

Recuperación de Información y Recomendaciones en la Web



**FACULTAD DE
INGENIERÍA**



**UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY**

Curso 2022

Grupo 12

Twitter Bot para encontrar canciones

Alumnos:

Gonzalo Silva

Mateo González

Santiago Parodi

Docente:

Libertad Tansini

Índice

Índice	2
Problema	3
Enfoque de la Solución	3
Obtener información	3
Twitter Bot	4
Diseño e implementación	4
Arquitectura	4
Componentes	4
Backend	4
Bot	5
Web Scraping	5
Base de datos	5
Funcionalidades	5
Trabajo Futuro	6
Conclusiones	8

Problema

Una problemática frecuente en el día a día de las personas se da cuando tienen un fragmento de la letra de una canción presente en la mente en determinado momento pero no saben, o se acuerdan, el nombre de esta. Para solucionar esto existen distintas aplicaciones que, dado un fragmento de la letra de una canción, proporcionan el nombre de esta, pero requieren ser descargadas, añadiendo un paso extra de complejidad y logrando con esto que muchas veces se decida permanecer con la duda antes de pasar por este proceso.

Por lo tanto, el problema radica en que se carece, en muchos casos, de una manera rápida y eficaz de obtener el nombre de la canción dado un fragmento de la letra de esta.

Enfoque de la Solución

Para solucionar la problemática planteada, se creó un bot de twitter que, al etiquetarlo en un tweet indicando el fragmento de la letra de una canción, este contesta en el menor tiempo posible el nombre de una canción que posee dicho fragmento en su letra.

A la hora de hacer la aplicación, se encontraron dos problemas separados para atacar, uno es de donde conseguir la información de las canciones, y la otra es como hacer que nuestro bot se notifique cuando alguien lo etiqueta.

Obtener información

Para obtener las letras de las canciones se investigaron varias posibilidades, pero finalmente se optó por utilizar el sitio web [musica.com](http://www.musica.com) para obtener las letras de las canciones, ya que el sitio utiliza un id numérico en el url para acceder a cada canción, por ejemplo el link <https://www.musica.com/letras.asp?letra=2065103> es la canción “*Algo mejor que hacer*” de la banda el Cuarteto de Nos, y el link <https://www.musica.com/letras.asp?letra=2065105> es “*El balcón de Paul*” también de esta banda, y simplemente incrementando este id se puede llegar a todas las canciones que tienen almacenadas.

Luego haciendo web scraping se puede obtener la letra de las canciones y almacenar estas en la base de datos.

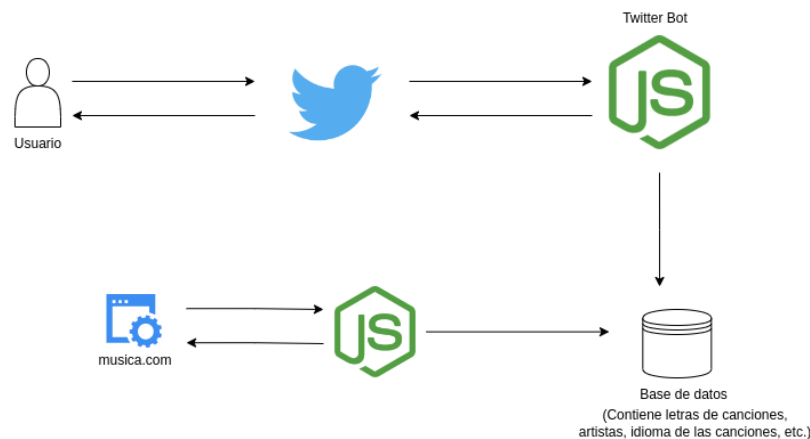
Twitter Bot

Para hacer que el bot se notifique cuando alguien lo etiquete se utilizó la API v2 de Twitter. Cuando un usuario publica un tweet etiquetando al bot ([@webirbot](#)) e indicando la letra de la canción a buscar, este busca la/las canciones que contienen esta letra mediante el uso de índices y luego retorna, en caso de haber encontrado, el nombre de la canción en cuestión, esto lo hace mediante una respuesta al tweet que lo etiquetó.

Diseño e implementación

Arquitectura

Para el funcionamiento de este bot, se requiere de la base de datos con las letras de las canciones almacenadas y una cuenta de Twitter Developers asociada al bot. En principio, el bot funciona únicamente local, es decir, para que el bot esté funcionando como se planea es necesario que se corra localmente en algún equipo. En un futuro, en caso de desearlo, se puede deployar el bot escrito en javascript en alguna plataforma que lo soporte, por ejemplo *Heroku*.



Componentes

Backend

El backend se separa en dos funcionalidades claves y bien separadas, una de ellas es el Bot que estará corriendo siempre escuchando eventos de twitter para saber cuando lo etiquetan en una publicación, y la otra es una función que se corre una única vez para poblar la base de datos con la información obtenida del sitio web.

Bot

El bot se conecta vía stream con la api de twitter utilizando el paquete “*twitter-api-v2*” donde escucha el evento “*data*” de la api de twitter. Cuando llega un evento se obtiene el texto del tweet, se procesa para limpiar el texto (quitar comas, puntos, pasar todo a minúsculas, etc.). Luego se envía una consulta a ElasticSearch utilizando el paquete que este ofrece para node “*@elastic/elastic-search*”, donde se busca entre las canciones almacenadas alguna que contenga el string enviado por el usuario.

Web Scraping

Para poblar la base de datos se utilizó la librería “*puppeteer*” para obtener la información de la página web musica.com, donde se accedió con un loop a las distintas canciones de este sitio web, y se extrajo de estas el artista/banda que las compuso y la letra de cada una. Luego la letra de las canciones se procesa para normalizarla un poco (se quitan comas y puntos, se pasa toda la letra a minúsculas, se quitan los textos descriptivos, etc), luego se obtiene el idioma de la canción utilizando la librería “*languagedetect*”

Base de datos

Los datos recopilados mediante *web scraping* son almacenados en una base de datos de elasticsearch, la que luego será consultada cuando se busqué un fragmento de letra de una canción.

Funcionalidades

Como se mencionó anteriormente, la funcionalidad principal de este bot es, dado un tweet donde el bot es etiquetado con una frase de una canción, devolver las canciones que contienen dicha frase, mediante el uso de distintas bases de datos y la utilización de índices invertidos.

Se aplicaron los conceptos aprendidos en el curso para ser capaces de encontrar coincidencias de canciones incluso en casos donde no hubo un “match” perfecto, con, por ejemplo, faltas de ortografía (en caso de que no se haya filtrado con el uso de comillas, donde se buscará únicamente aquellas que tengan un “match” perfecto).

Trabajo Futuro

Hay varias mejoras que se considera serían ideales a tener en cuenta para versiones futuras del bot:

- Conseguir una mejor fuente de información, ya que en este caso se usó musica.com, pero este sitio web no tiene un estándar a la hora de escribir la letra de las canciones, tiene algunas canciones repetidas, y algunas letras no están correctas, además de no estar siempre la letra completa o separada en párrafos para obtener las distintas estrofas de la canción. Por ejemplo se pueden ver estas dos imágenes:



LETRA



(Intro) How Yall Doin Out There Tonight? Yes, I Love You Too. I Just Wanna Talk To The Ladies Tonight. For All You Ladies, That Done Had Your Heart Broken, I Done Had My Heart Broken To, So I Can Relate, Ya Know. So If Its OK Wit You, Im Just Gonna Break It Down For The Fellas, That Dont Know That They Cant Just Break A Womens Heart. Alright? So Check It Out. I Gave You All Of My Love..... All To You. And You Dont Appreciate The Things That I Do. I Gave You All Of My Time..... Cause You Blew My Mind. I Let You Do Me Wrong, Cause Love Had Me So Blind. *I Was So Strong, But Im All Grown Up 'N', Im Dumb By Love 'N', I Used To Love You. You Tried To Play Me And Then Persuede me. I Still Remember The Pain You Gave Me. I Cant Believe You, I Gotta Leave You. Im Packin Up 'N' You Can Hit The Road Jack. You Can Take These Things That You Gave Me, Cause Im Never Comin Back.* All I Wanted..... Was A Little Love. But You Chose To Hang Out All Night With The Boys. So What Are You Saying?(Quit Wasting My Time) Trying To Claim Something That Aint Even Mine. *I Was So Strong, But Im All Grown Up 'N', Im Dumb By Love 'N', I Used To Love You. You Tried To Play Me And Then Persuede me. I Still Remember The Pain You Gave Me. I Cant Believe You, I Gotta Leave You. Im Packin Up 'N' You Can Hit The Road Jack. You Can Take These Things That You Gave Me, Cause Im Never Comin Back.* Why You Hurt Me?(Echo) Why You Hurting Me So Bad? Gave You My Time, You Played With My Mind. Why You Wanna Make Me Sad? (Repeat)



MORE THAN A WOMAN

Aaliyah

La canción 'More Than A Woman' se estrenó el 7 de julio de 2001. Este tema está incluido dentro del disco [Aaliyah](#)

LETRA



Passion, instant
Sweat me, feel me
Cupid's shot me
My heartbeat's racing
Tempt me, drive me
Feels so exciting
Thought of highly
It's yours entirely

I'll be
I'll be more than a lover
More than a woman
More than your lover
I'll be
I'll be more than another
More than a woman
More than another
I'll be
(I'll be more) more than your lover
(more) more than a woman

- Al ser canciones tienen muchos versos repetidos, los cuáles podrían quitarse al procesarlos, pero debido a la forma errática en la que están escritas las canciones en la fuente de información no se pudieron quitar las estrofas repetidas.
- Utilizar otro detector de lenguaje para reconocer el idioma de una letra dada, ya que actualmente se utiliza "languagedetect", la cual es una librería gratis pero no muy precisa. Una mejor alternativa sería utilizar la API de Google Translator.
- Implementar *features* extras que complementarán las funcionalidades del bot, por ejemplo, permitir al usuario añadir comodines para indicar que no se conoce ese fragmento de la canción.

Conclusiones

Se creó un *proof of concept* de un bot de twitter que contesta tweets en donde fue mencionado y que preguntan por el nombre de una canción dado un fragmento de la letra de esta. Se aplicaron los conceptos aprendidos a lo largo del curso, mediante la utilización de índices invertidos para consultar las letras de las canciones. Esto se realizó con *elasticsearch* que proporciona las herramientas necesarias para este tipo de consultas.

Si bien el bot es funcional, no deja de ser una *proof of concept*, por lo que se le pueden realizar mejoras en caso de realmente utilizarlo. Algunas de estas mejoras y agregados fueron mencionados a lo largo de este documento.