

Demystify Trends of Artists with Twitter, Reddit and Spotify API

Ketaki Khade
kkhade1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Nikita Shriram Navkar
nnavkar1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Swarali Patil
spatil32@binghamton.edu
Binghamton University
Binghamton, NY, USA

Tanmayee Kulkarni
tkulkar1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Yash Jawale
yjaware1@binghamton.edu
Binghamton University
Binghamton, NY, USA

ABSTRACT

Millions of people use Twitter and reddit as a social media web platform tool to express their thoughts and opinions on a range of subjects. As a result, a stream of text data, made up of enormous numbers of tweets, posts, comments, 'r/spotify', 'r/music', 'r/musician', 'r/singer' etc is generated. For this project, we're especially seeking tweets, posts etc that refer to certain Spotify tunes, music, albums and artists. We analyse those tweets, 'r/spotify', 'r/music', 'r/musician', 'r/singer', posts, and comments and gather track-related data using the Spotify API. Finally, we want to identify the artists who are now trending by using the information we've acquired in combination with a particular time frame.

KEYWORDS

Twitter API, Reddit API, Spotify API, MongoDB, Data Collection, Snowball Sampling

ACM Reference Format:

Ketaki Khade, Nikita Shriram Navkar, Swarali Patil, Tanmayee Kulkarni, and Yash Jawale. 2022. Demystify Trends of Artists with Twitter, Reddit and Spotify API. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

The music business is now constantly evolving, with various transitional music genres being popular at various points in time. For online music streaming, there are many different sources and platforms with a wide variety of genres and performers, such as Spotify, Apple Music, YouTube Music, Reddit etc. Everyone has varied musical preferences in terms of both genre and artist. Every time we hear something new, even if it's not always better, we usually share it with our friends, peers, and family, which eventually helps to create the music that is now in style or trending. We frequently have a tendency to tweet, publish, or share about it on social media using various hashtags, making it visible to the whole public. This

endless stream of unrestricted knowledge about music from numerous sources may be applied in novel ways in both academics and the music business. For this project, we'll be gathering real-time music-related data from social networking sites (particularly Twitter, and Reddit) and other music streaming services like Spotify, which provides a wealth of music-related data. To determine what musical style or performer is popular and trending when - we'll combine all this information from these three sources within a certain time frame.

2 DATA SOURCE

Here, in this project, we will be using Twitter data that we obtain from the Twitter Stream API and Reddit data from Reddit Stream API. We will collect information about top 10 streaming songs, or top Albums, or top Artists etc. and for this we will be using Spotify API. For implementation, we will use JavaScript and some of its libraries to fetch and handle the Requests.

2.1 Twitter API

Streaming HTTP protocol is used by Twitter for its streaming API. The App establishes a single connection between the client app and the API rather than returning data in batches via repeated requests from the client app, like a REST API would do. By using JavaScript, we will be able to build a client which sends requests using the streaming API and receives a continuous stream of data over, which will be filtered using the filtering key. Twitter data we have collected via sample stream (https://developer.twitter.com/en/docs/twitter-api/v1/tweets/sample-realtime/overview/get_statuses_sample) which provides an approximately 1

2.2 Spotify API

In this project we will be using Spotify web API to fetch data from a Spotify URL. This involves extracting the track ID from the URL and sending it as a parameter to the Spotify API to retrieve the details of any available track. This entails getting information such as artist name, album name. We can use this information to determine which artists are trading on twitter and reddit and which are top 5 songs which trading on twitter. We have collected Spotify data using <https://api.spotify.com/v1/search> endpoint.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

2.3 Reddit API

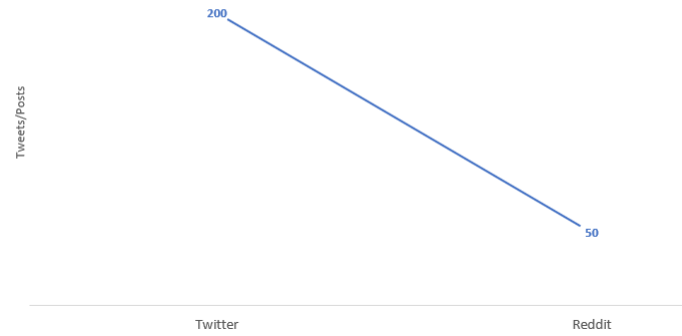
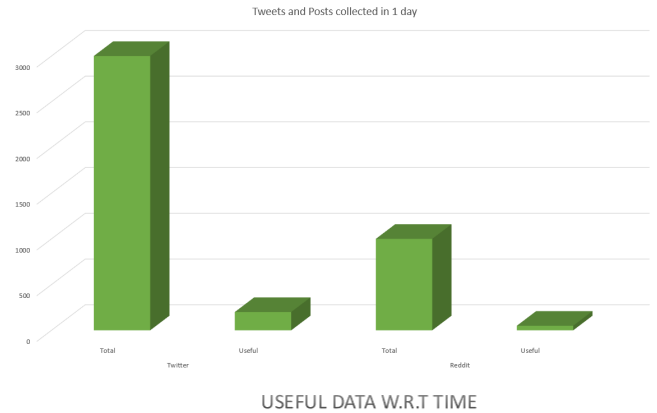
The Reddit REST API allows to access the user submitted and rated stories on reddit.com. We can use this information to collect people's reviews on trading songs on twitter. And we can also determine how many people are interested in talking about it. It also provides advanced functionality, including user account information and sub-reddit moderation. We have collected Reddit data using https://www.reddit.com/api/v1/access_token endpoint.

3 API METHODS

We will access data resources through standard HTTPS requests to an API endpoint. Web API uses appropriate HTTP verbs like; GET, POST, PUT and DELETE methods and returns all response data as a JSON object. As Twitter and Reddit API's allow us to find, retrieve and engage or create different resources, for this it uses two HTTP methods: POST and GET. In this case using GET method we can retrieve data from Twitter and Reddit API's. As Spotify Offers us API methods based on REST principles. We begin receiving a stream of data and then the data will be in JSON format in response.

4 IMPLEMENTATION

The data was repeatedly fetched from the Twitter Stream API using python. A counter will be added so that we can monitor the quantity of new tweets that arrive each day. The ideal method to utilize is Snowball Sampling, which illustrates a form of search approach to locate the most pertinent tweets, posts on Twitter and Reddit using Python. Relevant keywords include spotify, music, musicians, etc. Thus, "spotify" keyword is being used to filter the incoming tweets, which are subsequently sent to our MongoDB database. Additionally, we will be keeping track of the daily counter in the MongoDB database. There is no data gathering required for the Spotify API. We will solely make use of it to classify and validate the data we obtain from the Twitter API. The only thing we need to supply is the Track ID, which is the text that comes after "track/" and before the question mark in the track url. Using this track ID, we can then send a GET request to the Spotify API to obtain all the pertinent track information, including the artist, album, genre, and other details that we will need to evaluate our data. The loop that sends the requests to the Twitter and Reddit API endpoints has been implemented using System Cron job. Then, using our keyword (in this case, "spotify"), we wrote a function that filters incoming tweets and posts pushing the ones that contain the phrase to the MongoDB database. Additionally, we have created a method that updates the MongoDB database with the number of tweets and posts which will be analyzed each second.



5 MEASUREMENTS AND ANALYSIS

We will be planning to do analysis of musical data related to trending artist or genre based on artist or album popularity on Twitter or Reddit social media sites. We will be collecting data from Twitter, Reddit and Spotify using different APIs and update each new entry into database at real-time. To analyze data collected, we will retrieve data from database periodically and plot the graphs or charts using plotting and visualization libraries in python. We can also perform sentiment analysis based on positive or negative reaction got from different tweets on trending artist or album

6 PRELIMINARY EXPLORATION OF DATA

The information is all text and is comprised of tweets and posts with the keyword "spotify". We will utilize the track ID found in the Spotify URL to send GET queries to the Spotify API. An individual ID, the tweet/post content, and the URL object (which contains both the abbreviated and extended URL) are all contained in each MongoDB item. Additionally, we have a MongoDB storage location for the daily count.

7 DATA COLLECTION ESTIMATES

According to preliminary testing, not every tweet that contains the essential keywords is usable after gathering 100 tweets per hour on average. Only around 25 per of tweets are genuinely pertinent and useful for gathering information that we will later use in our study. After successfully deploying on VM, now we are able to process

around 2 million tweets per day. From these 2 million tweets, we are storing around 0.62 percent tweets which are filtered based on “spotify”. Similar to Reddit, just 20 per of posts that include relevant terms are collected after 50 posts/hour, which can ultimately be used to extract significant data.

8 DATA FLOW

As shown in Figure 1, we first retrieve the data from the Twitter and Reddit APIs using those respective APIs. To do this, we’ll utilize a JS script and the Needle library, which gives us a simple way to repeatedly send API calls to the Twitter and Reddit Stream APIs. In order to save these tweets and posts in our MongoDB database, we first search the text field of the received data for pertinent key terms like “spotify.” The URL will provide us the Spotify track ID, which we will use to get the track information information using the Spotify API, such as the artist’s name, genre, etc. After that, we may save these specifics in a different MongoDB database and use them for analysis and visualization.

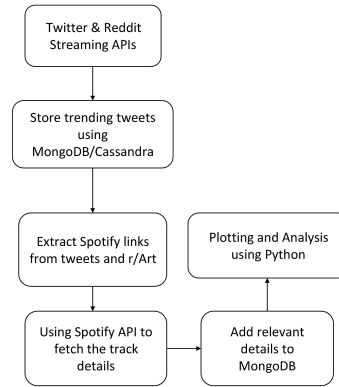


Figure 1

9 CHANGES COMPARED TO PROPOSAL

We have added required API endpoints of Twitter, Reddit and Spotify. For expanding the keyword search, we have analysed a method named Snowball Sampling. We have changed the keyword r/Art to r/Spotify, r/Music, r/Singer, r/musician which will help us to get more relevant data.

10 REFERENCES

Martin Pichl, Eva Zangerle, Günther Specht. Combining Spotify and Twitter Data for Generating a Recent and Public Dataset for Music Recommendation, 2014

Vipul Mishra, Kongmeng Liew, Romain Hennequin, Elena V. Epure, Eiji Aramaki. Are Metal Fans Angrier than Jazz Fans? A Genre-Wise Exploration of the Emotional Language of Music Listeners on Reddit, 2021