Dataset de contenidos musicales de video, basado en emociones

Luis Alejandro Solarte Moncayo
Estudiante Ingeniería Electrónica y telecomunicaciones

Mauricio Sánchez Barragán
Estudiante Ingeniería Electrónica y telecomunicaciones

Gabriel Elías Chanchí Golondrino
Estudiante Ingeniería Electrónica y telecomunicaciones

Diego Fabián Duran Dorado
Ingeniero en Electrónica y Telecomunicaciones
Magister en Ingeniería Telemática
Candidato a Doctor en Ingeniería Telemática

José Luis Arciniegas Herrera
Ingeniero en Electrónica y Telecomunicaciones
Especialista en Redes y Servicios Telemáticos
Doctor Ingeniero de Telecomunicación

Universidad del Cauca
{lasolartem, mauriciosanchez, gabrielc, dduran, jlarci}@unicauca.edu.co

(Tipo de Artículo: Investigación Científica y Tecnológica. Recibido el 11/06/2016. Aprobado el 22/06/2016)

Resumen. Agilizar el acceso al contenido, disminuyendo los tiempos de navegación por los catálogos multimedia, es uno de los retos del servicio de video bajo demanda (VoD), el cual es consecuencia del incremento de la cantidad de contenidos en las redes actuales. En este artículo, se describe el proceso de conformación de un dataset de videos musicales. Este dataset fue usado para el diseño e implementación de un servicio de VoD, el cual busca mejorar el acceso al contenido, mediante la clasificación musical de emociones. Así, en este trabajo se presenta la adaptación de un modelo de clasificación de emociones a partir del modelo de arousal-valence. Además, se describe el desarrollo de una herramienta Java para la clasificación de contenidos, la cual fue usada en la conformación del dataset. Finalmente, con el propósito de evaluar el dataset construido, se muestra la estructura funcional del servicio de VoD desarrollado.

Palabras clave. Arousal, clasificación de contenidos, dataset, multimedia, servicio de VoD, valence.

Dataset of music video content, based on emotions

Abstract. *Agile access to content by reducing navigation time by the multimedia content catalogs is one of the challenges of the video on demand service (VoD), which it is the result of the increased amount of content in current networks. In this article, we describe the conformation process of a musical video content dataset. Such dataset was used for the design and implementation of a VoD service, which aims to improve access to content through the musical classification of emotions. So, in this paper, we present the adaptation of a classification model of emotions taking into account the arousal-valence model. Furthermore, we describe the development of a Java tool for classifying content, which was used in the conformation of the dataset. Finally, with the purpose of evaluating the dataset built, we show the functional structure of the developed VoD service.*

Keywords. *Arousal, contents classification, dataset, media, VoD service, valence.*

1. Introducción

En los últimos años, el servicio de video bajo demanda (VoD) ha sido ampliamente difundido en internet [1]. Este servicio es definido en [2] como la entrega de contenido de vídeo a través del protocolo de internet (IP) de banda ancha al espectador, lo que posibilita al usuario, el pedido de videos a la carta, acceder a contenidos multimedia de alta calidad, además de permitirle controlar el modo de reproducción, siendo los proveedores más representativos del servicio: Netflix, YouTube y Vimeo.

A pesar de los beneficios del servicio de VoD, existen un conjunto de problemas que dificultan la experiencia e interacción del usuario en el entorno de televisión, dentro de estos están: el crecimiento de los catálogos de contenidos multimedia, el tiempo que puede emplear un usuario navegando por estos, los limitados métodos de entrada para navegar a través de ellos, entre otros. De esta manera, los principales retos del servicio de VoD son: agilizar el acceso y permitir el consumo adecuado del contenido multimedia en entornos televisivos [3].

Los contenidos multimedia y en especial los videos, tienen la capacidad de influir en el estado de ánimo de una persona [4]. Esto se debe a que su contenido es captado por dos órganos sensitivos como son la vista y el oído, los cuales reciben constantemente estímulos del medio externo que permiten que una persona esté alerta ante cualquier situación. Esto indica que es posible utilizar las emociones como método para la clasificación de contenido de video siendo el musical el más óptimo para utilizar.

La música tiene la capacidad de transmitir emociones y esto se debe a la influencia de diferentes características inherentes a la música como son: *arousal*, *valence*, ritmo, tempo, tono, modo, entre otros. Estas características son importantes ya que ayudan a clasificar el contenido en un determinado estado de ánimo. Además, la música por lo general está asociada con diferentes patrones de señales acústicas, por ejemplo, mientras el *arousal* se relaciona con el tempo (rápido/lento), la intensidad del tono (alta/baja), *pitch* (alto/bajo), *loudness* (alto/bajo), el timbre (brillante/suave), la valencia está relacionado con el modo (mayor/menor) y armonía (consonante/disonante) [5].

De igual manera, se observa que la percepción de la emoción raramente depende de una única característica musical, sino de una combinación de ellas. Por ejemplo, los acordes fuertes y los acordes agudos pueden sugerir

más valencia positiva que acordes suaves y acordes de tono bajo, cualquiera que sea el modo [6]. De esta manera, considerar las emociones como método de entrada para la navegación de contenidos, puede ayudar a mejorar el problema de agilizar el acceso a ellos en el servicio de VoD.

En este artículo se describe el proceso de conformación de un *dataset* de contenidos de video musicales, el cual es necesario para la clasificación emocional de los contenidos multimedia a ser difundidos mediante un servicio de VoD. Un *dataset* se define como un conjunto de datos representativos, que contiene variables o características de un determinado elemento. El *dataset* de video conformado en este trabajo es una representación de los videos musicales y sus características, que proporciona un modelo de programación relacional, coherente e independiente del origen de datos [7].

Así, el aporte principal de este artículo es la conformación de un *dataset* de contenidos musicales de video, el cual fue generado según el modelo de clasificación de emociones propuesto en la sección 2.3.1, donde el espacio cartesiano musical es dividido en cinco emociones o estados de ánimo: excitado, feliz, relajado, triste y enojado. De igual manera, para la verificación de los contenidos clasificados en el *dataset*, se desarrolló una herramienta de clasificación de contenidos musicales en el lenguaje Java. Finalmente, a modo de evaluación del *dataset* de contenidos musicales de video, se diseñó e implementó un servicio de VoD basado en emociones donde es usado para clasificar el contenido, el cual se describe en la sección 4.2.

El resto del artículo está organizado de la siguiente manera: en la sección 2 se presenta la metodología usada para la conformación del *dataset*. En la sección 3 se describe la propuesta del *dataset* de contenidos de video musical. En la sección 4 se describen los resultados obtenidos como son: la herramienta Java para reconocimiento de emociones y el servicio de VoD basado en emociones. Finalmente, en la sección 5 se presentan las conclusiones y trabajos futuros derivados de la presente investigación.

2. Metodología

Para la conformación del *dataset* de contenidos de video musicales construido en este artículo, se proponen cinco fases a saber: estudio de características musicales y modelos de emociones, exploración de soluciones comerciales basadas en emociones, propuesta de un modelo basado en emociones, diseño y construcción del *dataset* de contenidos multimedia musicales y evaluación del *dataset* de contenidos musicales de video (ver Figura 1).

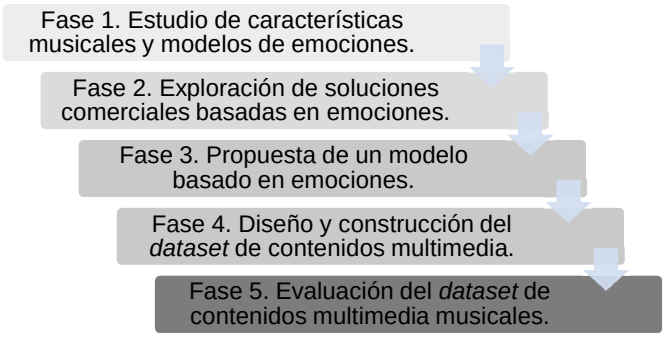

Figura 1. Metodología para la generación del dataset, fuente: propia

En la primera fase se realizó el estudio de las principales características musicales y los principales modelos de emociones, con el fin de identificar claramente la relación entre los contenidos multimedia y las emociones. En la segunda fase se hizo una exploración de las herramientas, librerías y tecnologías para la clasificación de contenidos multimedia de acuerdo a las emociones, con el propósito de identificar posibles opciones a nivel de desarrollo para la construcción del *dataset*. En la fase tres se presenta la propuesta adaptada del modelo de clasificación de emociones, obtenida a partir de los modelos presentados en la fase uno. El modelo de clasificación propuesto fue utilizado para conformar el *dataset* a partir de las tecnologías identificadas en la fase dos. Por su parte, en la fase cuatro se realiza la construcción del *dataset* de contenidos multimedia, considerando el modelo de clasificación propuesto en la fase tres y las herramientas seleccionadas en la fase dos. Finalmente, en la fase cinco, el *dataset* de contenidos multimedia es evaluado mediante la implementación de un servicio de VoD basado en emociones.

La fase uno de la metodología es abordada en la sección de características musicales y modelos de emociones (sección 2.1). La fase dos es considerada en la sección de soluciones comerciales (sección 2.2). La fase tres de la metodología se aborda en la sección de propuesta de modelo clasificación de emociones (sección 2.3). La fase cuatro es incluida en la sección de procedimiento de generación del *dataset* (sección 3). Finalmente, la fase cinco es abordada en la sección de resultados obtenidos del *dataset* de contenidos multimedia (sección 4).

2.1 Estudio de características musicales y modelos de emociones

Existen diferentes características musicales que permiten asociar un contenido multimedia con el estado de ánimo que puede presentar una persona en determinado momento. Así mismo, dichas características pueden ser usadas para inferir la emoción que percibe una persona al escuchar ese contenido musical. En este apartado se describen un conjunto de características musicales, que comúnmente

son usadas para determinar el estado de ánimo de una persona. Dentro de estas variables se encuentran: *valence*, *arousal,* ritmo, tempo, *speechiness*, *liveness*, *acousticness*, *danceability* y modo (ver Tabla 1).

Tabla 1. Características musicales del contenido, fuente: propia

Característica	Descripción
Valence	Esta propiedad musical describe la positividad musical transmitida por una pista de audio. Las pistas con alta valencia están asociadas con emociones positivas tales como: estar feliz, alegre eufórico, entre otros. Por otra parte, las pistas de con baja valencia están asociadas a emociones negativas o estados de ánimo tales como: tristeza, depresión, enojo [8].
Arousal	Representa una medida de percepción de la intensidad y la actividad a lo largo de la pista musical. Típicamente las pistas rápidas que cuentan con sonidos fuertes y ruido, como por ejemplo el rock pesado, tendrían una alta energía, mientras que una pista de música clásica como: Air de Beethoven, estaría bajo en la escala de energía. Otras características que contribuyen en este atributo son el rango dinámico musical, percepción del volumen, timbre, entropía general [8].
Ritmo	El ritmo es el patrón de pulsos o notas de fuerza variable, que se describe a menudo en términos del tempo, métrica o fraseo. Una canción con un ritmo rápido a menudo se percibe como una alto arousal, además un ritmo fluido se asocia generalmente con una valencia positiva, mientras que un ritmo firme está asociado con una valencia negativa [5].
Tempo	Esta propiedad sirve en una pieza musical para transmitir emociones, de tal modo que la música rápida según estudios es percibida o relacionada con emociones activas (felicidad), mientras que la música lenta tiende a percibirse como una emoción pasiva (tristeza). El tempo varía usualmente entre 20ppm y 240ppm, aunque puede tomar valores menores o mayores a estos [8].
Speechiness	Esta es una característica que permite detectar la presencia de las palabras habladas en una pista. Los valores que son superiores a 0.66 describen pistas que probablemente están hechas totalmente de palabras. Por su parte, valores comprendidos entre 0.33 y 0.66 describen pistas que pueden contener tanto música como interpretación de palabras. Finalmente, los valores por debajo de 0.33 probablemente puedan ser solo música [8].
Liveness	Esta propiedad musical permite detectar la presencia de audiencia en la grabación. De este modo, con valores del 1.0 la probabilidad de que la pista sea en vivo es alta. Los valores entre 0.6 y 0.8 describen pistas que pueden o no estar en vivo o contener audiencia simulada, por lo general está simulación es colocada al inicio o al final de la pista musical. Por último, los valores que estén por debajo de 0.6 son grabaciones que se han hecho en un estudio [8].
Acousticness	Es la característica que permite medir la probabilidad de que una grabación se haya creado únicamente con elementos tales como la voz e instrumentos acústicos, en lugar de utilizar elementos electrónicos. Las pistas con valores bajos suelen incluir guitarras eléctricas, distorsión, sintetizadores, entre otros. Así mismo, los valores cercanos a 1 indican que la canción presenta esta característica [8].
Danceability	Característica que permite conocer que contenido es el más adecuado para el baile. Las pistas con valores cercanos a 1, son los que mayormente reflejan esta característica en todo su desarrollo. Según [8], algunos elementos que permiten caracterizarla son el tempo, la estabilidad del ritmo, la regularidad que presente la pista, entre otros.
Modo	Es la propiedad que indica la modalidad (mayor o menor) de una pista, es decir el tipo de escala de la que deriva su contenido melódico [8].

2.1.1 Modelos de emociones

Las emociones pueden ser reducidas a un núcleo afectivo específico: placer o displacer. Otros estudios sugieren modelos en dos dimensiones, tales como el modelo de *arousal-valence*, donde el *arousal* y la valencia pueden ser positivas o negativas pudiendo caracterizar cualquier emoción por sus coordenadas, en un espacio de dos dimensiones, existiendo también modelos en tres dimensiones. A continuación, se describen algunos de los modelos de emociones más difundidos, dentro de los que se encuentran: modelo circunflejo de Russell, modelo de Hevner, modelo de Plutchick.

Modelo circunflejo: El modelo circunflejo o de Rusell (ver Figura. 2), es tal vez uno de los modelos más investigados, una estructura circular de dos dimensiones (valencia/activación), que parte el espacio en cuatro cuadrantes, en el cual las emociones son trazadas basándose en su nivel de actividad (activo/pasivo) y su valencia (positiva/negativa). El modelo circunflejo del afecto, muestra que los estados afectivos surgen de interpretaciones cognitivas de sensaciones nerviosas centrales, que son el producto de dos sistemas neurofisiológicos independientes, uno está relacionado con la valencia (placer/desagrado) y el otro al *arousal* (estado de alerta). Cada emoción puede ser vista como una combinación lineal de estas dos dimensiones o diversos grados de valencia [9]**.**

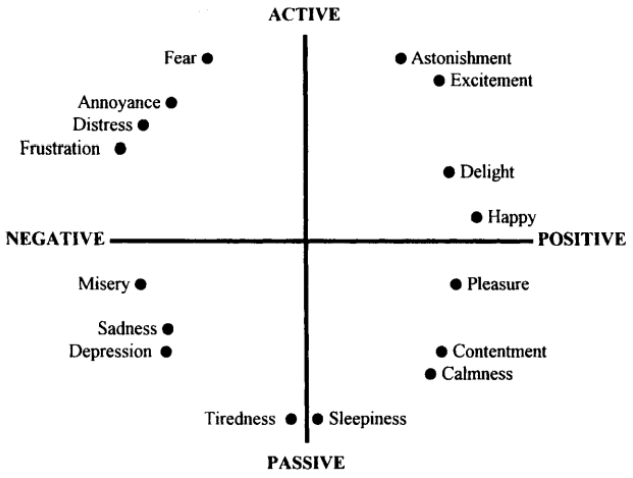

Figura 2. Modelo circunflejo de las emociones adaptado por Russell, tomado de [9]

El modelo tiene una estructura clara que tiene implicaciones en la forma en la que las emociones se experimentan en términos de la evaluación cognitiva, es decir como bueno y como malo o deseado y no deseado, a partir de respuestas fisiológicas o niveles de *arousal* y *valence*.

Modelo de Hevner: Este modelo se desarrolló con base a una experimentación en donde se manipularon características musicales como la melodía, armonía, modo, ritmo y tempo, a piezas musicales. El experimento consistió en darle a escuchar a un grupo de personas, la versión original y una versión modificada, en la cual se alteró una sola característica. Después de escuchar las dos versiones, se les preguntaba a las personas acerca de la emoción que describía mejor la pieza escuchada, y se les pedían señalar en el modelo dicha emoción. Lo anterior permitió realimentar el modelo inicial con diferentes emociones, tal como se muestra en la Figura. 3 [10].

Según el estudio realizado, se concluyó que utilizando la característica e identificando el estado de ánimo de los participantes, es posible modificar el estado emocional de una persona de forma gradual, siendo el tempo y el modo las variables que más impacto generaron en los oyentes.

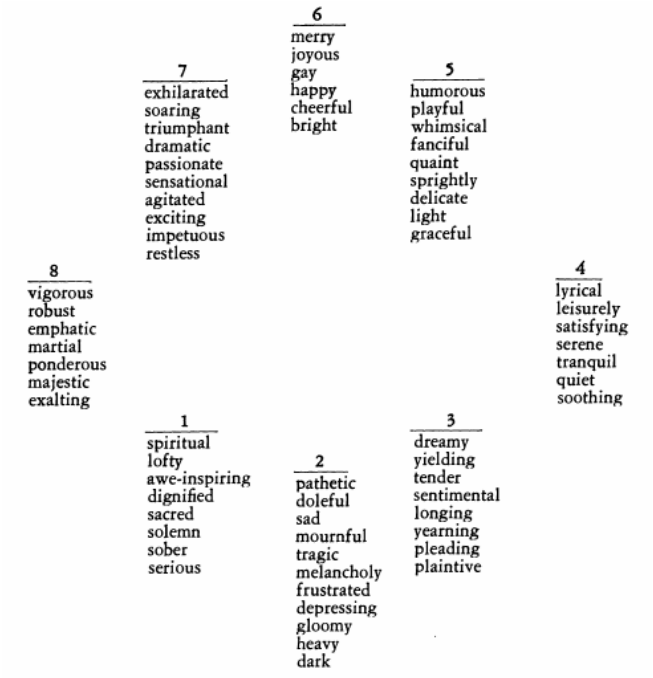

Figura 3. Modelo de las emociones de Hevner, tomado de [10]

Modelo de plutchick: Es un modelo tridimensional en el cual se postulan 8 emociones primarias (tristeza, sorpresa, miedo, ira, entre otras) de las cuales se derivan las demás, proponiendo que las emociones varían en polaridad, intensidad y grado de similitud, donde las emociones se intensifican a medida que se mueve desde el exterior hacia el centro de la circunferencia. A modo de ejemplo, una emoción como el aburrimiento, puede ser intensificada en odio. Además, cada sector del círculo tiene su emoción opuesta correspondiente, lo contrario de la alegría y al mismo tiempo opuesto a su posición en la figura se encuentra la tristeza (ver Figura. 4). Las emociones con ningún color representan un estado de ánimo que es una mezcla de dos estados emocionales primarios por ejemplo la anticipación y la alegría se combinan para ser el optimismo [11].

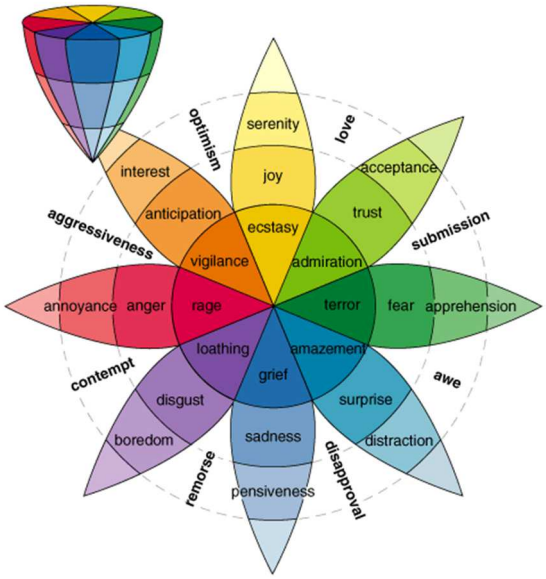

Figura 4. Modelo tridimensional de plutchick, tomado de [11]

En este trabajo, se decidió escoger el modelo circunflejo de dos dimensiones planteado por Russell, debido a su sencillez, amplio estudio y aceptación por parte de diferentes profesionales del área. Así la propuesta del modelo de clasificación usada en este trabajo parte de este modelo y es presentada con detalle en la sección 2.3.

2.2 Exploración de soluciones comerciales basadas en emociones

A continuación, se muestran un conjunto de soluciones comerciales y/o librerías de desarrollo que utilizan las emociones con base en las características del contenido musical, para proporcionar datos que pueden usarse en la generación de recomendaciones de contenidos multimedia. Dentro de las API's exploradas se encuentran: Musicovery, Gracenote y EchoNest.

Musicovery: La API de Musicovery[1] proporciona datos para generar recomendaciones de música y listas de reproducción de todo tipo: desde un estado de ánimo, un artista, una pista, un género / estilo, un tema, un período / año (ver fig. 5). La respuesta, una lista de pistas / artistas, se puede filtrar y personalizar de acuerdo a varios factores como son: popularidad, país oyente, tipo de similitud. Se puede acceder libremente a la API Musicovery, con la restricción de solo poder realizar 200 consultas en total. De este modo, es necesario solicitar una clave de API, permitiendo extender a 5000 las consultas. Para consultar sin limitación Musicovery API, necesita una clave de API Premium. La API tiene soporte para realizar los desarrollos en lenguaje PHP y el formato de respuesta del api es en JSON (JavaScript Object Notation) o XML (eXtensible Markup Language) [12].

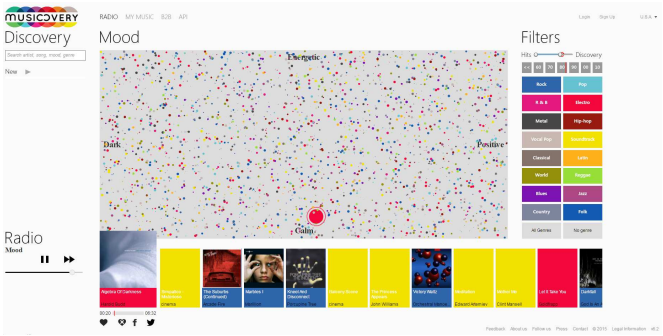

Fig. 5. Interfaz de usuario Musicovery, tomado de [12]

La interfaz de usuario de musicovery muestra un servicio web de radio. Este trabaja en forma similar al modelo de *arousal-valence*, en el cual basta con presionar cualquier de los puntos en su interior y este le proporciona la canción para escuchar, además de poder realizar las búsquedas por género (ver Figura 5).

Gracenote: Esta API ofrece un conjunto de metadatos de música a través del protocolo de internet HTTP (Hypertext Transfer Protocol), permitiendo búsquedas tales como género, región de origen y el estado de ánimo. Además, ofrece las búsquedas de portadas de disco, imágenes biografía, artículos y otros contenidos relacionados. Para hacer uso de esta API es necesario conseguir una ID de cliente y una clave API que le autoriza hacer llamadas al servicio gracenote con un número limitado de peticiones. Este API ofrece clientes para aplicaciones web y móviles, el formato de respuesta del api es XML o JSON; y permite el desarrollo en lenguajes como: Python, PHP, Ruby, Java, pero no presta soporte para estos por ser proyectos no oficiales de gracenote [13].

EchoNest: Es una herramienta desarrollada en los laboratorios del MIT por Tristán Jehan y Brian Whitman, la cual es propiedad de Spotify. Esta API permite obtener características del contenido tales como: valencia, energía, tempo, duración, popularidad, entre otros (ver fig. 6). Para hacer uso de la API de EchoNest[2], es necesario obtener una llave (clave) de la API, la cual da acceso o permite hacer 20 peticiones por minuto de manera ilimitada.

Esta herramienta además proporciona los mensajes de respuestas en formato JSON o XML. Entre los lenguajes a los que da soporte se encuentran: Python, Java, Ruby, PHP, Objective-C/iOS, C++, Javascript, para los cuales se encuentran librerías que permiten implementar las funcionalidades básicas que provee EchoNest. Algunas de las librerías oficiales son: pyechonest la cual es la biblioteca para Python, Jen para clientes Java, Enios para Objective-C. Así mismo, dentro de las librerías no oficiales se encuentran: PHP echonest, Ruby echonest, node echonest para Javascript, entre otros [8].

Figura 6. Página web oficial de EchoNest, tomado de [8]

En este trabajo se optó por utilizar la API de echonest debido a que es ampliamente utilizada por servicios como spotify y twitter, además del hecho de contar con peticiones ilimitadas con respecto a las otras dos soluciones presentadas.

1 Proporciona datos para la recomendación de música y generación de listas de reproducción. Página web oficial de Musicovery: http://musicovery.com/.

2 EchoNest es una plataforma inteligente de música, encargada del análisis de datos musicales. Página web oficial de EchoNest: http://the.echonest.com.

2.3 Propuesta del modelo de clasificación de emociones

Para la conformación del *dataset* de video, se eligió trabajar con la API de EchoNest, puesto que a pesar de que permite realizar 20 consultas por minuto, estas no tienen un límite en cantidad, a diferencia de gracenote y musicovery. Además de lo anterior, esta API presenta soporte para trabajar con lenguajes como Python y Java, los cuales han sido usados para el desarrollo de los distintos prototipos software realizados en el presente trabajo. Para el desarrollo con Python se hizo uso de la librería pyechonest3, mientras que en el caso de Java se utilizó la librería Jen4.

2.3.1 Modelo de emociones

El *dataset* de contenidos fue conformado teniendo en cuenta las cinco emociones consideradas por la API de EchoNest, y tomando como base el modelo de emociones de dos dimensiones (*arousal-valence*). Los estados de ánimo considerados en la representación del modelo son: excitado, feliz, relajado, triste y enojado. Además, se ha elegido trabajar con el modelo de *arousal-valence*, ya que es un buen marco de referencia que facilita su uso y adaptación a nuevos diseños, debido a su simplicidad de dos coordenadas para la identificación de la emoción asociada al contenido. El nuevo modelo que se plantea en este artículo es una adaptación de *arousal-valence*, considerado también en otros trabajos e investigaciones sobre el tema [5, 6, 14, 15]. El nuevo modelo generado se puede apreciar en la Figura 7.

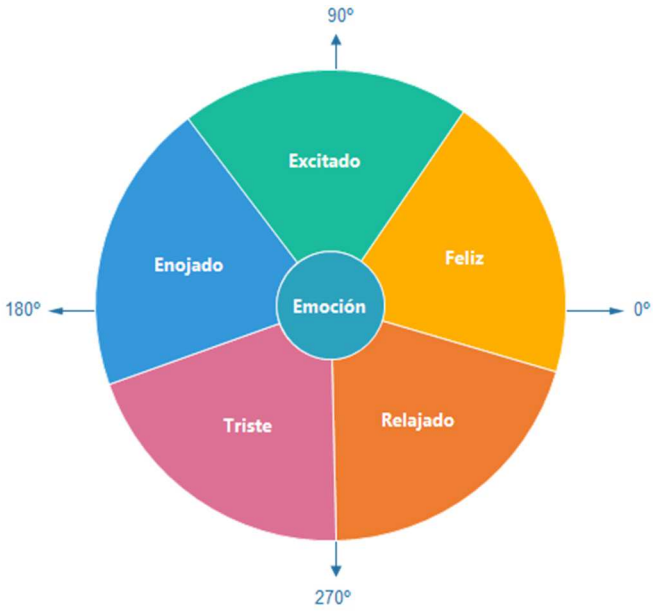

Figura 7. Modelo de arousal-valence adaptado a 5 emociones, fuente: propia

Cada emoción descrita en la Figura 7 tiene una amplitud de 72º, y el rango para el que está determinado cada estado de ánimo se puede observar en la Tabla 2.

Tabla 2. Estados de emoción, según el rango de ángulos, fuente: propia

Rango de ángulos	Emoción asociada
<54º y ≥342º	feliz
≥54º y <126º	excitado
≥126º y <198º	enojado
≥198° y <270º	triste
≥270º y <342º	relajado

3. Propuesta del dataset de contenidos multimedia

A continuación, se describen los pasos realizados para la conformación del *dataset* de contenidos de video musical. Además, se muestran las librerías y demás herramientas utilizadas para su generación. En la Figura 8, se presenta el diagrama de flujo del proceso de generación del *dataset*, donde se incluyen las siguientes etapas: generación de listado de contenidos populares, obtención de parámetros musicales, asociación de contenidos por emociones, obtención URL youtube, consolidación del catálogo y descarga de contenidos.

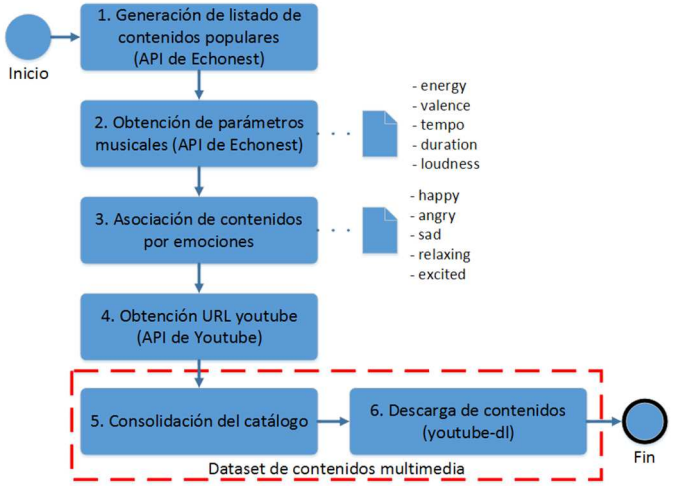

Figura 8. Diagrama de flujo del dataset de contenidos multimedia, fuente: propia

Para la generación del *dataset* de contenidos multimedia de video, se hizo uso de la librería *pyechonest* de Python, para lo cual es necesario adquirir una llave de acceso a la API, a través del registro en la página web oficial de EchoNest. Además, se utilizaron varias librerías adicionales de Python que fueron necesarias para la conformación del *dataset* como: math, youtube_dl, urllib, cookielib, sys, time, entre otros. A continuación, se describen cada uno de los pasos

3 Librería pyechonest de Python para el api de echonest, online: http://echonest.github.io/pyechonest/.

4 Librería Jen de Java para el api echonest, online: https://github.com/echonest/jEN.

seguidos para la generación del *dataset* de contenidos multimedia de video.

En el paso “1” a través de la API de Echonest, se procede a generar un listado con las canciones más populares del catálogo de contenidos de Echonest. Para lo anterior, se genera un script encargado de realizar 18 peticiones cada 60 segundos, dadas las restricciones de acceso provistas por la API.

En el paso “2” se obtiene con ayuda de la API de Echonest un conjunto de características asociadas al contenido multimedia, dentro de estas se encuentran: *title*, *artist*, *energy*, *valence*, tempo, entre otras. En la Figura 9 se visualizan cada una de las características mencionadas, las cuales son pieza clave para para la asociación de las emociones a cada video musical. Además de lo anterior, en este paso se realiza la discriminación de las canciones que puedan resultar repetidas en el proceso de búsqueda.

```
if(cur not in lista_repetidos):
    title=cancion.title
    artist=cancion.artist_name
    energy=cancion.audio_summary['energy']
    valence=cancion.audio_summary['valence']
    liveness=cancion.audio_summary['liveness']
    tempo=cancion.audio_summary['tempo']
    speechiness=cancion.audio_summary['speechiness']
    acousticness=cancion.audio_summary['acousticness']
    instrumentalness=cancion.audio_summary['instrumentalness']
    mode=cancion.audio_summary['mode']
    time_signature=cancion.audio_summary['time_signature']
    duration=cancion.audio_summary['duration']
    loudness=cancion.audio_summary['loudness']
    danceability=cancion.audio_summary['danceability']
    url=''
    visua=''
```

Figura 9. Obtención de las características del contenido, fuente: propia

En el paso “3” se realiza la asociación de las emociones con el contenido multimedia de video, usando para ello los valores de los parámetros de *arousal* y *valence*, los cuales permiten obtener el ángulo trigonométrico sobre el modelo de 5 emociones presentado en la sección 2.3.1 (ver Figura 10).

Dado que los valores obtenidos con la API de Echonest de las propiedades de *arousal* y *valence* están comprendidos entre 0 y 1, pero en el modelo de las emociones estos valores oscilan entre -1 y 1, se procede a normalizar los valores obtenidos con la API usando las ecuaciones 1 y 2. De esta forma se puede visualizar de una mejor manera si el contenido presenta una *valence* o *arousal* negativo.

$$arousal\_Normalizado = 2(arousal - 0.5) \quad (1)$$

$$valence\_Normalizado = 2(valence - 0.5) \quad (2)$$

```
e_n=2*(energy-0.5)
v_n=2*(valence-0.5)
ang=math.degrees(math.atan2(e_n,v_n))
if ang<0:
    ang=ang+360
else:
    ang=ang

# Estados de animo : 5 estados
if (ang>0 and ang<=54) or (ang>342):
    ani2='Happy'
elif ang>54 and ang<=126:
    ani2='Excited'
elif ang>126 and ang<=198:
    ani2='Angry'
elif ang>198 and ang<=270:
    ani2='Sad'
elif ang>270 and ang<=342:
    ani2='Relaxing'
```

Figura 10. Obtención de la emoción del contenido, fuente: propia

En el paso “4” se procede a obtener para cada contenido musical, la URL disponible en youtube del video asociado a este. Para lo anterior, se hizo uso de la librería youtube-dl[5] y de la API de youtube. La librería de youtube-dl permite filtrar, seleccionar, descargar y manipular el contenido de youtube por medio de las características como la calidad, titulo, formato, categorías, fechas, entre otros. Adicionalmente, esta librería permite la captura de imágenes de cada uno de los videos, las cuales serán presentadas al momento de visualizar el catálogo.

Con ayuda de dicha librería se realiza la búsqueda de los títulos para todos los videos en la categoría de música de la API de youtube y se hace una comparación con los proporcionados por la API de EchoNest. Para ello se filtran las canciones con una relación de correlación más alta a nivel del texto, haciendo uso del método SequenceMatcher de la librería difflib de Python. Este método retorna un valor entre 0 y 1, el cual representa el porcentaje de correlación entre los títulos del contenido. Una vez constatados los valores de correlación, se seleccionan las URL que sobrepasen el valor de 0.6 y que cumpla las otras condiciones. Para acotar un poco más el listado de videos encontrados, se agrega la variable de los videos más visitados, que en su gran mayoría son los videos oficiales de las canciones y los que mejor resolución presentan.

Una vez obtenidas las URL de los videos en el paso “5”, se procede a realizar la consolidación del *dataset* en un archivo JSON (ver Figura. 11), el cual es un formato

5 Es un programa de línea de comandos pequeña para descargar videos de YouTube.com y otros pocos sitios. Se encuentra online en: https://github.com/rg3/youtube-dl.

más liviano y flexible que XML. El archivo JSON se administra por medio del gestor de base de datos “tinydb”, el cual está diseñado para ser simple de usar y no necesita de un servidor externo, además de ser compatible con todas las versiones modernas de Python.

Finalmente, en el paso “6” se realiza la descarga de los contenidos multimedia de video en formato .mp4 haciendo uso de la librería youtube-dl.

{"_default": {"1": {"liveness": 0.094833, "energy": 0.717726, "tempo": 129.882, "speechiness": 0.106457, "currency": 0.22104064240294416, "instrumentalness": 0.338526, "duration": 313.94844, "view_count": 822693136, "mood": "Excited", "artist": "Major Lazer", "url": "http://www.youtube.com/watch?v=YqeW9_5kURI", "title": "Lean On", "acousticness": 0.033117, "danceability": 0.862321, "time_signature": 4, "loudness": -5.691, "valence": 0.421374, "mode": 0}, "2": {"liveness": 0.064907, "energy": 0.480501, "tempo": 80.025, "speechiness": 0.081511, "currency": 0.16323143005491134, "instrumentalness": 1e-06, "duration": 229.52576, "view_count": 118733053, "mood": "Angry", "artist": "Wiz Khalifa", "url": "http://www.youtube.com/watch?v=RgKAFK5djSk", "title": "See You Again (feat. Charlie Puth)", "acousticness": 0.369444, "danceability": 0.689399, "time_signature": 4, "loudness": -7.503, "valence": 0.26631, "mode": 1}, "3": {"liveness": 0.515838, "energy": 0.873779, "tempo": 168.901, "speechiness": 0.191693, "currency": 0.0005813953488372093, "instrumentalness": 7.04e-05, "duration": 233.10512, "view_count": 782095, "mood": "Excited", "artist": "Rage Against The Machine", "url": "http://www.youtube.com/watch?v=3L4YrGaR8E4", "title": "Bulls on Parade", "acousticness": 0.0250803, "danceability": 0.398765, "time_signature": 4, "loudness": -10.218, "valence": 0.510731, "mode": 1},

Figura 11. Muestra del archivo JSON generado, fuente: propia

De esta manera, se descargaron un total de 200 videos para la conformación del *dataset* de contenidos musicales de video, agrupando alrededor de 40 canciones por emoción. En la Figura 11 se presenta una muestra de los campos asociados a cada uno de los contenidos de video musicales que constituyen el archivo JSON generado.

Tabla 3. Características musicales para cada video, archivo JSON, fuente: propia

Característica	Valor	Color
id	1	
liveness	0.094833	
energy	0.717726	
tempo	129.882	
speechiness	0.106457	
currency	0.22104064240294416	
instrumentalness	0.338526	
duration	313.94844	
view_count	822693136	
mood	Excited	
artist	Major Lazer	
url	http://www.youtube.com/watch?v=YqeW9_5kURI	
title	Lean On	
acousticness	0.033117	
danceability	0.862321	
time_signature	4	
loudness	-5.691	
valence	0.421374	
mode	0	

Cada contenido de video musical del *dataset* contiene los campos mostrados en la tabla 3. Dentro de estos campos se destaca las propiedades de *energy* (*arousal*) y *valence*, las cuales fueron usadas para la determinar la emoción asociada a cada contenido multimedia.

4. Resultados y discusión

Para verificar el funcionamiento del modelo de clasificación de la sección 2.3 y explorar la API de Echonest, se desarrolló una herramienta en lenguaje Java para el reconocimiento del estado de ánimo asociado a un contenido multimedia musical. En la aplicación desarrollada se pueden observar propiedades musicales tales como: *energy* (*arousal*), *valence* (valencia), tempo, el ángulo formado por la valencia y energía en el plano cartesiano, emoción, entre otros; parámetros los cuales se pueden apreciar en la Figura 12.

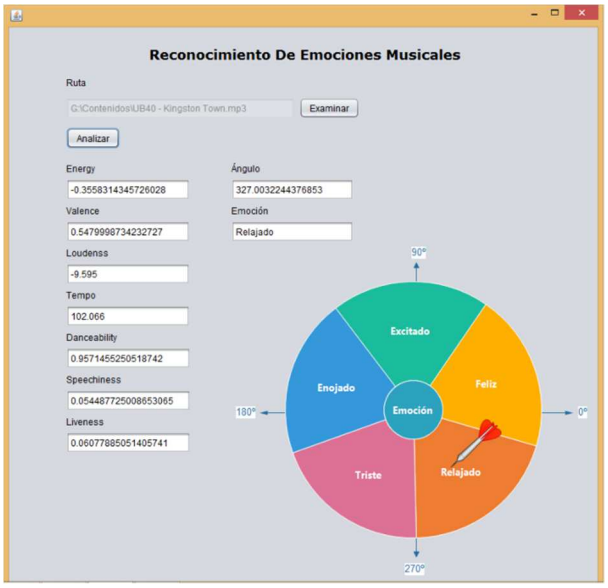

Figura 12. Reconocimiento de emociones musicales, fuente: propia

A modo de ejemplo, en la Figura 12 se observa como usando la librería Jen de Java (provista por Echonest), se analizaron las características musicales para la canción del grupo pop-reggae UB40 llamada Kingston Town. Los parámetros que permiten conocer a que emoción está asociada la canción son el *arousal*, el cual tiene un valor de -0.3558 lo que permite ubicarlo en la mitad inferior del modelo. Al observar el valor de la *valence* que arroja la API, se sabe que la canción presenta una emoción positiva, por tanto, se encuentra en el cuarto cuadrante del gráfico. El ángulo resultante entre la propiedad *arousal* (eje y) y la propiedad *valence* (eje x), permite analizar de acuerdo a la tabla 1 la emoción asociada a la canción. El anterior proceso puede evidenciarse en el diagrama de bloques de la herramienta de clasificación Java de clasificación (ver Figura 13).

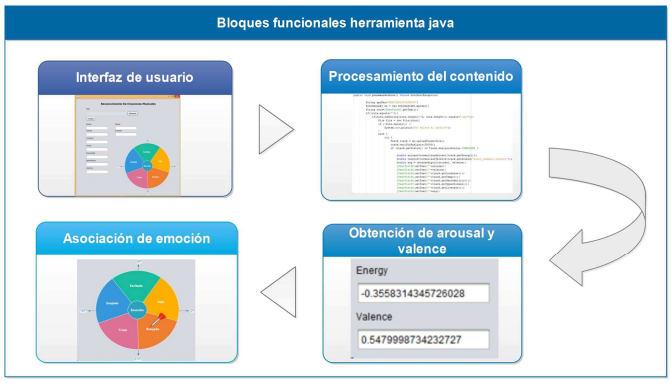

Figura 13. Bloques funcionales de la herramienta Java, fuente: propia

4.1 Sistema de VoD para validar el dataset

Con el propósito de evaluar el *dataset* de contenidos multimedia de video, se diseñó e implemento un servicio de VoD basado en emociones, el cual presenta un listado de contenidos multimedia de video clasificados por emoción. En la fig. 14 se presenta la arquitectura básica del servicio de VoD basado en emociones, la cual está formada por los siguientes módulos funcionales: proveedor de contenido multimedia, proveedor del servicio, proveedor de red y consumidor de contenido.

El proveedor de contenido multimedia es el módulo encargado de proporcionar los contenidos de videos al servicio de VoD, a través del servidor Apache vía protocolo HTTP; es allí donde se alojan los contenidos multimedia asociados al *dataset*. El proveedor del servicio, es el módulo gestor del servicio de VoD, allí el usuario puede consumir los contenidos de video a través de la web; para ello se hizo uso de tecnologías del lado del servidor como: PHP y el gestor base de datos MySQL, encargado de la gestión y almacenamiento de información de usuarios y del *dataset*. El módulo proveedor de red se encarga de brindar soporte de red (acceso a internet) al usuario, para la distribución del contenido multimedia. Finalmente, el módulo consumidor o cliente se hizo uso de tecnologías como: HTML5, JavaScript y el framework Bootstrap para el diseño, lógica de la interfaz y la reproducción del contenido. Para el consumo del servicio se requiere un navegador con soporte para JavaScript y HTML5, ejemplo: Google Chrome, Firefox, Opera, entre otros.

Figura 14. Arquitectura para el despliegue del servicio de VoD, fuente: propia

En la Figura 15 se presenta la interfaz web del servicio de VoD basado en emociones. Esta se encuentra formada por cuatro componentes principales: en “1” se encuentran un conjunto de botones que representan las cinco emociones de entrada, que permiten al usuario seleccionar manualmente el tipo de contenidos musicales que desea visualizar. Una vez el usuario escoge la emoción de entrada, se le presentan un conjunto de contenidos que han sido previamente clasificados en el *dataset* de videos musicales (“2”), según el modelo de clasificación de emociones presentado en la sección 2.3.1. A partir del catálogo presentado en “2”, el usuario puede escoger el contenido a visualizar, mediante el componente de reproducción (“3”). Finalmente, asociado al componente de reproducción existe un panel de control de reproducción que permite adelantar, atrasar, controlar el volumen y la resolución del video.

Figura 15. Interfaz de usuario del servicio de VoD, fuente: propia

5. Conclusiones y trabajos futuros

Por medio de características musicales como el *arousal* y la *valence* es posible clasificar un contenido multimedia en el espacio cartesiano de las emociones. Lo anterior permite agilizar el acceso a los contenidos multimedia del servicio de VoD, en el sentido que se personalizan las preferencias del usuario mediante una emoción de entrada.

El *dataset* de contenidos multimedia de video propuesto, constituye un aporte importante para el diseño e implementación de servicios de video basados en emociones, integrando las ventajas del *dataset* musical de Echonest y la información de video provista por la API de youtube. Así mismo, el modelo de clasificación de emociones presentado en la sección 2.3 puede ser considerado para el desarrollo de servicios de contenidos multimedia basados en emociones, que hagan uso de la API de Echonest (sistemas de recomendaciones, buscadores servicios publicitarios, entre otros).

La herramienta Java para la clasificación de contenidos musicales, permitió verificar la correcta clasificación en el proceso de conformación del *dataset* de contenidos musicales de video. Así mismo, esta herramienta permitió evaluar la pertinencia del modelo de clasificación de contenidos propuesto en la sección 2.3.

Las API´s provistas por Echonest en los lenguajes Python y Java, se constituyen como buenas alternativas para la implementación de servicios, que consideren la relación de los contenidos multimedia con los modelos emocionales. Lo anterior considerando que estas librerías permiten acceder a un conjunto relevante de características musicales, sin restricciones a nivel de la cantidad de consultas realizadas.

A modo de trabajo futuro, se pretende extender el modelo emocional de los contenidos multimedia, considerando el análisis sobre las letras de las canciones. Así mismo, se espera realimentar el servicio de VoD basado en emociones, mediante el diseño e implementación de un módulo para la captura implícita de emociones de entrada. De igual forma, se pretende vincular el catálogo de emociones presentado en este trabajo, con un sistema de recomendaciones clásico de contenidos multimedia musicales, buscando así atacar el problema de arranque en frío presente en los recomendadores.

6. Referencias

[1] K. Pripuzic, I. Zarko, V. Podobnik, I. Lovrek, M. Cavka, I. Petkovic, P. Stulic y M. Gojceta, «Building an IPTV VoD Recommender System: An Experience Report,» Telecommunications (ConTEL), 2013 12th International Conference on, pp. 155-162, 2013.

[2] J. Altgeld y D. Z. John, «The IPTV/VoD Challenge: Upcoming Business Models,» de Achieving the Triple Play: Technologies and Business Models for Success, Intl. Engineering Consortiu, 2006, p. 3.

[3] S. Paul, Digital video distribution in broadband, television, mobile and converged networks: trends, challenges and solutions, New york: Wiley, 2011.

[4] P. Andersen y L. Guerrero, Handbook of communication and emotion, California : Academic Press, 1997.

[5] H. Chen y Y. Yang, Music Emotion Recognition, Taiwan: CRC Press, 2011.

[6] R. Fay y M. Jones, Music Perception, Ohio: Springer, 2010.

[7] J. Rivera, «Acceso a Datos con DataSets en Visual Web Developer 2008/2010,» Conciencia Tecnológica, pp. 47-51, 2011.

[8] T. Jehan y B. Whitman, «The Echonest,» MIT Media Lab, Junio 2005. [En línea]. Available: http://the.echonest.com/. [Último acceso: 13 Enero 2016].

[9] J. Russell, «A Circumplex Model of Affect,» de Personality and Social Psychology, Vancouver , APA journals , 1980, pp. 1161-1178.

[10] K. Hevner, «Experimental Studies of the Elements of Expression in Music,» de The American Journal of Psychology, Illinois, Jstor, 1936, pp. 246-268.

[11] R. Plutchik, «The Nature of Emotions,» de American Scientist, Florida, American Scientist, 2001, pp. 344-350.

[12] V. Castaignet y V. Frederic, «Musicovery,» Junio 2006. [En línea]. Available: http://musicovery.com/api/doc/documentation.php. [Último acceso: 2 Marzo 2016].

[13] T. Kan y S. Scherf, «Gracenote,» Tribune Media Company, 1998. [En línea]. Available: https://developer.gracenote.com/web-api. [Último acceso: 2 Marzo 2016].

[14] J. Posner y J. Russell, «The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology, » de Dev Psychopathol, New York, NIH, 2005, p. 715–734.

[15] O. Meyers, A Mood-Based Music Classification and Exploration System, Massachusetts: Massachusetts Institute of Technology, 2004.