la aplicación</li> <li>2. El usuario selecciona la opción clasificar usuarios por género.</li> <li>3. El usuario selecciona un género</li> </ol>                                                                                                             |
|                                            | Respuesta Del Sistema                                                                                                                                                                                                                                                                                 | <ol style="list-style-type: none"> <li>1. El sistema verifica que el vector característico del usuario haya sido extraído previamente.</li> <li>2. El sistema concede la clasificación.</li> <li>5. El sistema calcula las distancias euclidianas y las ordena.</li> <li>6. El sistema despliega los resultados.</li> </ol> |
| <b>MANEJO DE SITUACIONES EXCEPCIONALES</b> | <p>✓ Si el vector característico del usuario no ha sido generado previamente se le notifica al usuario que debe hacerlo para continuar con el paso de clasificación. La clasificación no se efectúa.</p>                                                                                              |                                                                                                                                                                                                                                                                                                                             |

### 3.4 Diagramas de secuencia general

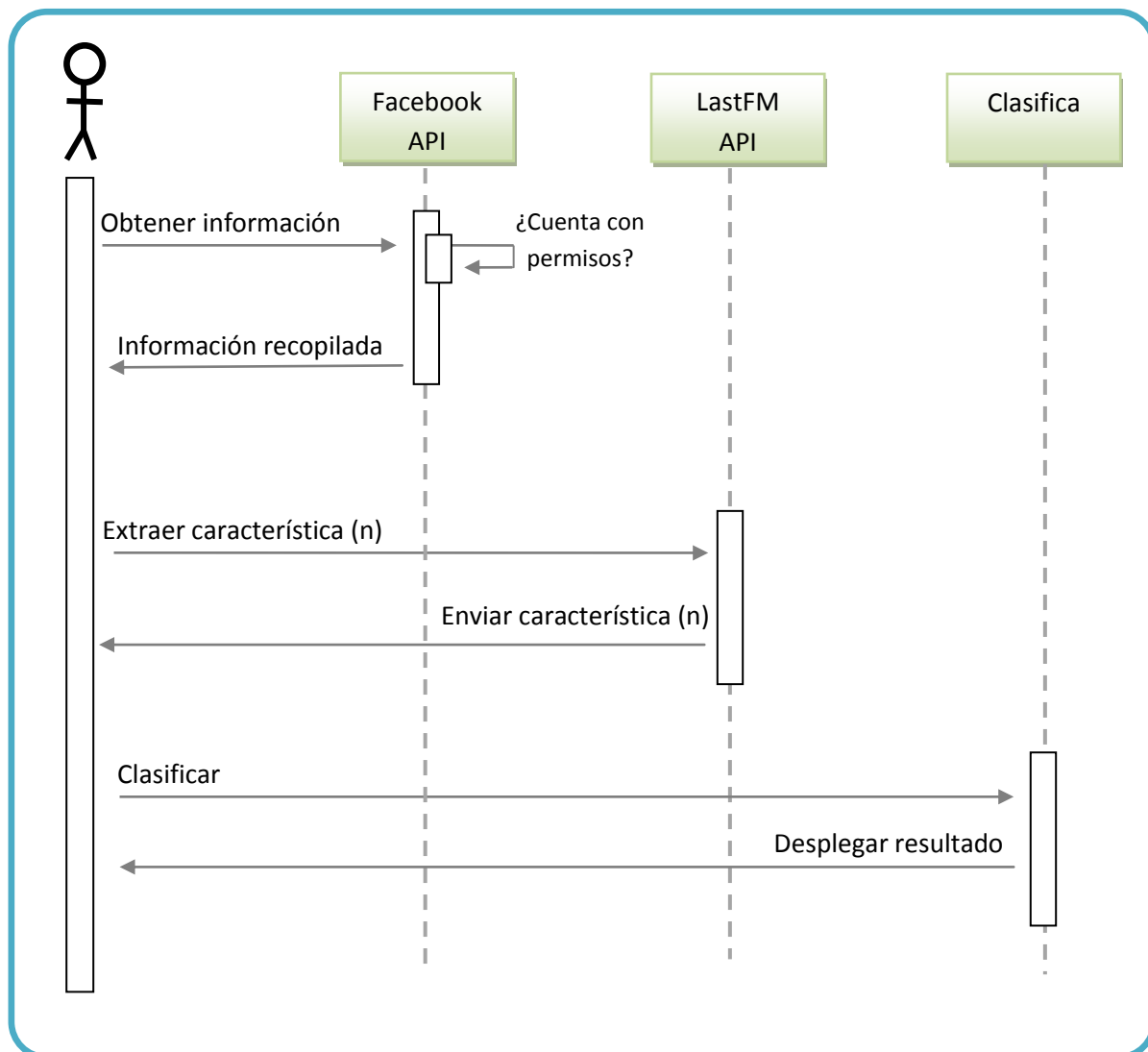


Figura 3.29 Diagrama de secuencia

### 3.5 Conclusiones

La metodología propuesta resulta adecuada para el análisis de la similitud de usuarios, a su vez resulta flexible debido a que cada uno de los módulos puede ser implementado de diversas maneras; cabe mencionar que al guardar en base de datos los resultados de cada una de las etapas sirve para aumentar la eficiencia del procesamiento ya que se deja parte del trabajo al sistema gestor de base de datos en lugar de mantener peticiones al servidor de aplicaciones

# CAPÍTULO IV

## PRUEBAS Y RESULTADOS

***" Me sentaré sin rodeos,  
esperando el regalo  
del sonido y la visión."  
David Bowie***

## Capítulo 4 Pruebas y resultados

### 4.1 Introducción

En el presente capítulo nos enfocaremos en crear un escenario para poder realizar pruebas a los casos de uso y con esto realizar un análisis sobre los resultados de la metodología propuesta.

Para las pruebas se publicó la aplicación en Facebook y se dejó libre a que cualquier persona pudiera hacer uso de ella, sin embargo los usuarios deben cumplir una restricción la cual es que contengan 19 instancias musicales en su perfil; es decir, deben tener indicados 19 músicos que sean de su agrado, se selecciona este número debido a que el vector característico que hemos diseñado cuenta con este mismo número de posiciones donde cada elemento corresponde a un género por lo cual si el usuario brinda 19 ejemplos de músicos que le interesan el peor de los escenarios sería que se distribuya uno por cada género al momento de extraer las características, lo que ya nos indica una preferencia aunque esta sea la misma por cada género.

Es importante resaltar el hecho que al final de cada una de las etapas del sistema la información del usuario se va almacenando en base de datos por lo cual el primer resultado interesante consiste en analizar el número de usuarios por etapa para conocer el grado de interés en la aplicación, dicha información se muestra en la Tabla 4.1

Tabla 4.1 Usuarios por etapa

| <b>Etapas</b>                  | <b>Número de usuarios</b> | <b>Porcentaje del Total</b> |
|--------------------------------|---------------------------|-----------------------------|
| <b>Visitaron la Aplicación</b> | 71                        | 100%                        |
| <b>Recopilación</b>            | 54                        | 76%                         |
| <b>Extracción</b>              | 50                        | 70%                         |
| <b>Clasificación</b>           | 50                        | 70%                         |

Con los resultados de la tabla anterior, las siguientes pruebas se enfocarán en analizar resultados con los 50 usuarios que realizaron el proceso completo; el porcentaje que se indica en la tabla anterior hace referencia al cálculo tomando como 100% a los 71 usuarios que al menos visitaron la aplicación.

Para presentar los resultados se le asignará un alias a los usuarios que hicieron uso de la aplicación, los cuales cumplirán con la nomenclatura  $U_n$  siendo  $n$  su número de aparición únicamente, este subíndice no es aginado por ningún otro motivo; en la Tabla 4.2 se presenta la relación de los usuarios con los cuales se realizan las pruebas.

**Tabla 4.2 Datos de usuarios de prueba**

| Número de usuario | Nombre de usuario               | ID usuario en Facebook | #Instancias musicales |
|-------------------|---------------------------------|------------------------|-----------------------|
| U1                | Isaac Husaí Arroyo Mejía        | 571863636              | 173                   |
| U2                | Tatiana Gs                      | 618809010              | 184                   |
| U3                | Reex Gallardo                   | 1522720504             | 406                   |
| U4                | Roberto Castañeda               | 851834603              | 21                    |
| U5                | Liza Simpson                    | 592823248              | 174                   |
| U6                | Miri Balbuena                   | 777998233              | 32                    |
| U7                | Daniela Medina                  | 813333367              | 71                    |
| U8                | Sara Estudillo Arriaga          | 540484867              | 91                    |
| U9                | Samuel Mejía                    | 100001586883631        | 69                    |
| U10               | Xochitl Purple                  | 791633544              | 26                    |
| U11               | Teran Corona Liliana            | 100003356872849        | 46                    |
| U12               | Krla Martx                      | 587411837              | 87                    |
| U13               | Luis Pacheco                    | 678063011              | 66                    |
| U14               | Mariana Wolf                    | 519925490              | 94                    |
| U15               | Alesi Morquecho                 | 100002215682731        | 21                    |
| U16               | Michell Itallekzi               | 722334972              | 85                    |
| U17               | Adagioo Adkiins                 | 100001748546453        | 20                    |
| U18               | Daniela Avila Vázquez           | 765242717              | 99                    |
| U19               | Sergio Marquez                  | 776709792              | 21                    |
| U20               | Raúl A. García Cárdenas         | 649311912              | 81                    |
| U21               | Axl Camacho                     | 617333195              | 74                    |
| U22               | Jessica Santiago                | 618418444              | 60                    |
| U23               | BränDy González                 | 100002681766038        | 76                    |
| U24               | Marcelo Barrera                 | 567967574              | 57                    |
| U25               | Jorge Mendoza                   | 548334940              | 60                    |
| U26               | Flavia Zorrilla Drago           | 658654173              | 94                    |
| U27               | Fernando Rocha                  | 100001046993654        | 32                    |
| U28               | Antonio De Jesús Méndez         | 100000490407299        | 72                    |
| U29               | Raul Antonio Villamil Rodriguez | 1259811024             | 59                    |
| U30               | Katt Marañosky                  | 1006897230             | 72                    |
| U31               | Daniel Correa                   | 840484003              | 56                    |
| U32               | Dara Rivera                     | 577323716              | 139                   |
| U33               | Deivid Estévez Campos           | 1087791576             | 77                    |
| U34               | Eugenio Vargas Becerril         | 1676751945             | 188                   |
| U35               | Jorge Israel Toledo Alvarado    | 515068207              | 23                    |
| U36               | Miriam Haller                   | 614165604              | 148                   |
| U37               | K Gaona M. Oom                  | 668141890              | 52                    |
| U38               | Feer Luckeer                    | 100003063429181        | 102                   |
| U39               | Luis Cabrera Rivera             | 100003093019476        | 33                    |
| U40               | Adhyn Velasco                   | 725338447              | 56                    |
| U41               | Coneja Perez                    | 647351963              | 28                    |
| U42               | Juan Castelán                   | 719006208              | 19                    |
| U43               | Jessica Calixto                 | 622164531              | 103                   |
| U44               | Luis Rivera                     | 1294494611             | 27                    |
| U45               | Ava Dew                         | 100000401684873        | 156                   |
| U46               | Poni K Montalvo                 | 662635542              | 65                    |
| U47               | Israel Murillo                  | 100001106636751        | 66                    |
| U48               | Leonardo Garcia                 | 100004492050001        | 24                    |
| U49               | Marco Moreno Ibarra             | 1220539913             | 61                    |
| U50               | Mar Leiden                      | 100002557899549        | 84                    |

En la figura 4.1 se muestran los resultados de las preferencias de los usuarios según cada uno de los géneros seleccionados para representar nuestro vector característico, cabe señalar que los usuarios están ordenados según la distancia Euclidiana respecto a  $U_1$  de izquierda a derecha.

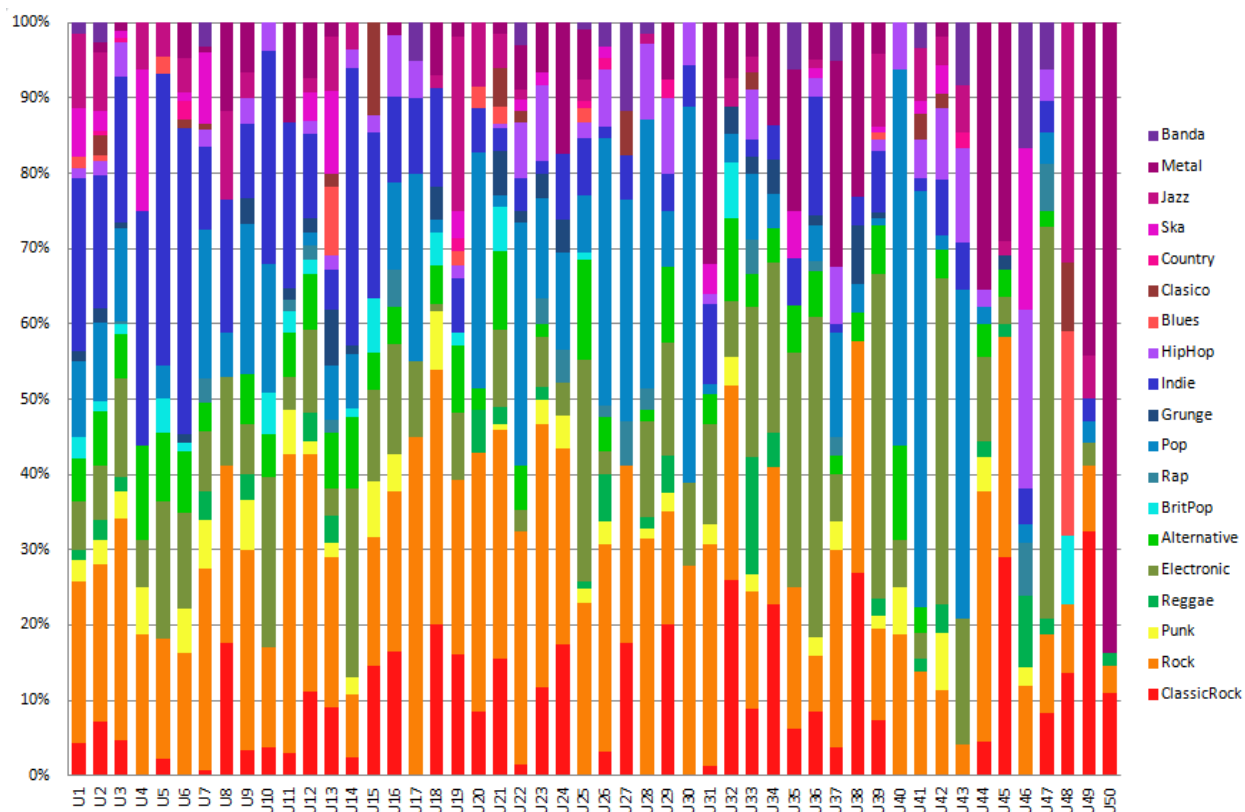


Figura 4.1 Vectores característicos normalizados ordenados por distancia

Como se muestra en la gráfica, se puede visualizar que mientras las barras se alejan del origen del eje Y la distribución de sus géneros va cambiando lo que indica diferencia en las preferencias lo cual coincide con el cálculo de su distancia euclidiana. Los vectores más cercanos al lado izquierdo tienen una mayor preferencia indicada en tonos rojos mientras que los vectores más alejados tienen una mayor incidencia de tonos morados. Lo anterior se puede observar de manera más clara en la figura 4.2 la cual muestra los cinco usuarios más cercanos y los cinco más lejanos a  $U_1$

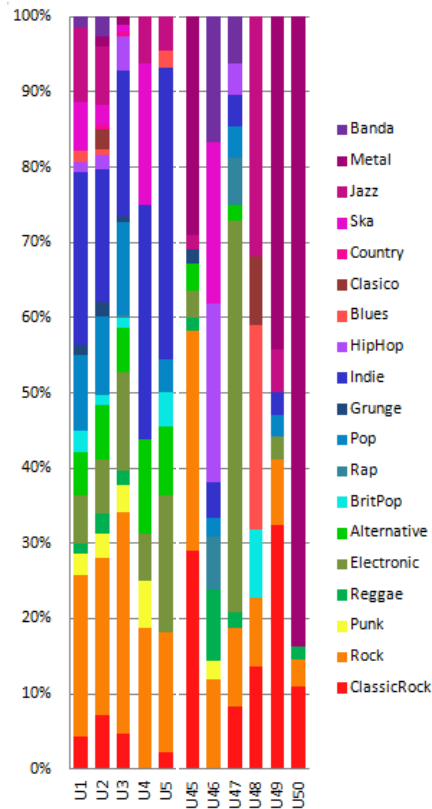
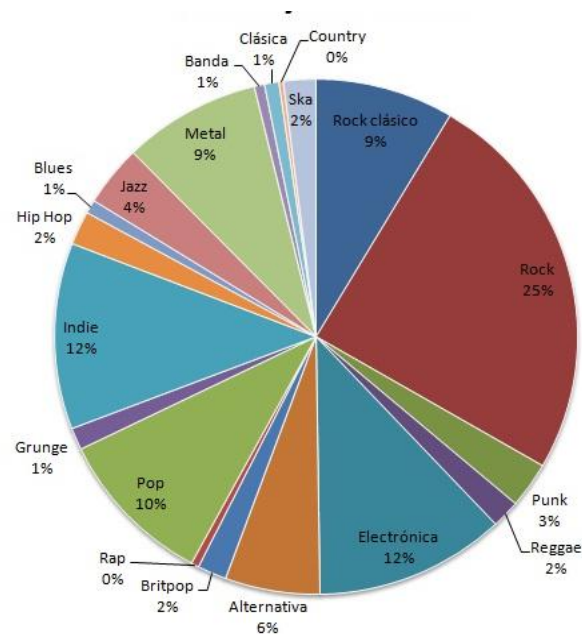


Figura 4.2 Vectores más cercanos y más lejanos según distancia euclidiana

En la figura 4.3 se muestra el porcentaje de ocurrencia de cada uno de los géneros utilizados para el vector característico.

Figura 4.3 Total de géneros en gráfica



En la Tabla 4.3 se presenta la tabla que contiene el vector característico de cada uno de los usuarios utilizados en las pruebas de los casos de uso, tomando como punto de partida al usuario U<sub>1</sub>.

**Tabla 4.3 Vectores característicos de prueba**

| Número de usuario | ClassicRock | Rock | Punk | Reggae | Electronic | Alternative | BritPop | Rap | Pop | Grunge | Indie | HipHop | Blues | Clasico | Country | Ska | Jazz | Metal | Banda |
|-------------------|-------------|------|------|--------|------------|-------------|---------|-----|-----|--------|-------|--------|-------|---------|---------|-----|------|-------|-------|
| U1                | 6           | 30   | 4    | 2      | 9          | 8           | 4       | 0   | 14  | 2      | 32    | 2      | 2     | 0       | 0       | 9   | 14   | 0     | 2     |
| U2                | 11          | 32   | 5    | 4      | 11         | 11          | 2       | 0   | 16  | 3      | 27    | 3      | 1     | 4       | 1       | 4   | 12   | 2     | 4     |
| U3                | 14          | 90   | 11   | 6      | 40         | 18          | 4       | 1   | 38  | 2      | 59    | 14     | 0     | 0       | 2       | 3   | 0    | 3     | 0     |
| U4                | 0           | 3    | 1    | 0      | 1          | 2           | 0       | 0   | 0   | 0      | 5     | 0      | 0     | 0       | 0       | 3   | 1    | 0     | 0     |
| U5                | 1           | 34   | 8    | 5      | 10         | 5           | 0       | 4   | 25  | 0      | 14    | 3      | 0     | 1       | 0       | 12  | 0    | 1     | 4     |
| U6                | 1           | 8    | 2    | 1      | 2          | 2           | 0       | 0   | 6   | 1      | 3     | 1      | 0     | 0       | 0       | 0   | 1    | 2     | 0     |
| U7                | 2           | 7    | 0    | 0      | 12         | 3           | 3       | 0   | 9   | 0      | 15    | 2      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| U8                | 2           | 7    | 2    | 0      | 21         | 8           | 1       | 0   | 6   | 1      | 31    | 2      | 0     | 0       | 0       | 0   | 3    | 0     | 0     |
| U9                | 10          | 13   | 3    | 0      | 9          | 3           | 0       | 3   | 7   | 0      | 7     | 5      | 0     | 0       | 0       | 0   | 0    | 1     | 0     |
| U10               | 0           | 9    | 0    | 0      | 2          | 0           | 0       | 0   | 5   | 0      | 2     | 1      | 0     | 0       | 0       | 0   | 0    | 0     | 1     |
| U11               | 3           | 12   | 0    | 2      | 0          | 1           | 0       | 0   | 11  | 0      | 2     | 0      | 1     | 0       | 0       | 0   | 3    | 0     | 0     |
| U12               | 1           | 21   | 0    | 0      | 2          | 4           | 0       | 0   | 22  | 1      | 3     | 5      | 0     | 1       | 0       | 1   | 1    | 4     | 2     |
| U13               | 7           | 21   | 2    | 1      | 4          | 1           | 0       | 2   | 8   | 2      | 1     | 6      | 0     | 0       | 0       | 1   | 0    | 4     | 0     |
| U14               | 2           | 18   | 2    | 4      | 2          | 3           | 0       | 1   | 23  | 0      | 1     | 5      | 0     | 0       | 1       | 1   | 0    | 0     | 2     |
| U15               | 3           | 4    | 0    | 0      | 0          | 0           | 0       | 1   | 5   | 0      | 1     | 0      | 0     | 1       | 0       | 0   | 0    | 0     | 2     |
| U16               | 0           | 22   | 1    | 1      | 9          | 1           | 0       | 2   | 25  | 0      | 0     | 7      | 0     | 0       | 0       | 0   | 1    | 0     | 1     |
| U17               | 0           | 5    | 0    | 0      | 2          | 0           | 0       | 0   | 9   | 0      | 1     | 1      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| U18               | 3           | 21   | 3    | 0      | 5          | 2           | 0       | 2   | 11  | 0      | 1     | 6      | 0     | 0       | 0       | 0   | 0    | 22    | 4     |
| U19               | 0           | 3    | 1    | 0      | 1          | 2           | 0       | 0   | 8   | 0      | 0     | 1      | 0     | 0       | 0       | 0   | 0    | 0     | 0     |
| U20               | 0           | 8    | 0    | 1      | 2          | 2           | 0       | 0   | 32  | 0      | 1     | 3      | 0     | 2       | 0       | 1   | 4    | 0     | 2     |
| U21               | 0           | 6    | 4    | 2      | 23         | 2           | 0       | 0   | 1   | 0      | 4     | 5      | 0     | 1       | 0       | 2   | 2    | 1     | 0     |
| U22               | 0           | 2    | 0    | 0      | 8          | 0           | 0       | 0   | 21  | 0      | 3     | 6      | 0     | 0       | 1       | 0   | 3    | 0     | 4     |
| U23               | 0           | 5    | 1    | 4      | 0          | 0           | 0       | 3   | 1   | 0      | 2     | 10     | 0     | 0       | 0       | 9   | 0    | 0     | 7     |
| U24               | 4           | 5    | 0    | 1      | 25         | 1           | 0       | 3   | 2   | 0      | 2     | 2      | 0     | 0       | 0       | 0   | 0    | 0     | 3     |
| U25               | 1           | 7    | 0    | 0      | 8          | 4           | 2       | 0   | 2   | 0      | 17    | 0      | 1     | 0       | 0       | 0   | 2    | 0     | 0     |
| U26               | 0           | 14   | 5    | 0      | 11         | 7           | 1       | 0   | 0   | 1      | 35    | 0      | 0     | 1       | 2       | 1   | 4    | 4     | 0     |
| U27               | 3           | 4    | 0    | 0      | 2          | 0           | 0       | 0   | 1   | 0      | 3     | 0      | 0     | 0       | 0       | 0   | 2    | 2     | 0     |
| U28               | 2           | 27   | 4    | 0      | 3          | 4           | 2       | 1   | 0   | 1      | 15    | 0      | 0     | 0       | 0       | 0   | 0    | 9     | 0     |
| U29               | 6           | 17   | 1    | 2      | 6          | 4           | 1       | 1   | 1   | 1      | 6     | 1      | 0     | 0       | 0       | 2   | 1    | 4     | 0     |
| U30               | 5           | 11   | 1    | 2      | 2          | 4           | 0       | 1   | 4   | 4      | 3     | 1      | 5     | 1       | 0       | 6   | 4    | 1     | 0     |
| U31               | 6           | 7    | 3    | 0      | 5          | 2           | 3       | 0   | 0   | 0      | 9     | 1      | 0     | 5       | 0       | 0   | 0    | 0     | 0     |
| U32               | 23          | 39   | 9    | 0      | 1          | 6           | 5       | 0   | 2   | 5      | 15    | 0      | 0     | 0       | 0       | 0   | 2    | 8     | 0     |
| U33               | 9           | 13   | 0    | 0      | 5          | 5           | 1       | 0   | 0   | 0      | 4     | 1      | 1     | 0       | 1       | 2   | 13   | 1     | 0     |
| U34               | 21          | 41   | 1    | 3      | 14         | 14          | 8       | 0   | 2   | 8      | 4     | 1      | 3     | 7       | 0       | 0   | 6    | 2     | 0     |
| U35               | 4           | 6    | 1    | 0      | 1          | 0           | 0       | 1   | 3   | 1      | 2     | 0      | 0     | 0       | 0       | 0   | 0    | 4     | 0     |
| U36               | 0           | 24   | 2    | 1      | 31         | 14          | 1       | 0   | 8   | 0      | 8     | 2      | 2     | 0       | 1       | 0   | 3    | 7     | 1     |
| U37               | 8           | 6    | 1    | 2      | 6          | 4           | 0       | 0   | 3   | 0      | 2     | 4      | 0     | 0       | 1       | 0   | 0    | 3     | 0     |
| U38               | 1           | 22   | 2    | 0      | 10         | 3           | 0       | 0   | 1   | 0      | 8     | 1      | 0     | 0       | 0       | 3   | 0    | 24    | 0     |
| U39               | 7           | 7    | 1    | 0      | 2          | 3           | 2       | 0   | 1   | 1      | 0     | 0      | 0     | 0       | 0       | 0   | 1    | 2     | 0     |
| U40               | 4           | 7    | 1    | 7      | 9          | 2           | 0       | 2   | 4   | 1      | 1     | 3      | 0     | 1       | 0       | 0   | 1    | 2     | 0     |
| U41               | 5           | 4    | 0    | 1      | 5          | 1           | 0       | 0   | 1   | 1      | 1     | 0      | 0     | 0       | 0       | 0   | 0    | 3     | 0     |
| U42               | 1           | 3    | 0    | 0      | 5          | 1           | 0       | 0   | 0   | 0      | 1     | 0      | 0     | 0       | 0       | 1   | 0    | 3     | 1     |
| U43               | 7           | 6    | 2    | 0      | 35         | 5           | 0       | 1   | 4   | 1      | 13    | 2      | 0     | 0       | 0       | 1   | 1    | 4     | 0     |
| U44               | 7           | 8    | 0    | 0      | 0          | 1           | 0       | 0   | 1   | 2      | 1     | 0      | 0     | 0       | 0       | 0   | 0    | 6     | 0     |
| U45               | 9           | 15   | 2    | 3      | 53         | 8           | 0       | 0   | 1   | 1      | 10    | 2      | 1     | 0       | 0       | 1   | 12   | 5     | 0     |
| U46               | 2           | 15   | 2    | 1      | 5          | 2           | 0       | 0   | 1   | 0      | 0     | 1      | 0     | 0       | 0       | 0   | 0    | 16    | 0     |
| U47               | 16          | 16   | 0    | 1      | 2          | 2           | 0       | 0   | 0   | 1      | 0     | 0      | 0     | 0       | 0       | 0   | 1    | 16    | 0     |
| U48               | 3           | 2    | 0    | 0      | 0          | 0           | 2       | 0   | 0   | 0      | 0     | 0      | 6     | 2       | 0       | 0   | 7    | 0     | 0     |
| U49               | 11          | 3    | 0    | 0      | 1          | 0           | 0       | 0   | 1   | 0      | 1     | 0      | 0     | 0       | 0       | 0   | 2    | 15    | 0     |
| U50               | 6           | 2    | 0    | 1      | 0          | 0           | 0       | 0   | 0   | 0      | 0     | 0      | 0     | 0       | 0       | 0   | 0    | 46    | 0     |

En la Tabla 4.3 se presenta la tabla que contiene el vector característico normalizado de cada uno de los usuarios utilizados en las pruebas de los casos de uso, tomando como punto de partida al usuario U<sub>1</sub>.

**Tabla 4.4 Vectores característicos normalizados**

| Usuario | ClassicRock | Rock  | Punk  | Reggae | Electronic | Alternative | BritPop | Rap   | Pop   | Grunge | Indie | HipHop | Blues | Clasico | Country | Ska   | Jazz  | Metal | Banda  |
|---------|-------------|-------|-------|--------|------------|-------------|---------|-------|-------|--------|-------|--------|-------|---------|---------|-------|-------|-------|--------|
| U1      | 0,117       | 0,585 | 0,078 | 0,039  | 0,1755     | 0,156       | 0,078   | 0     | 0,273 | 0,039  | 0,624 | 0,039  | 0,039 | 0       | 0       | 0,176 | 0,273 | 0     | 0,039  |
| U2      | 0,2144      | 0,624 | 0,097 | 0,078  | 0,2144     | 0,2144      | 0,039   | 0     | 0,312 | 0,0585 | 0,526 | 0,0585 | 0,02  | 0,078   | 0,0195  | 0,078 | 0,234 | 0,039 | 0,078  |
| U3      | 0,1123      | 0,722 | 0,088 | 0,0481 | 0,3209     | 0,1444      | 0,0321  | 0,008 | 0,305 | 0,016  | 0,473 | 0,1123 | 0     | 0       | 0,016   | 0,024 | 0     | 0,024 | 0      |
| U4      | 0           | 0,424 | 0,141 | 0      | 0,1414     | 0,2828      | 0       | 0     | 0     | 0      | 0,707 | 0      | 0     | 0       | 0       | 0,424 | 0,141 | 0     | 0      |
| U5      | 0,0481      | 0,337 | 0     | 0      | 0,3849     | 0,1925      | 0,0962  | 0     | 0,096 | 0      | 0,818 | 0      | 0,048 | 0       | 0       | 0     | 0,096 | 0     | 0      |
| U6      | 0           | 0,344 | 0,123 | 0      | 0,2703     | 0,172       | 0,0246  | 0     | 0     | 0,0246 | 0,86  | 0      | 0     | 0,0246  | 0,0491  | 0,025 | 0,098 | 0,098 | 0      |
| U7      | 0,0205      | 0,697 | 0,164 | 0,1025 | 0,205      | 0,1025      | 0       | 0,082 | 0,513 | 0      | 0,287 | 0,0615 | 0     | 0,0205  | 0       | 0,246 | 0     | 0,021 | 0,082  |
| U8      | 0,4376      | 0,584 | 0     | 0      | 0,2917     | 0           | 0       | 0     | 0,146 | 0      | 0,438 | 0      | 0     | 0       | 0       | 0     | 0,292 | 0,292 | 0      |
| U9      | 0,0877      | 0,702 | 0,175 | 0,0877 | 0,1754     | 0           | 0       | 0     | 0,526 | 0,0877 | 0,263 | 0,0877 | 0     | 0       | 0       | 0     | 0,088 | 0,175 | 0      |
| U10     | 0,0873      | 0,306 | 0     | 0      | 0,5237     | 0,1309      | 0,1309  | 0     | 0,393 | 0      | 0,655 | 0,0873 | 0     | 0       | 0       | 0     | 0     | 0     | 0      |
| U11     | 0,0607      | 0,819 | 0,121 | 0      | 0,091      | 0,1214      | 0,0607  | 0,03  | 0     | 0,0303 | 0,455 | 0      | 0     | 0       | 0       | 0     | 0     | 0,273 | 0      |
| U12     | 0,2847      | 0,807 | 0,048 | 0,0949 | 0,2847     | 0,1898      | 0,0475  | 0,048 | 0,048 | 0,0475 | 0,285 | 0,0475 | 0     | 0       | 0       | 0,095 | 0,048 | 0,19  | 0      |
| U13     | 0,2921      | 0,643 | 0,058 | 0,1168 | 0,1168     | 0,2337      | 0       | 0,058 | 0,234 | 0,2337 | 0,175 | 0,0584 | 0,292 | 0,0584  | 0       | 0,351 | 0,234 | 0,058 | 0      |
| U14     | 0,0504      | 0,176 | 0,05  | 0      | 0,5293     | 0,2016      | 0,0252  | 0     | 0,151 | 0,0252 | 0,781 | 0,0504 | 0     | 0       | 0       | 0     | 0,076 | 0     | 0      |
| U15     | 0,3881      | 0,453 | 0,194 | 0      | 0,3234     | 0,1294      | 0,1941  | 0     | 0     | 0      | 0,582 | 0,0647 | 0     | 0,3234  | 0       | 0     | 0     | 0     | 0      |
| U16     | 0,4468      | 0,581 | 0,134 | 0      | 0,4021     | 0,134       | 0       | 0,134 | 0,313 | 0      | 0,313 | 0,2234 | 0     | 0       | 0       | 0     | 0     | 0,045 | 0      |
| U17     | 0           | 0,836 | 0     | 0      | 0,1857     | 0           | 0       | 0     | 0,464 | 0      | 0,186 | 0,0928 | 0     | 0       | 0       | 0     | 0     | 0     | 0,0928 |
| U18     | 0,4586      | 0,778 | 0,18  | 0      | 0,0199     | 0,1196      | 0,0997  | 0     | 0,04  | 0,0997 | 0,299 | 0      | 0     | 0       | 0       | 0     | 0,04  | 0,16  | 0      |
| U19     | 0,4049      | 0,585 | 0     | 0      | 0,225      | 0,225       | 0,045   | 0     | 0     | 0      | 0,18  | 0,045  | 0,045 | 0       | 0,045   | 0,09  | 0,585 | 0,045 | 0      |
| U20     | 0,1753      | 0,701 | 0     | 0,1168 | 0          | 0,0584      | 0       | 0     | 0,643 | 0      | 0,117 | 0      | 0,058 | 0       | 0       | 0     | 0,175 | 0     | 0      |
| U21     | 0,3989      | 0,779 | 0,019 | 0,057  | 0,266      | 0,266       | 0,152   | 0     | 0,038 | 0,152  | 0,076 | 0,019  | 0,057 | 0,133   | 0       | 0     | 0,114 | 0,038 | 0      |
| U22     | 0,0316      | 0,663 | 0     | 0      | 0,0631     | 0,1262      | 0       | 0     | 0,694 | 0,0316 | 0,095 | 0,1578 | 0     | 0,0316  | 0       | 0,032 | 0,032 | 0,126 | 0,0631 |
| U23     | 0,2771      | 0,831 | 0,079 | 0,0396 | 0,1584     | 0,0396      | 0       | 0,079 | 0,317 | 0,0792 | 0,04  | 0,2375 | 0     | 0       | 0       | 0,04  | 0     | 0,158 | 0      |
| U24     | 0,4339      | 0,651 | 0,109 | 0      | 0,1085     | 0           | 0       | 0,109 | 0,325 | 0,1085 | 0,217 | 0      | 0     | 0       | 0       | 0     | 0     | 0,434 | 0      |
| U25     | 0           | 0,546 | 0,046 | 0,0227 | 0,7047     | 0,3183      | 0,0227  | 0     | 0,182 | 0      | 0,182 | 0,0455 | 0,046 | 0       | 0,0227  | 0     | 0,068 | 0,159 | 0,0227 |
| U26     | 0,0658      | 0,593 | 0,066 | 0,1317 | 0,0658     | 0,0987      | 0       | 0,033 | 0,757 | 0      | 0,033 | 0,1646 | 0     | 0       | 0,0329  | 0,033 | 0     | 0     | 0,0658 |
| U27     | 0,3974      | 0,53  | 0     | 0      | 0          | 0           | 0       | 0,133 | 0,662 | 0      | 0,133 | 0      | 0     | 0,1325  | 0       | 0     | 0     | 0     | 0,2649 |
| U28     | 0           | 0,623 | 0,028 | 0,0283 | 0,2548     | 0,0283      | 0       | 0,057 | 0,708 | 0      | 0     | 0,1981 | 0     | 0       | 0       | 0     | 0,028 | 0     | 0,0283 |
| U29     | 0,5714      | 0,429 | 0,071 | 0,1429 | 0,4286     | 0,2857      | 0       | 0     | 0,214 | 0      | 0,143 | 0,2857 | 0     | 0       | 0,0714  | 0     | 0     | 0,214 | 0      |
| U30     | 0           | 0,473 | 0     | 0      | 0,189      | 0           | 0       | 0     | 0,85  | 0      | 0,095 | 0,0945 | 0     | 0       | 0       | 0     | 0     | 0     | 0      |
| U31     | 0,0283      | 0,623 | 0,057 | 0      | 0,283      | 0,0849      | 0       | 0     | 0,028 | 0      | 0,226 | 0,0283 | 0     | 0       | 0       | 0,085 | 0     | 0,679 | 0      |
| U32     | 0,6312      | 0,631 | 0,09  | 0      | 0,1803     | 0,2705      | 0,1803  | 0     | 0,09  | 0,0902 | 0     | 0      | 0     | 0       | 0       | 0     | 0,09  | 0,18  | 0      |
| U33     | 0,2598      | 0,455 | 0,065 | 0,4547 | 0,5846     | 0,1299      | 0       | 0,13  | 0,26  | 0,065  | 0,065 | 0,1949 | 0     | 0,065   | 0       | 0     | 0,065 | 0,13  | 0      |
| U34     | 0,559       | 0,447 | 0     | 0,1118 | 0,559      | 0,1118      | 0       | 0     | 0,112 | 0,1118 | 0,112 | 0      | 0     | 0       | 0       | 0     | 0     | 0,335 | 0      |
| U35     | 0,1443      | 0,433 | 0     | 0      | 0,7217     | 0,1443      | 0       | 0     | 0     | 0      | 0,144 | 0      | 0     | 0       | 0       | 0,144 | 0     | 0,433 | 0,1443 |
| U36     | 0,1779      | 0,153 | 0,051 | 0      | 0,8896     | 0,1271      | 0       | 0,025 | 0,102 | 0,0254 | 0,33  | 0,0508 | 0     | 0       | 0       | 0,025 | 0,025 | 0,102 | 0      |
| U37     | 0,0885      | 0,619 | 0,089 | 0      | 0,1474     | 0,059       | 0       | 0,059 | 0,324 | 0      | 0,03  | 0,1769 | 0     | 0       | 0       | 0     | 0     | 0,649 | 0,118  |
| U38     | 0,5604      | 0,641 | 0     | 0      | 0          | 0,0801      | 0       | 0     | 0,08  | 0,1601 | 0,08  | 0      | 0     | 0       | 0       | 0     | 0     | 0,48  | 0      |
| U39     | 0,1528      | 0,255 | 0,034 | 0,0509 | 0,8999     | 0,1358      | 0       | 0     | 0,017 | 0,017  | 0,17  | 0,034  | 0,017 | 0       | 0       | 0,017 | 0,204 | 0,085 | 0      |
| U40     | 0           | 0,335 | 0,112 | 0      | 0,1118     | 0,2236      | 0       | 0     | 0,894 | 0      | 0     | 0,1118 | 0     | 0       | 0       | 0     | 0     | 0     | 0      |
| U41     | 0           | 0,238 | 0     | 0,0297 | 0,0594     | 0,0594      | 0       | 0     | 0,951 | 0      | 0,03  | 0,0892 | 0     | 0,0594  | 0       | 0,03  | 0,119 | 0     | 0,0594 |
| U42     | 0           | 0,237 | 0,158 | 0,079  | 0,9084     | 0,079       | 0       | 0     | 0,04  | 0      | 0,158 | 0,1975 | 0     | 0,0395  | 0       | 0,079 | 0,079 | 0,04  | 0      |
| U43     | 0           | 0,083 | 0     | 0      | 0,3322     | 0           | 0       | 0     | 0,872 | 0      | 0,125 | 0,2491 | 0     | 0       | 0,0415  | 0     | 0,125 | 0     | 0,1661 |
| U44     | 0,0876      | 0,657 | 0,088 | 0,0438 | 0,2191     | 0,0876      | 0       | 0     | 0,044 | 0      | 0     | 0,0438 | 0     | 0       | 0       | 0     | 0     | 0,701 | 0      |
| U45     | 0,5733      | 0,573 | 0     | 0,0358 | 0,0717     | 0,0717      | 0       | 0     | 0     | 0,0358 | 0     | 0      | 0     | 0       | 0       | 0     | 0,036 | 0,573 | 0      |
| U46     | 0           | 0,296 | 0,059 | 0,2365 | 0          | 0           | 0       | 0,177 | 0,059 | 0      | 0,118 | 0,5913 | 0     | 0       | 0       | 0,532 | 0     | 0     | 0,4139 |
| U47     | 0,1514      | 0,189 | 0     | 0,0379 | 0,9463     | 0,0379      | 0       | 0,114 | 0,076 | 0      | 0,076 | 0,0757 | 0     | 0       | 0       | 0     | 0     | 0     | 0,1136 |
| U48     | 0,2914      | 0,194 | 0     | 0      | 0          | 0           | 0,1943  | 0     | 0     | 0      | 0     | 0      | 0,583 | 0,1943  | 0       | 0     | 0,68  | 0     | 0      |
| U49     | 0,5781      | 0,158 | 0     | 0      | 0,0526     | 0           | 0       | 0     | 0,053 | 0      | 0,053 | 0      | 0     | 0       | 0       | 0     | 0,105 | 0,788 | 0      |
| U50     | 0,1292      | 0,043 | 0     | 0,0215 | 0          | 0           | 0       | 0     | 0     | 0      | 0     | 0      | 0     | 0       | 0       | 0     | 0     | 0,991 | 0      |

## 4.2 Pruebas de casos de uso

En los siguientes apartados se realiza una prueba para cada uno de los cuatros casos de uso previamente definidos: recopilar información, extraer características, clasificar usuarios, clasificar usuarios por género.

### 4.2.1 Prueba para recopilar información

La recopilación de información del usuario se realiza en dos clases: la primera es la recopilación de la información implícita (ubicación geográfica) la cual se obtiene una vez que el usuario ingresa al sistema como se muestra en la figura 4.4, por otro lado la información explícita se recopila ingresando a la opción *Recopila*, la cual valida que sea la primera vez que el usuario realiza dicha acción, en caso contrario se lo indica por medio de un mensaje y se le invita a la etapa de extracción como se muestran en las figuras 4.5 y 4.6

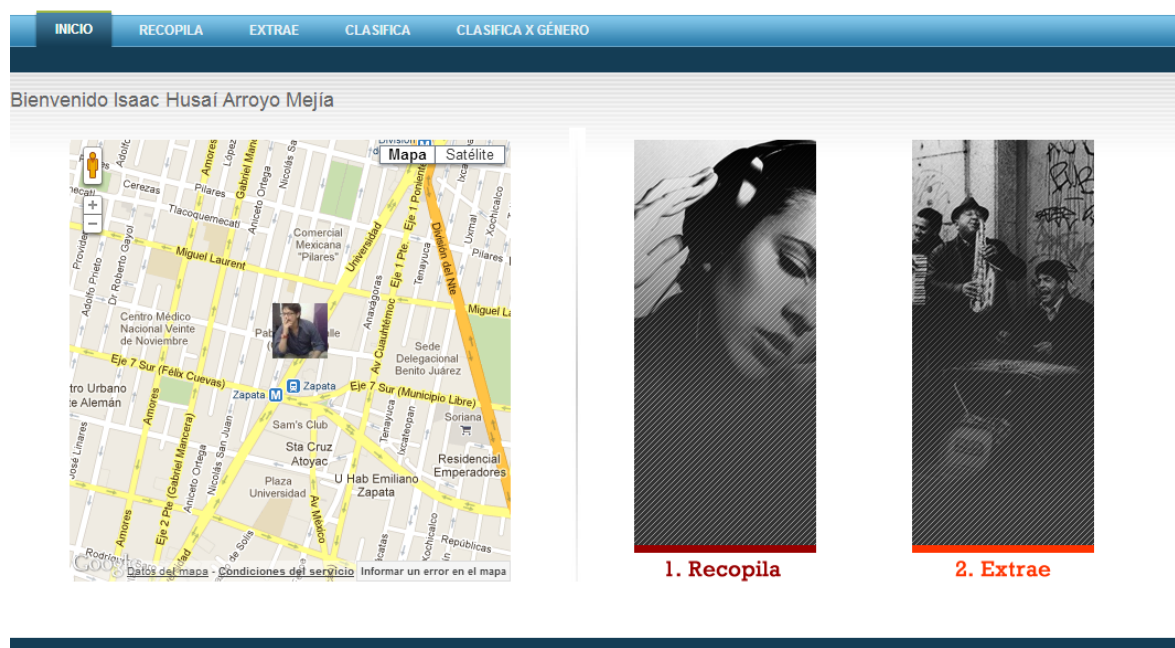


Figura 4.4 Recopilación implícita

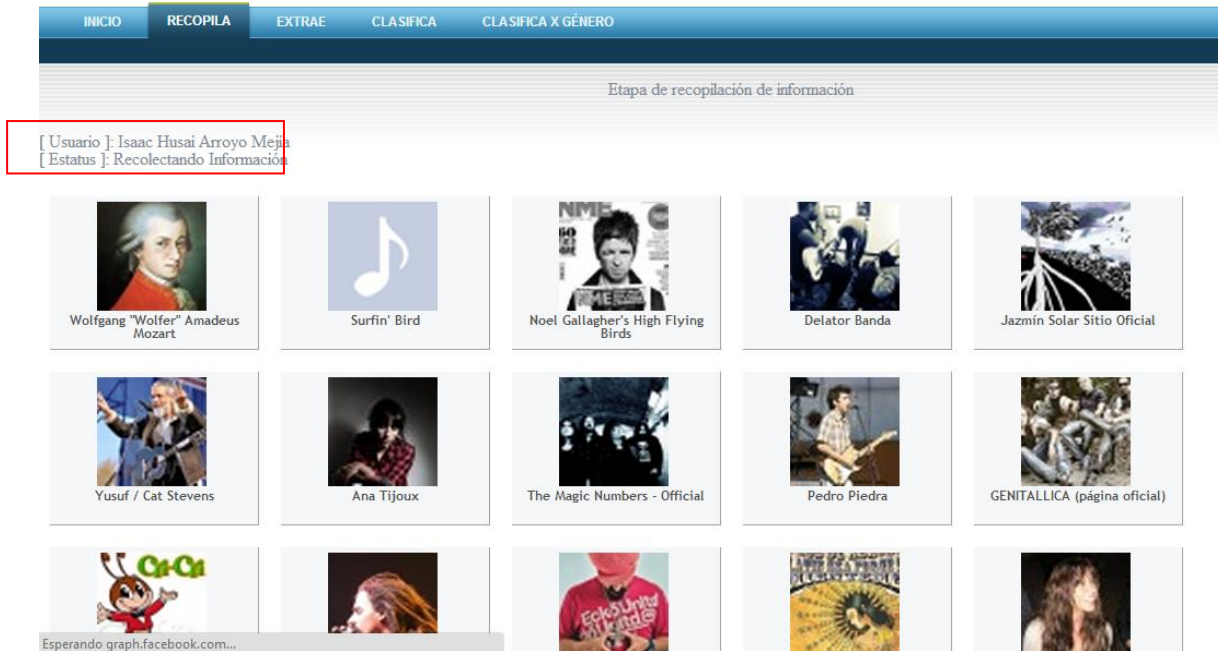


Figura 4.5 Recopilación explícita

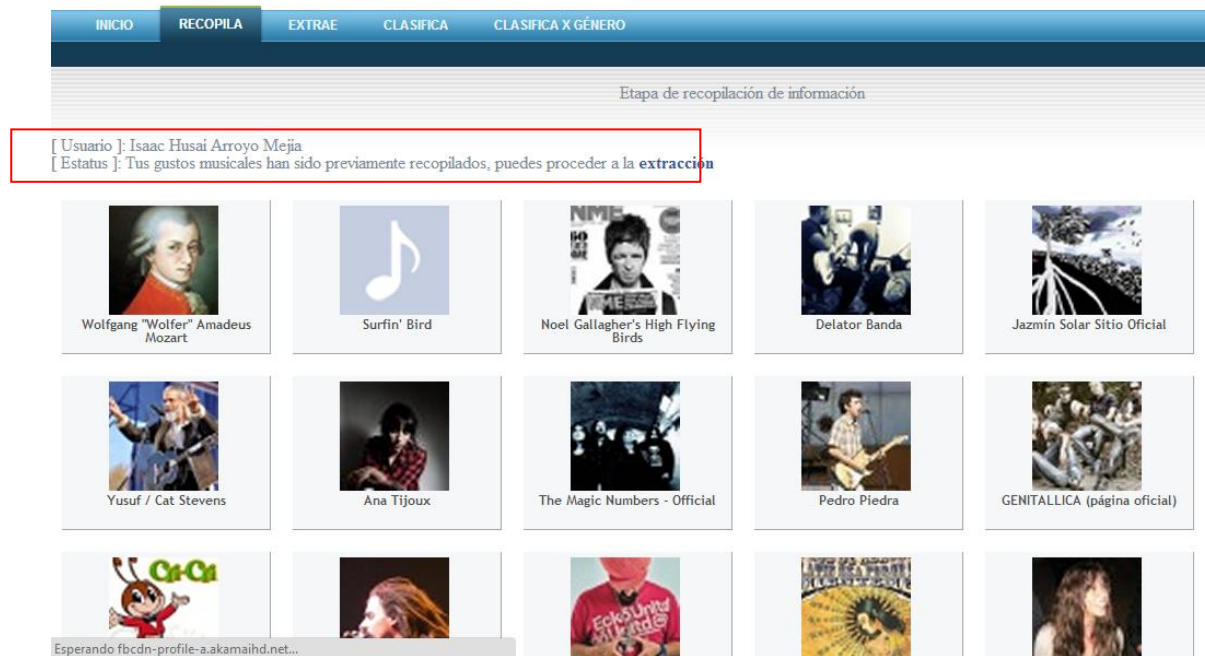


Figura 4.6 Recopilación explícita II

Existe un flujo alternativo el cual se desencadena cuando el usuario no cumple con el requisito de contar con al menos 19 instancias musicales, esto puede verse en la figura 4.7 por medio de otro usuario de prueba.

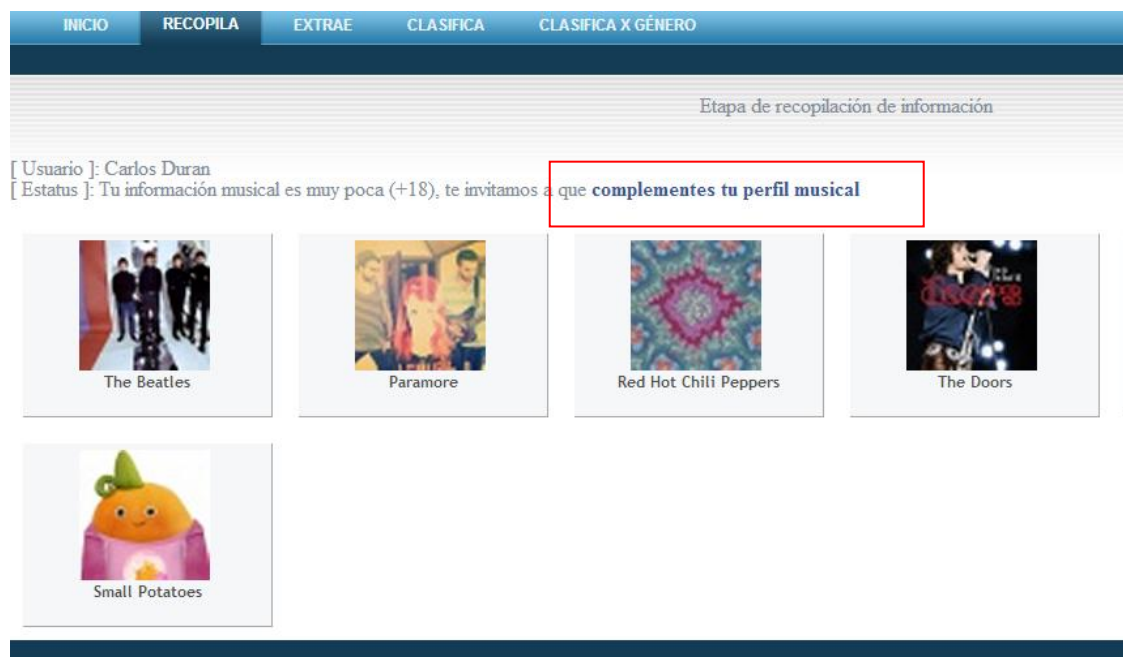


Figura 4.7 Recopilación explícita III

## 4.2.2 Prueba para extraer características

Al extraer características el sistema itera a través de las instancias musicales del usuario almacenadas en base de datos y busca que las etiquetas asociadas a cada uno de ellos correspondan a alguna posición de nuestro vector característico, el ejemplo de salida se despliega en la figura 4.8



Figura 4.8 Extracción de prueba

Al terminar el proceso se nos devuelve una tabla la cual contiene el vector característico del usuario y el porcentaje que tiene en cada género musical, dicho resultado se muestra en la figura 4.9

Clases ponderadas:

| Género       | Valor | Porcentaje |
|--------------|-------|------------|
| classic rock | 6     | 4.29 %     |
| rock         | 31    | 22.14 %    |
| punk         | 4     | 2.86 %     |
| reggae       | 2     | 1.43 %     |
| electronic   | 9     | 6.43 %     |
| alternative  | 8     | 5.71 %     |
| britpop      | 4     | 2.86 %     |
| rap          | 0     | 0 %        |
| pop          | 13    | 9.29 %     |
| grunge       | 2     | 1.43 %     |
| indie        | 32    | 22.86 %    |
| hip-hop      | 2     | 1.43 %     |
| blues        | 2     | 1.43 %     |
| classical    | 0     | 0 %        |
| country      | 0     | 0 %        |
| ska          | 9     | 6.43 %     |
| jazz         | 14    | 10 %       |
| metal        | 0     | 0 %        |
| banda        | 2     | 1.43 %     |

Figura 4.9 Porcentaje de géneros de usuario

Si la información del usuario ya ha sido extraída y se vuelve a seleccionar dicha acción el sistema nos indicará que previamente se han extraído las características del usuario y nos desplegará una tabla con los vectores característicos de todos los usuarios en el sistema, como se muestra en la figura 4.10

| Etapa de extracción de características                                                                              |             |      |      |        |            |             |         |     |     |        |       |        |       |          |         |     |      |       |       |
|---------------------------------------------------------------------------------------------------------------------|-------------|------|------|--------|------------|-------------|---------|-----|-----|--------|-------|--------|-------|----------|---------|-----|------|-------|-------|
| Tu información característica ya ha sido previamente extraída, a continuación se despliega la de todos los usuario. |             |      |      |        |            |             |         |     |     |        |       |        |       |          |         |     |      |       |       |
| Usuario                                                                                                             | ClassicRock | Rock | Punk | Reggae | Electronic | Alternative | BritPop | Rap | Pop | Grunge | Indie | HipHop | Blues | Classico | Country | Ska | Jazz | Metal | Banda |
| Liza Simpson                                                                                                        | 1           | 34   | 8    | 5      | 10         | 5           | 0       | 4   | 25  | 0      | 14    | 3      | 0     | 1        | 0       | 12  | 0    | 1     | 4     |
| Jorge Israel Toledo Alvarado                                                                                        | 4           | 6    | 1    | 0      | 1          | 0           | 0       | 1   | 3   | 1      | 2     | 0      | 0     | 0        | 0       | 0   | 0    | 4     | 0     |
| Miri Balbuena                                                                                                       | 1           | 8    | 2    | 1      | 2          | 2           | 0       | 0   | 6   | 1      | 3     | 1      | 0     | 0        | 0       | 0   | 1    | 2     | 0     |
| Antonio De Jesús Méndez                                                                                             | 2           | 27   | 4    | 0      | 3          | 4           | 2       | 1   | 0   | 1      | 15    | 0      | 0     | 0        | 0       | 0   | 0    | 9     | 0     |
| Eugenio Vargas Becerra                                                                                              | 21          | 41   | 1    | 3      | 14         | 14          | 8       | 0   | 2   | 8      | 4     | 1      | 3     | 7        | 0       | 0   | 6    | 2     | 0     |
| Reex Gallardo                                                                                                       | 14          | 90   | 11   | 6      | 40         | 18          | 4       | 1   | 38  | 2      | 59    | 14     | 0     | 0        | 2       | 3   | 0    | 3     | 0     |
| Mar Leiden                                                                                                          | 6           | 2    | 0    | 1      | 0          | 0           | 0       | 0   | 0   | 0      | 0     | 0      | 0     | 0        | 0       | 0   | 0    | 46    | 0     |
| Ava Dew                                                                                                             | 9           | 15   | 2    | 3      | 53         | 8           | 0       | 0   | 1   | 1      | 10    | 2      | 1     | 0        | 0       | 1   | 12   | 5     | 0     |
| Deivid Estévez Campos                                                                                               | 9           | 13   | 0    | 0      | 5          | 5           | 1       | 0   | 0   | 0      | 4     | 1      | 1     | 0        | 1       | 2   | 13   | 1     | 0     |
| Fernando Rocha                                                                                                      | 3           | 4    | 0    | 0      | 2          | 0           | 0       | 0   | 1   | 0      | 3     | 0      | 0     | 0        | 0       | 0   | 2    | 2     | 0     |
| Raúl A. García Cárdenas                                                                                             | 0           | 8    | 0    | 1      | 2          | 2           | 0       | 0   | 32  | 0      | 1     | 3      | 0     | 2        | 0       | 1   | 4    | 0     | 2     |
| Daniela Medina                                                                                                      | 2           | 7    | 0    | 0      | 12         | 3           | 3       | 0   | 9   | 0      | 15    | 2      | 0     | 0        | 0       | 0   | 0    | 0     | 0     |
| Sergio Marquez                                                                                                      | 0           | 3    | 1    | 0      | 1          | 2           | 0       | 0   | 8   | 0      | 1     | 0      | 0     | 0        | 0       | 0   | 0    | 0     | 0     |
| Marco Moreno Ibarra                                                                                                 | 11          | 3    | 0    | 0      | 1          | 0           | 0       | 0   | 1   | 0      | 1     | 0      | 0     | 0        | 0       | 0   | 2    | 15    | 0     |
| Daniel Correa                                                                                                       | 6           | 7    | 3    | 0      | 5          | 2           | 3       | 0   | 0   | 0      | 9     | 1      | 0     | 5        | 0       | 0   | 0    | 0     | 0     |
| Luis Rivera                                                                                                         | 7           | 8    | 0    | 0      | 0          | 1           | 0       | 0   | 1   | 2      | 1     | 0      | 0     | 0        | 0       | 0   | 0    | 6     | 0     |
| Israel Munillo                                                                                                      | 16          | 16   | 0    | 1      | 2          | 2           | 0       | 0   | 0   | 1      | 0     | 0      | 0     | 0        | 0       | 0   | 1    | 16    | 0     |
| Raul Antonio Villamil Rodriguez                                                                                     | 6           | 17   | 1    | 2      | 6          | 4           | 1       | 1   | 1   | 1      | 6     | 1      | 0     | 0        | 0       | 2   | 1    | 4     | 0     |
| Teran Corona Liliana                                                                                                | 3           | 12   | 0    | 2      | 0          | 1           | 0       | 0   | 11  | 0      | 2     | 0      | 1     | 0        | 0       | 0   | 3    | 0     | 0     |
| Roberto Castañeda                                                                                                   | 0           | 3    | 1    | 0      | 1          | 2           | 0       | 0   | 0   | 0      | 5     | 0      | 0     | 0        | 0       | 3   | 1    | 0     | 0     |
| Juan Castellán                                                                                                      | 1           | 3    | 0    | 0      | 5          | 1           | 0       | 0   | 0   | 0      | 1     | 0      | 0     | 0        | 0       | 1   | 0    | 3     | 1     |
| Feer Luckeer                                                                                                        | 1           | 22   | 2    | 0      | 10         | 3           | 0       | 0   | 1   | 0      | 8     | 1      | 0     | 0        | 0       | 3   | 0    | 24    | 0     |
| Coneja Perez                                                                                                        | 5           | 4    | 0    | 1      | 5          | 1           | 0       | 0   | 1   | 1      | 1     | 0      | 0     | 0        | 0       | 0   | 0    | 3     | 0     |
| Dara Rivera                                                                                                         | 23          | 39   | 9    | 0      | 1          | 6           | 5       | 0   | 2   | 5      | 15    | 0      | 0     | 0        | 0       | 0   | 2    | 8     | 0     |
| Mariana Wolf                                                                                                        | 2           | 18   | 2    | 4      | 2          | 3           | 0       | 1   | 23  | 0      | 1     | 5      | 0     | 0        | 1       | 1   | 0    | 0     | 2     |
| Miriam Haller                                                                                                       | 0           | 24   | 2    | 1      | 31         | 14          | 1       | 0   | 8   | 0      | 8     | 2      | 2     | 0        | 1       | 0   | 3    | 7     | 1     |
| Leonardo Garcia                                                                                                     | 3           | 2    | 0    | 0      | 0          | 0           | 2       | 0   | 0   | 0      | 0     | 0      | 6     | 2        | 0       | 0   | 7    | 0     | 0     |

Figura 4.10 Vectores característicos de usuario

### 4.2.3 Prueba para clasificar usuarios

Al realizar la prueba de clasificación nos arroja en orden descendente a los usuarios según su similitud musical, a su vez nos indica el número de grupo al que ha sido asignado por el algoritmo K-mean, por otro lado los usuarios son representados en un mapa según su última ubicación geográfica conocida, lo anterior se muestra en la figura 4.11

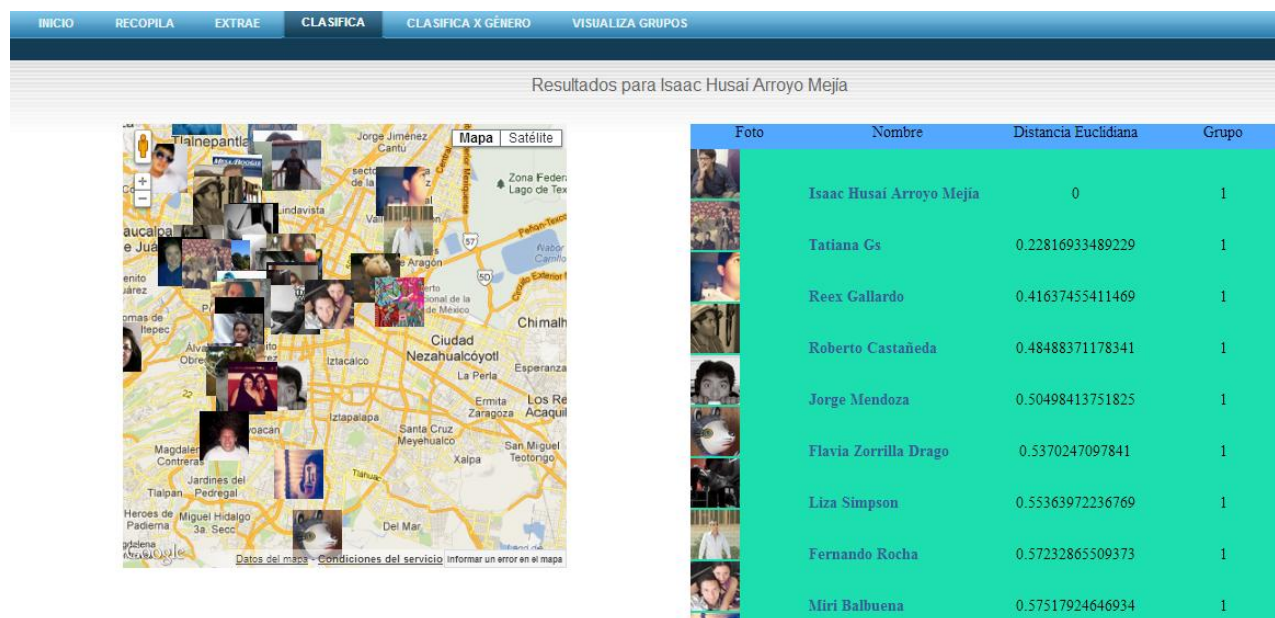


Figura 4.11 Prueba de clasificación

Para realizar el análisis de la salida en cuanto a la distancia euclidiana procederemos a interpretar exclusivamente los primeros tres usuarios más similares, cuyos alias son:  $U_2$ ,  $U_3$ ,  $U_4$  y los tres menos similares, cuyos alias son:  $U_{48}$ ,  $U_{49}$  y  $U_{50}$ .

En la figura 4.12 se muestra la comparativa de la gráfica de los vectores con menor distancia Euclidiana.

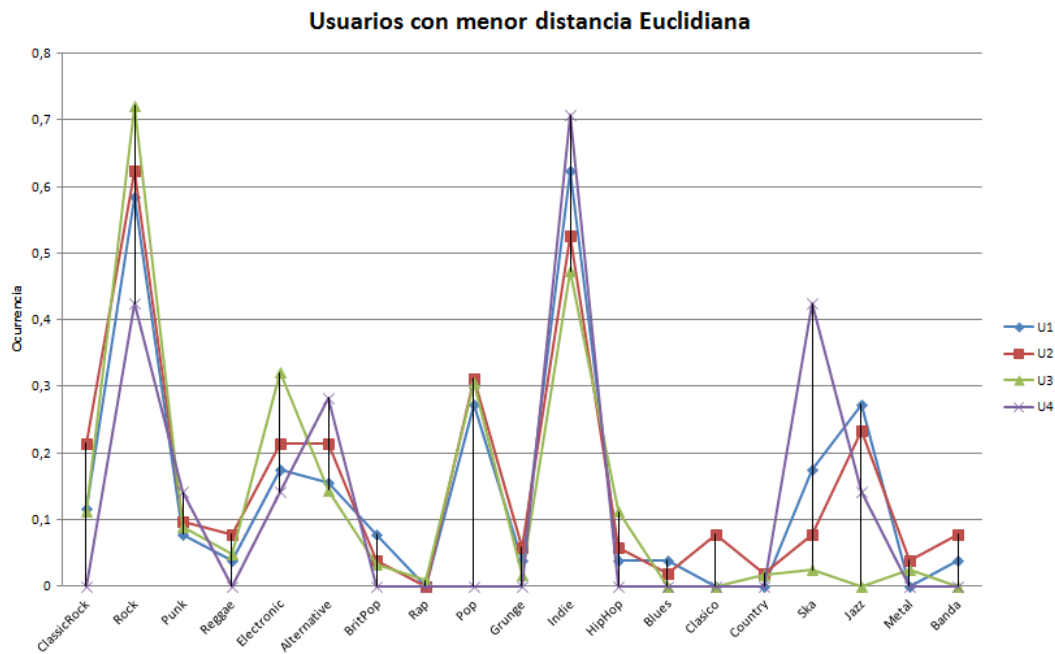


Figura 4.12 Usuarios con menor distancia Euclidiana

En la figura 4.13 se muestra la comparativa de la gráfica de los vectores con menor distancia Euclidiana.

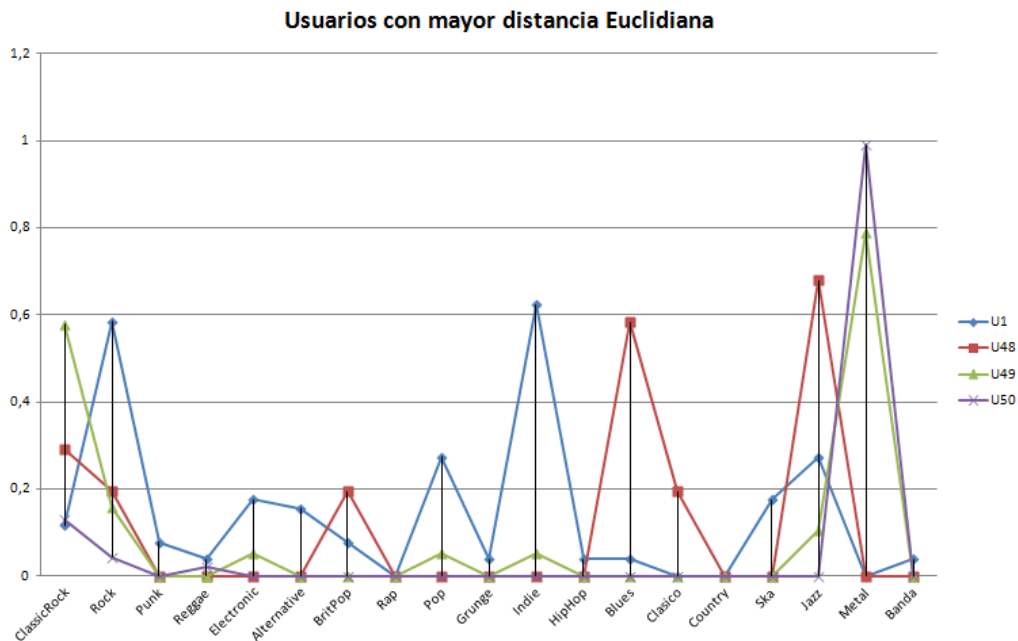


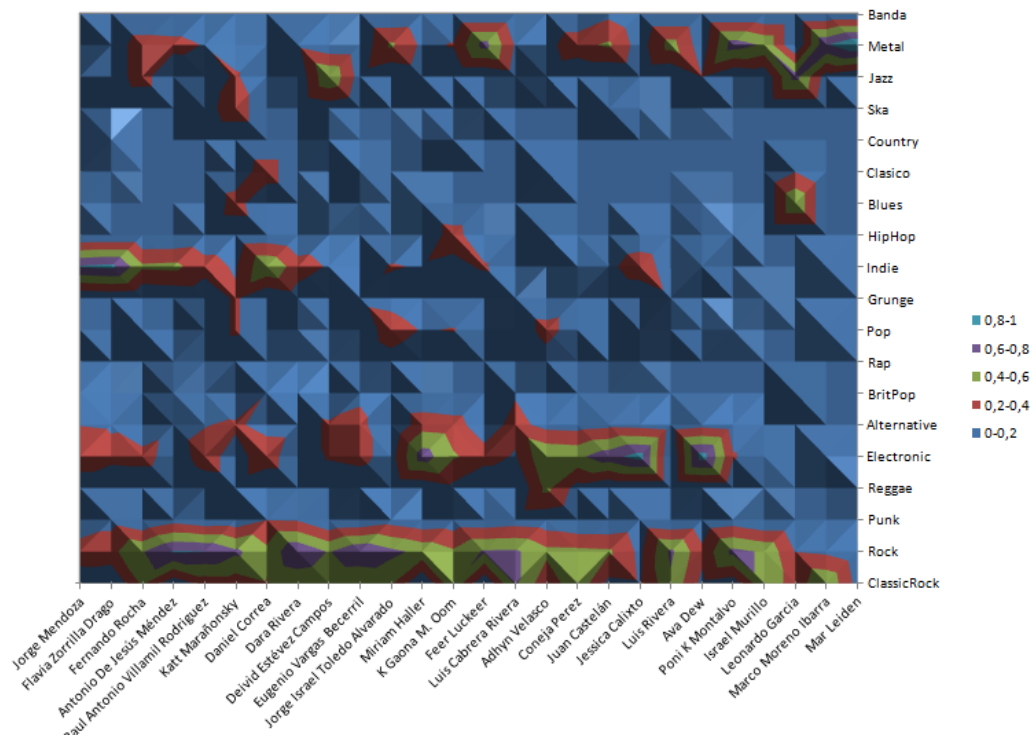
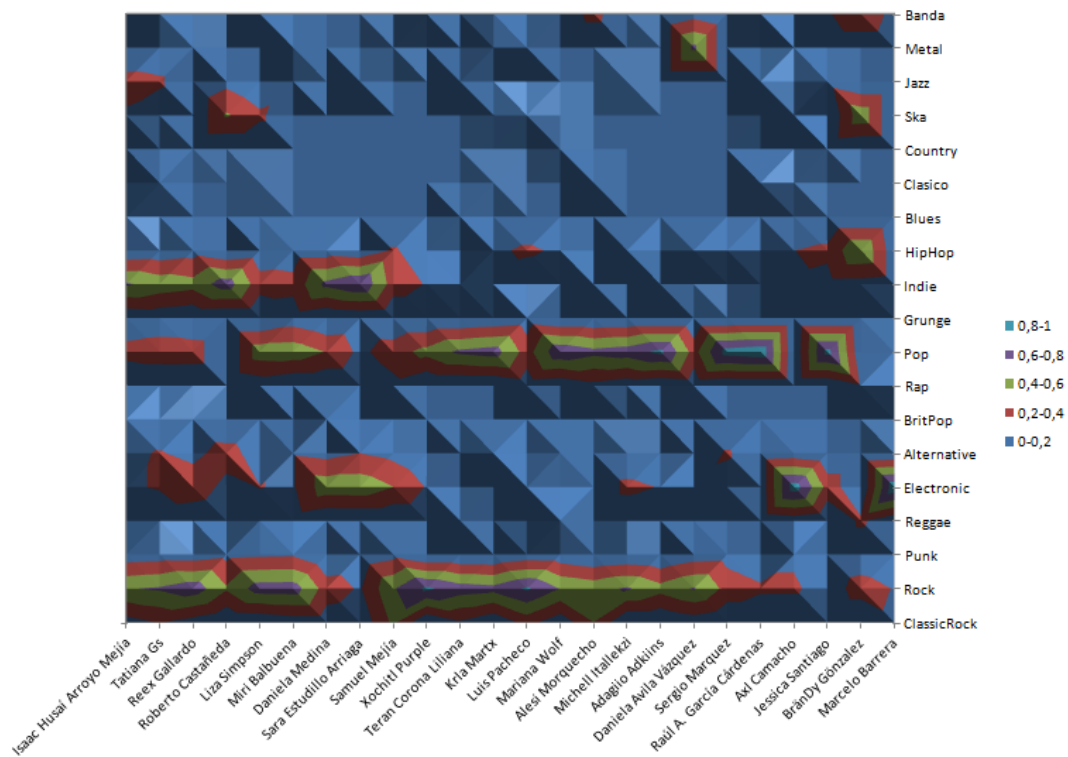
Figura 4.13 Usuarios con mayor distancia Euclidiana

Como se puede apreciar en las dos gráficas anteriores los usuarios con una distancia euclidiana menor mantienen curvas similares debido a que su preferencia musical indicada por sus vectores característicos es similar, mientras que los usuarios con una distancia euclidiana mayor presentan curvas muy diferentes.

Por otro lado se puede observar que mediante el algoritmo K-mean se particionan los usuarios en dos grupos como se observa en la figura 4.14 para los cuales hemos graficado las superficies que son generadas por los vectores de los usuarios que los componen y se observan los resultados en las figura 4.15 y 4.16.

| Nombre                          | Distancia Euclidiana | Grupo |
|---------------------------------|----------------------|-------|
| Isaac Husai Arroyo Mejía        | 0                    | 1     |
| Tatiana Gs                      | 0.22816933489229     | 1     |
| Reex Gallardo                   | 0.41637455411469     | 1     |
| Roberto Castañeda               | 0.48488371178341     | 1     |
| Jorge Mendoza                   | 0.50498413751825     | 0     |
| Flavia Zorrilla Drago           | 0.5370247097841      | 0     |
| Liza Simpson                    | 0.55363972236769     | 1     |
| Fernando Rocha                  | 0.57232865509373     | 0     |
| Miri Balbuena                   | 0.57517924646934     | 1     |
| Daniela Medina                  | 0.58206408931066     | 1     |
| Antonio De Jesús Méndez         | 0.59883459272679     | 0     |
| Raul Antonio Villamil Rodriguez | 0.6001557087024      | 0     |
| Katt Marañosky                  | 0.6365240149553      | 0     |
| Sara Estudillo Arriaga          | 0.64529118385935     | 1     |
| Daniel Correa                   | 0.65836779970728     | 0     |
| Samuel Mejía                    | 0.65850973622174     | 1     |
| Xochitl Purple                  | 0.67636763137515     | 1     |
| Dara Rivera                     | 0.68583244808983     | 0     |
| Deivid Estévez Campos           | 0.69443086664365     | 0     |
| Teran Corona Liliana            | 0.71730264552385     | 1     |
| Eugenio Vargas Becerril         | 0.76965798787796     | 0     |
| Krla Martx                      | 0.78130270146136     | 1     |
| Luis Pacheco                    | 0.78621723758618     | 1     |
| Jorge Israel Toledo Alvarado    | 0.79072292699601     | 0     |
| Miriam Haller                   | 0.79388522395382     | 0     |
| Mariana Wolf                    | 0.85616814755828     | 1     |
| Alesi Morquecho                 | 0.86000653757558     | 1     |
| Michell Itallekzi               | 0.86430894329567     | 1     |
| K Gaona M. Oom                  | 0.88626378062704     | 0     |
| Adaggio Adkiins                 | 0.88955290164019     | 1     |
| Feer Luckeer                    | 0.89470635637701     | 0     |
| Luis Cabrera Rivera             | 0.90470220446772     | 0     |

Figura 4.14 Ejemplo de clasificación con grupos



Como se observa en las dos superficies anteriores, los grupos generados son consistentes ya que a pesar de que ambos comparten un alto interés por géneros como el rock o el classic rock presentan diferencias en otros, por ejemplo los miembros del grupo 1 mantienen un interés alto en géneros tales como el Hip-Hop, Indie, Grunge y Pop mientras que los usuarios del grupo 2 se encuentran más interesados en géneros como el Metal, Banda y Jazz.

Por otro lado también se realiza un análisis respecto a las ubicaciones geográficas de los usuarios y sus preferencias, cuyos resultados se pueden visualizar en la superficie que se encuentra en la figura 4.17

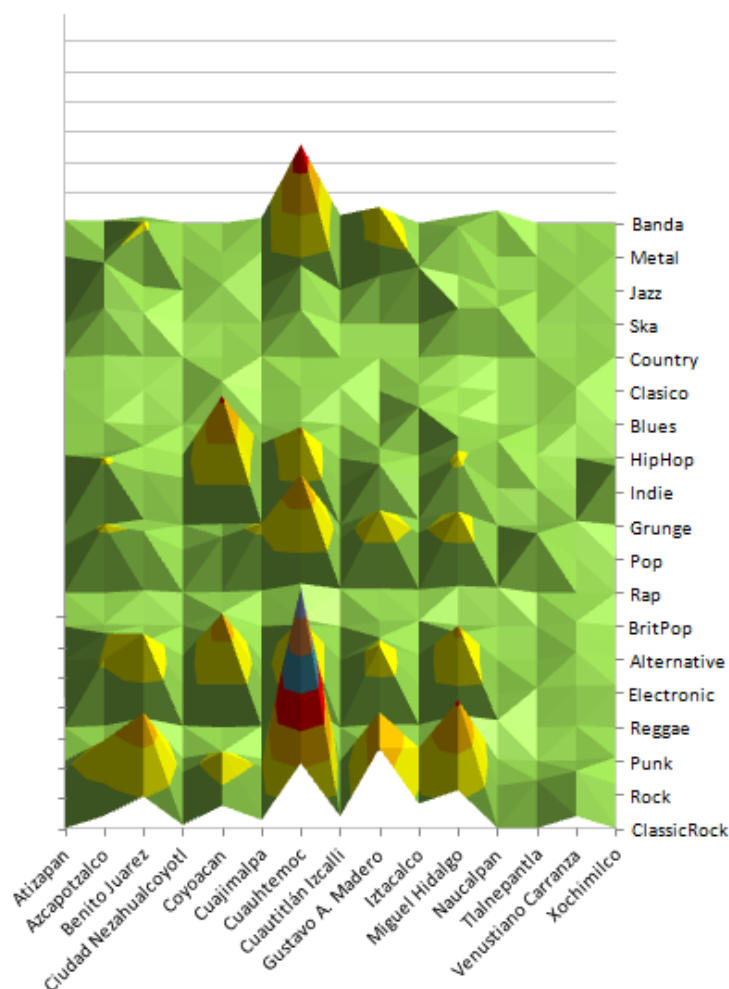


Figura 4.17 Superficie de preferencias por municipio

La aplicación también cuenta con un mapa que nos permite visualizar cómo se encuentran distribuidos los miembros de cada uno de los grupos como se muestra en la figura 4.18

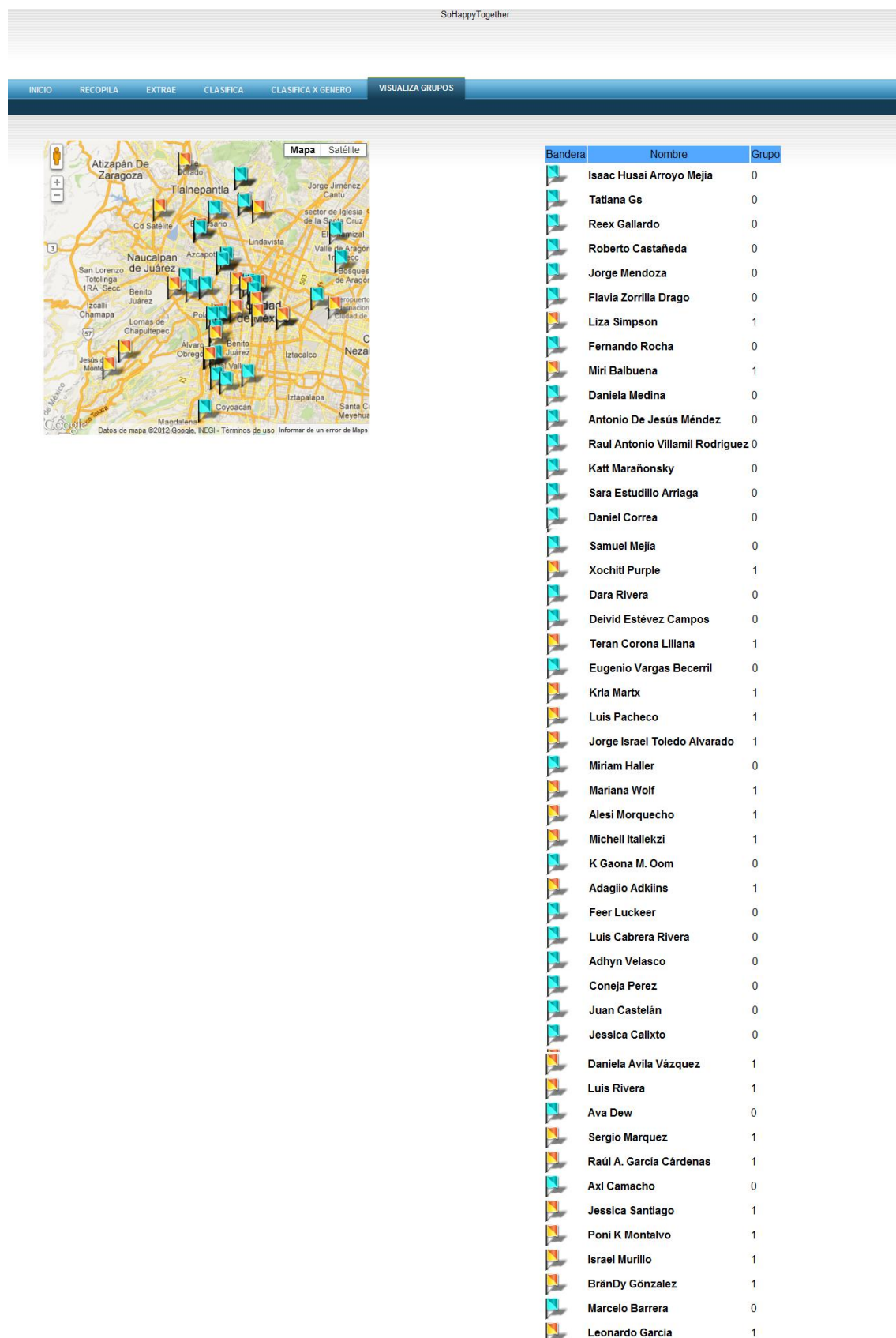


Figura 4.18 Miembros de grupos visualizados en mapa

#### 4.2.4 Prueba para clasificar usuarios por género

La opción de clasificar por género surge de la necesidad de eliminar la influencia que tienen las posiciones del vector característico que no sean de nuestro interés, por ejemplo para la prueba realizada en el punto anterior se observan dos casos interesantes, en primera instancia si se analiza de manera general la similitud entre el usuario  $U_1$  y el  $U_4$  es alta, pero al observar su gráfica en la figura 4.12 se puede observar que la clase Jazz no es tan similar por lo cual si clasificamos con ese filtro la similitud con el usuario ya no será alta y como se puede visualizar en la figura 4.19 el usuario  $U_4$  quedaría en los últimos niveles de similitud.

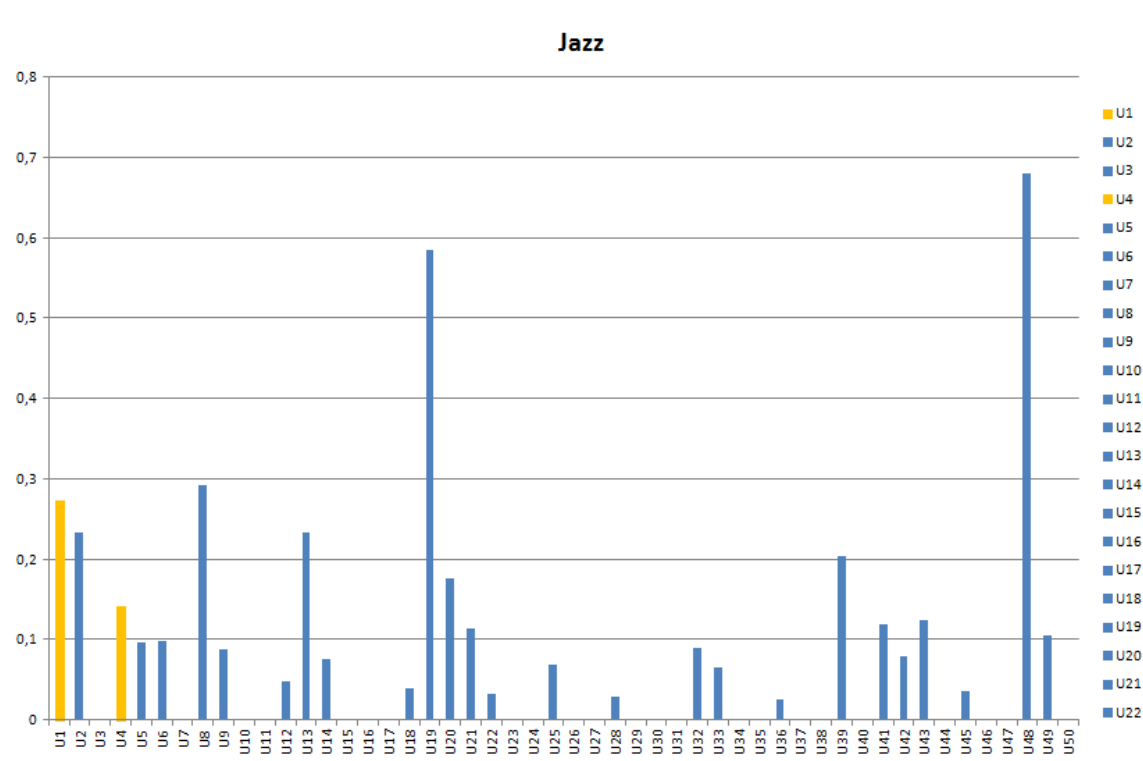


Figura 4.19 Gráfica jazz para comparar U1 y U4

Finalmente se realiza una comparativa para el orden de los usuarios según la distancia euclidiana seleccionando la métrica de la reciprocidad; es decir el número de conexiones que comparten en Facebook respecto a la música. Con lo anterior se generó una gráfica donde se compara el orden en que se ordenan a los usuarios por un lado tomando en cuenta el número de conexiones musicales que comparten y por otro lado la distancia de su vector característico que representa la preferencia musical de los usuarios, los resultados pueden verse en la figura 4.20

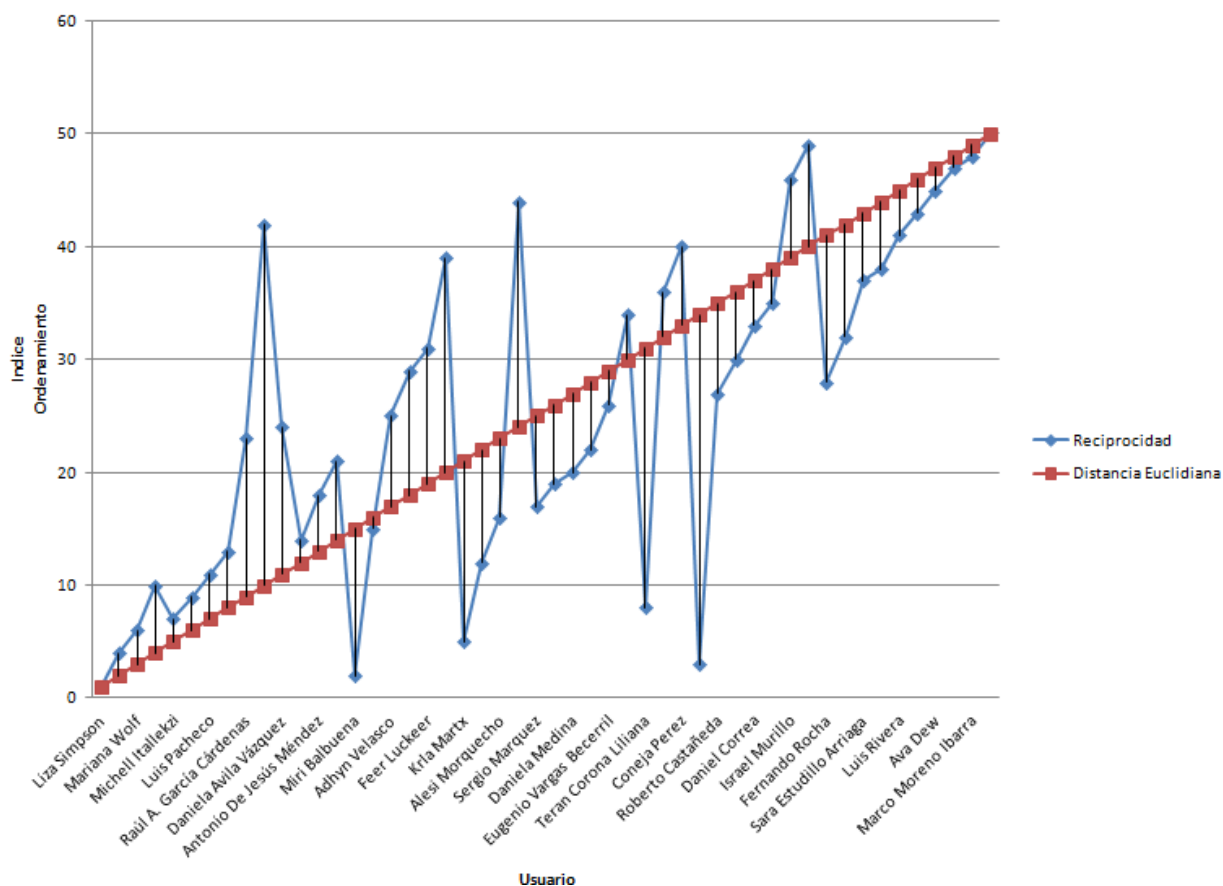


Figura 4.20 Comparativa de salida reciprocidad contra distancia euclidiana

Con la gráfica anterior se puede concluir que los resultados son consistentes respecto a la métrica de la reciprocidad, las diferencias que existen se debe que a pesar de que en ocasiones la reciprocidad no sea alta puede influir que los géneros de los artistas sean similares entre dos usuarios, en el caso contrario una reciprocidad alta no garantiza que no existan otros usuarios cuyas preferencias generales sean más similares ya que se depende del vector característico completo.



# CAPÍTULO V

## CONCLUSIONES

***“No siempre se consigue lo que se quiere,  
pero si se intenta,  
se consigue lo que se necesita.”***

Mick Jagger

## Capítulo 5 Conclusiones

### 5.1 Introducción

El presente trabajo ha mostrado un nuevo enfoque para construir grupos de usuarios en redes sociales, enfocándose en los intereses expuestos por los mismos más que en las conexiones existentes entre ellos.

Mediante la metodología expuesta se podría analizar la similitud entre usuarios de una red social tomando en cuenta cualquiera de sus intereses siempre y cuando tengamos una fuente de información especializada en el interés seleccionado.

En la actualidad la creación y uso de redes sociales especializadas va creciendo debido a que las personas buscan encontrar a otros individuos con los cuales compartan sus gustos y no solamente conocidos en común, sin embargo no todas las personas están dispuestas a hacer uso de varias redes sociales al mismo tiempo por lo desgastante que puede resultar mantener activo más de un perfil de usuario a la vez.

Por otro lado el gran número de usuarios que utilizan Facebook van construyendo un perfil rico en instancias de sus preferencias, sin embargo estas instancias por si solas únicamente sirven para comparar la similitud entre dos usuarios por medio de la multiplicidad, es decir el número de instancias que tienen en común, lo anterior sumado al hecho de que la multiplicidad puede ser calculada entre dos instancias exclusivamente idénticas ya que en Facebook dos instancias son diferentes sin importar lo similar que resulte el nombre entre ellas, vuelve necesario crear un mecanismo que sirva como puente entre toda esta información y un significado para poder explotarla.

La metodología propuesta resulta ser un enfoque novedoso para poder crear grupos de usuarios que estén fuertemente relacionados, esta metodología también resulta flexible ya que permite implementar libremente cada uno de los módulos.

### 5.2 Trabajos futuros

El presente proyecto genera la oportunidad de diversos trabajos a futuro principalmente en el área de la mercadotecnia ya que podemos encontrar tipos de usuarios que sean de nuestro interés, esto puede servir para ofrecer servicios o productos a un mercado meta, algunos ejemplos de trabajos se listan a continuación:

1. Sistema para notificación de conciertos de tu interés cerca de tu ubicación geográfica.
2. Probar otras técnicas de clasificación para la metodología.
3. Implementar una taxonomía de la música para realizar la clasificación.
4. Sistema para recomendación de artistas mediante tiendas de música en línea.
5. Análisis de la relación entre los intereses musicales y la ubicación geográfica.
6. Sistema para detección de mercados meta.
7. Sistema para publicidad dirigida.
8. Clasificación de usuarios mediante redes semánticas.
9. Implementación de metodología en otros tipos de interés por ejemplo el cine.

El presente trabajo tiene como motivación personal aportar un poco para disminuir la brecha que separa a las personas a veces paradójicamente más distantes mientras más se conectan en las redes, tomando como pretexto y caso de estudio la música ya que como diría Robert Browning: “el que escucha música siente que su soledad, de repente, se puebla”.



## REFERENCIAS

- [1] Javier Sorella. (2008). Árbol genealógico de la música contemporánea. Recuperado el 10 de Junio del 2011, del sitio web: <<http://i49.tinypic.com/2nm29tc.gif>>
- [2] Lada A. Adamic, Orkut Buyukkokten, Eytan Adar. (2003). A social network caught in the Web. *First Monday*, 8(6).
- [3] Barry Wellman, Jeffrey Boase, Wenhong Chen (2002). The networked nature of community: Online and Offline. *IT & Society*, 151-165.
- [4] Xiang Xiao, Yu Zheng, Qiong Luo, Xing Xie. (2009). Finding similar users using category-based location history. Microsoft Research Asia.
- [5] Aristóteles. (1920). Libro primero. *La Política versión castellana de Nicolás Estévez* (5-6). Paris: Editorial Garnier Hermanos.
- [6] Goodchild, M.F. (2007): "Citizens as sensors: the world of volunteered geography". *GeoJournal* 69, 4, pp. 211–221.
- [7] Valdis Krebs. (2000). "The Social Life of Routers." *Internet Protocol Journal*, 14-25.
- [8] Philip Kotler. (2008). *Fundamentos de Marketing* (235-255). México: Pearson.
- [9] Enkh-Amgalan Baatarjav, Santi Phitakkitnukoon, Ram Dantu. (2008). Group Recommendation System for Facebook. Department of Computer Science and Engineering. Universidad del Norte de Texas.
- [10] Yi-Kuang Ko, Jing-Kai Lou, Cheng-Te Li, Shou-De Lin, Shyh-Kang Jeng. (2010). A social network evolution model based on seniority. Springer.
- [11] Barbara Caci, Maurizio Cardaci, Marco E. Tabacchi. (2010). Facebook as a Small World: a topological hypothesis. Springer.
- [12] Eszter Hargittai. (2007). Whose space? Differences among users and nonusers of social network sites. *Journal of Computer-Mediated Communication* 13(1):artículo 14
- [13] Donath J, Boyd D. (2004). Public displays of connection. *BT Technology Journal* 2(4):71–82

- [14] Eran Hammer. (2012). Hueniver. Recuperado el 02 de Febrero de 2011, del sitio web de Hueniverse: < <http://hueniverse.com/oauth/>>
- [15] Facebook. (2012). Graph API. Recuperado el 10 de Febrero de 2012, del sitio web de Desarrolladores Facebook: <<https://developers.facebook.com/docs/reference/api>>
- [16] Brad McCarty. (2011). LastFM 50 billion scrobbles and the return of the mix tape. Recuperado el 16 de Abril del 2011, del sitio web de The Next Web: <<http://thenextweb.com/media/2011/04/16/last-fm-50-billion-scrobbles-and-the-return-of-the-mix-tape/3>>
- [17] LastFM. (2012). LastFM Web Services. Recuperado el 20 de Diciembre de 2011, del sitio web de LastFM: <<http://www.lastfm.es/api/intro>>
- [18] Matt Oakes. (2012). PHP LastFM API. Recuperado el 2 de Enero del 2012, del sitio web de Matt Oakes: <<http://www.matto1990.com/projects>>
- [19] Faust K. Wasserman. (1994). Social network analysis. Methods and applications. Cambridge University Press, Cambridge
- [20] John Scott. (2011). Social network analysis: developments, advances, and prospects. Social Network Analysis and Mining 1(1):21–26
- [21] Miller McPherson, Lynn Smith-Lovin, James Cook. (2001). Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 27:415-444
- [22] Joel Podolny, James Baron. (1997). Resources and relationships: Social networks and mobility in the workplace. American Sociological Review, 62(5):673-693
- [23] Charles Kadushin. (2012). Understanding social networks: Theories, concepts, and findings. Oxford: Oxford University Press
- [24] Victor Robles. (2003). Clasificación supervisada basada en redes bayesianas. Aplicación en Biología computacional. Tesis de doctorado. Facultad de Informática. Universidad Politécnica de Madrid.
- [25] Ron Kohavi. (1995). Wrappers for Performance Enhancement and Oblivious Decision Graphs. Tesis de doctorado, Computer Science Department, Universidad de Stanford.

- [26] Thomas G. Dietterich. (1986). Learning at the knowledge level. Machine Learning, 1(3):287-315.
- [27]. Abdelmalik Moujahid, Iñaki Inza, Pedro Larrañaga. Clasificadores K-NN [en línea]. Euskadi: Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2010 [fecha de consulta: 10 de Febrero del 2012]. Disponible en: <<http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t9knn.pdf>>.
- [28] Queen, J.B. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1:281-297.
- [29] Xian-Ming Xu, Justin Zhan, Hai-Tao Zhu. (2008). Using Social Networks to Organize Researcher Community. Springer-Verlag.
- [30] Ivar Jacobson. Object Oriented Software Engineering. A Use Case Driven Approach. Addison Wesley, 1992.