**SONIC**

**A MUSIC VIDEO STREAMING APPLICATION USING SEQUENCE ASSOCIATION RULE MINING AND USER-BASED COLLABORATIVE FILTERING**

A Project Study Presented to the Faculty of the College of Information,

Computer and Communications Technology

University of San Jose – Recoletos

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In Partial Fulfillment of the

Requirements for Thesis 1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ruziel Jon D. Mantalaba

Regil Kent V. Sembrino

*Members*

Mrs. Marissa M. Buctuanon

*Adviser*

**Chapter 1**

# INTRODUCTION

# 1. RATIONALE

Everyone listens to music almost every day. Through music, people’s feelings are expressed. The music that a person listens to affects his day to day performance. Some people don’t rely on what they hear, they want to use the sense of sight in appreciating the music in a form of music videos.

In our current time, the undeniably growing number of music available in the internet can be overwhelming and affects our simple act of listening to music. Thus, resulting to several problems such as information and cognitive overload on the user’s side. In additional, researchers also noticed that existing applications have cold start problems. This problem occurs when new users enter the system or new items are added to the catalogue. In such cases, neither the taste of the new users can be predicted nor can the new items be rated or purchased by the users leading to less accurate recommendations [1].

Recommendation can be defined as “the process of utilizing the opinions of a community of customers to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices” [2].

With the goal of providing a more personalized experience, this study aims to develop a hybrid recommender system that implements a combination of content-based collaborative filtering and sequence association rule mining. Hence, this study aims to build a recommender system that will provide fine-tuned recommendations that will reduces the user’s effort and time in making decisions.

# 2. THEORETICAL BACKGROUND

### 2.1 Recommender System

Recommender system is defined as a decision-making strategy for users under complex information environments [3]. A recommendation system is used to generate meaningful information to a collection of users for items or products that might interest the users.

Recommendation systems have become increasingly popular in recent years and are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. There are many ways the recommendation system works, but for this application, it implements a combination of sequence association rule mining and user-based collaborative filtering.

**2.2 Multi-Criteria Based Recommendation**

Multi-Criteria Based recommendation is implemented in this system to solve the cold-start problem. In order to have an initial genre preference of the user, factors are considered like age, personality type, location, and the user input preference per genre. This is to obtain a genre weight based on factors that are calculated as:

is the resulting genre weight. is the corresponding genre weight with respect to the value that was set under that criteria. is the percentage value for that certain criteria. Consequently, the weight of each genre will be assumed with the interpretation that the higher the weight value, the more likely user prefers this genre, and the smaller weight value, the lesser the user prefers this genre. These values are then computed with the videos with respect to their likes. Video weights will be calculated as:

corresponds to the likes gained by the video. is the obtained genre value weight. Thus, giving us which is the resulting video weight.

# 2.2 User-based Collaborative Filtering // TO BE REVISED BY REGIL

The profile of each user is exploited to detect similarities among users. Collaborative filtering methods are based on collecting and analyzing of data on users’ behaviors, activities or preferences and predicting what users will like based on their similarity to other users.

In user-based collaborative filtering algorithm, the ratings provided by similar users to a target user *A* are used to make recommendations for *A*. The predicted ratings of user *A* are computed as the weighted average values of these “peer group” ratings for each item [4].

User-based neighborhoods are made in order to identify similar users to the target user. In order to determine the neighborhood of the target user *A*, the user’s similarity to all the other users is computed.

One of the most popular similarity measures in Collaborative Filtering is the Cosine Similarity. Cosine similarity involves comparing user profiles. If one can compare whether any two objects are similar, one can use the similarity as a building block to achieve more complex tasks such as classification, clustering, and recommendation.

Mathematically, the cosine similarity is expressed by the following formula:

Where and are users and represents the rating of the video that was given by the user*.* Thus indicating and are the ratings given by user and respectively.

**2.2.3 Rating Prediction // TO BE REVISED BY REGIL**

Collaborative filtering recommender systems implements ratings-based user profiles in order to make item recommendations or predictions about user ratings for items. The rating prediction problem is a widely studied problem in the domain of rating-based recommender systems, where the goal is to predict the rating that a user would assign to an item that has not rated in the past [5]. Prediction of rating can be calculated as

where is the predicted user x’s rating to item i, N is the set users most similar to x who have rated item i, which is the ratings of the other users belonging to N, and is the similarity measure of user *x* and *y*.

# 2.3 Sequence Association Rule Mining

Sequence Association Rule Mining is procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases. It aims to find rules which enable us to predict the occurrence of a specific item based on the occurrences of the other items in the transaction.

# 2.3.1 Apriori Algorithm

Apriori Algorithm is a classical algorithm in data mining. It is used for mining frequent item sets and relevant association rules. It is also used to operate on a database containing a lot of transactions, for example, items bought by customers in a store. There are multiple rules possible even from a very small database, so in order to select the interesting ones, we use constraints on various measures of interest and significance. It will look at some of these useful measures such as support, confidence, lift and conviction:

1. Support – is the proportion of transaction in the database in which video V appears. It signifies the popularity of an itemset.
2. Confidence – it signifies the likelihood of video Y being viewed when video X is viewed.
3. Lift – it signifies like hood of the itemset being viewed when item Is viewed while taking into account the populate
4. Conviction – the conviction states by what factor the correctness of the rule (as expressed by its confidence) would reduce if the antecedent and the consequent of the rule were independent. A high value therefore means that the consequent depends strongly on the antecedent.

# 3. REVIEW OF RELATED STUDIES // TO BE REVISED BY RUZ

There are many video or music recommender systems that are existing in order to maintain the entertainment brought by the application. It creates an environment for expressing thoughts and emotions. The application aims to implement Sequential Rule Mining and User Based Collaborative Filtering for the application’s music video recommending approach. These following researches are related to this research and the ways and approaches will be applied to the application.

**3.1 A Review of Factors Affecting Music Recommender Success**

According to this study, it was seen that age, origin, occupation, socio-economic background, personality factors (introversion/extraversion), gender, musical education, attitude toward music, and familiarity with the music or style are useful for producing music recommendations. This study also stated that music preference is also influenced by what they believe about the music they hear and their familiarity with it.

**3.2 Recommender Systems: Issues, Challenges, and Research Opportunities**

This study is one of the basis of the researchers in explicitly asking the user’s preference in the registration part. According to this, a cold-start problem can be addressed in many ways. These are asking the user at the beginning to rate some items, asking the users explicitly to state their taste in aggregate, and suggesting items to the user based on the collected demographic information.

**3.3 Combining Content-based and Collaborative Filters in an Online Newspaper**.

For Content Based Collaborative Filtering, a research was conducted wherein the Content-Based and Collaborative Filtering were combined to produce a single prediction in an online newspaper. Combining content-based and collaborative filtering by basing a prediction on a weighted average of the content-based prediction and the collaborative prediction [6].

**3.4 Tapestry: Using Collaborative Filtering to Weave an Information**

Another is Tapestry, it is one of the earliest implementations of collaborative filtering-based recommender systems. This system relied on the explicit opinions of people from a close-knit community, such as an office workgroup.

**3.5 An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model**

This study presented a hybrid music recommender system that ranks musical pieces by comprehensively considering collaborative and content-based data, i.e., rating scores derived from users and acoustic features derived from audio signals [7].

**3.3 Sequential Patterns Analysis in a Student Database**

For Sequential Rule Mining, a paper presents a data mining methodology to analyze the careers of students, where a career can be seen as a sequence of exams. The model is built using sequential pattern analysis and uses the algorithm SPAM (Sequential Pattern Mining). It consider an ideal career corresponding to a student which has taken each examination just after the end of the corresponding course, without delays. The frequent patterns identified by the sequential pattern analysis are then compared with the career of the ideal student. The most interesting patterns are then used to refine the analysis by using clustering techniques [8].

# 4. PROJECT OBJECTIVE

This study aims to develop a music video streaming application. Specifically, it aims to:

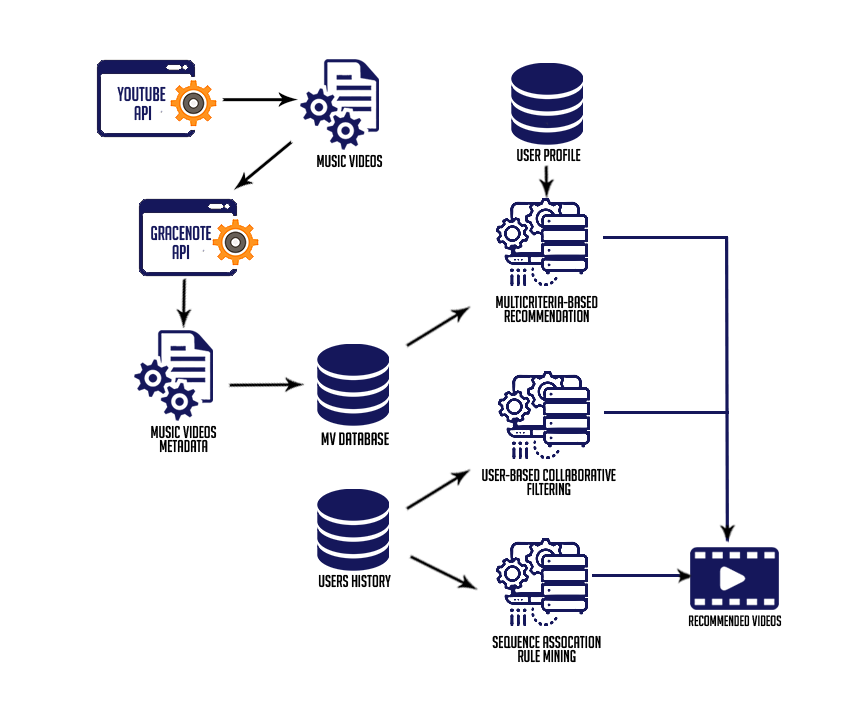
1. Produce a more personalized recommendation
2. Implements a hybrid recommender system that includes the following:
3. Sequence Association Rule Mining
4. Content Based Collaborative Filtering
5. Let user discover new music that matches his or her preference

# 5. SCOPE AND LIMITATION

This application gathers music videos only from Youtube using Youtube API. The metadata of the music are retrieved using Gracenote API. This application only runs on web with a J2EE backend. This application implements User-based Collaborative Filtering and Sequence Association Rule Mining.

# 6. RESEARCH METHODOLOGY

# 6.1 Conceptual Diagram



**Figure 1** Conceptual Framework

The system will use YouTube API for retrieving music videos and Gracenote API for the music metadata since YouTube can’t give information like genres and albums of the music. The retrieved data will be saved to the music video database.

In the user side, the user’s preference will be determined by inputting rating preference per genre, and other factors such as age and personality type. This will help the system to know what genres the user prefers. Afterwards, the system implements different recommendation techniques since this is a hybrid system. These processing techniques includes criteria-based recommendation, user-based filtering, and sequence association rule mining.

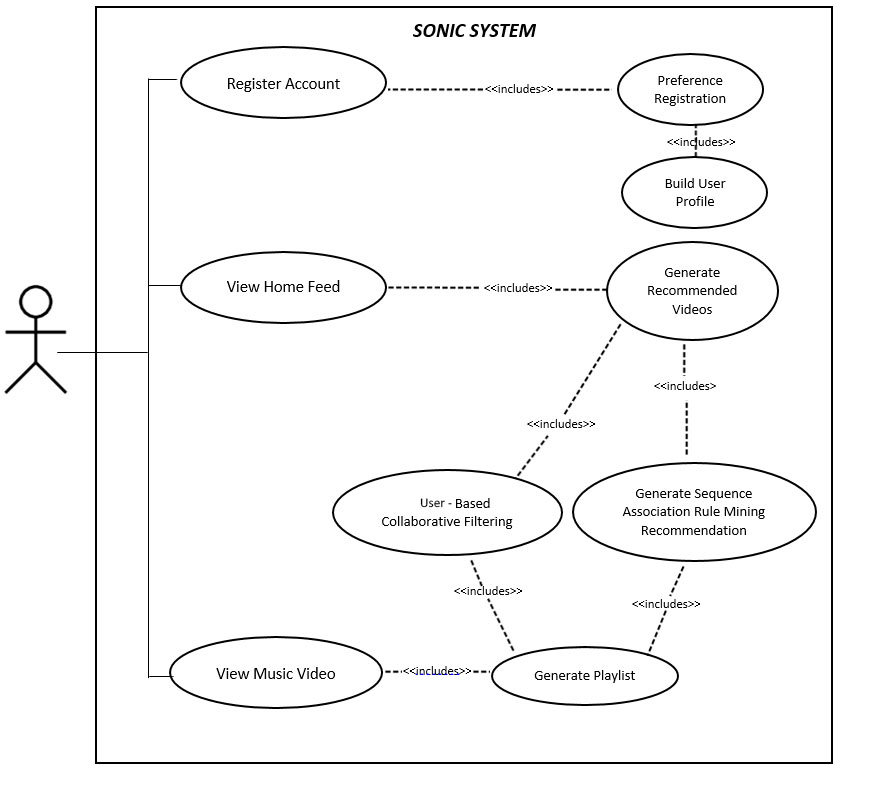
In Criteria-based recommendation, it is aimed that the system will be able to determine the possible genre preference of the user. The result is a list of ranked genres together with their corresponding score weight. These values will be used in computing the video score weight in relation to its number of views. Afterwards, User-based filtering and Sequence Association Rule Mining will follow.

**CHAPTER 2**

**SOFTWARE REQUIREMENTS AND DESIGN SPECIFICATIONS**

This chapter sets out the requirements that the system is expected to accomplish as well as the process of achieving those requirements. The section contains conceptual models for the design of the system that includes the Use Case Model, Use Case Scenarios, Activity Diagram and the Class Diagram.

**2.1 USE CASE DIAGRAM // TO BE REVISED BY RUZ**



**Figure 6** Use Case Diagram

The diagram above represents the use case diagram of the study. Upon registration; users are asked to input their personal information and their music preferences. After registering, YouTube API is being used to retrieve music videos and then generate recommended videos based from user’s music preferences. A playlist is generated after the user views more music videos.

**2.2 USE CASE NARRATIVE**

**// TO BE REVISED BY RUZ & REGIL \*FOR EVERY USECASE NAA DAPAT CORRESPONDING ACTIVITY DIAGRAM**

**Register Account**

|  |  |
| --- | --- |
| **Use Case** | Register Account |
| **Actors** | User, System |
| **Purpose** | System gets the information of the user and music preferences for recommendation of music videos. |
| **Overview** | After the system retrieves information, it will be saved to the database. The data will be used by the system to recommend music videos. |
| **Type** | Essential |
| **Pre-condition** | The user should be connected to the internet. |
| **Post-condition** | The system will create a user profile based on the data retrieved. |
| **Flow of Events** | |
| **Actor Action** | **System Response** |
| 1. The user registers. 2. The user inputs needed information. 3. The user inputs music preferences. | 1. The system will show the registration form. 2. The system will show the genre preferences registration. 3. The system will save the user’s information and preferences. 4. System will redirect the user to the home page. |

**Build User Profile (Genre Preference)**

|  |  |
| --- | --- |
| **Use Case** | Build user profile |
| **Actors** | System |
| **Purpose** | To determine user’s genre preference |
| **Overview** | The user’s data being saved are used in determining the personality of the user. |
| **Type** | Essential |
| **Pre-condition** | The user should complete the registration process |
| **Post-condition** | The system will determine user’s genre preference and ranked them according to score weight |
| **Flow of Events** |  |
| **Actor Action** | **System Response** |
|  | 1. The system will retrieve the user input preference 2. The system will compute each genre preference weight in relation to different criteria. |

**Preprocess Music Video Data**

|  |  |
| --- | --- |
| **Use Case** | Preprocess Music Video Data |
| **Actors** | Admin, System |
| **Purpose** | To get the Music Videos information |
| **Overview** | Music Videos are taken by using Youtube API and the metadata for a specific music are taken by using Gracenote API. |
| **Type** | Essential |
| **Pre-condition** |  |
| **Post-condition** | All music videos saved in the database should have their corresponding metadata. |
| **Flow of Events** |  |
| **Actor Action** | **System Response** |
| 1. The admin will send request to the Youtube API. 2. The admin will send request to the Gracenote API. | 1. The system receives the data. 2. The system will save the video ID, music video title, and the link for the thumbnail. 3. The system will save the music metadata to the database. |

**Compute User-based Collaborative Recommendation**

|  |  |
| --- | --- |
| **Use Case** | Compute Content-based Collaborative Recommendation |
| **Actors** | System |
| **Purpose** | To be able to recommend music videos based on music data and user’s profile |
| **Overview** | Determine top recommended music video based on criteria’s (cosine similarity, user’s personality, and user history) |
| **Type** | Essential |
| **Pre-condition** | User profile and music metadata must exist |
| **Post-condition** | The system must compute the weight of the music videos, filter it according to the set threshold and rank it according to the highest weight. |
| **Flow of Events** |  |
| **Actor Action** | **System Response** |
| 1. The user watches a video. | 1. The system implement a user to user Pearson correlation and saves calculations. 2. The system retrieves similar users that passed the set threshold. 3. The system retrieves suggested music videos that passed the set threshold from the top similar users. 4. The system will save the data. |

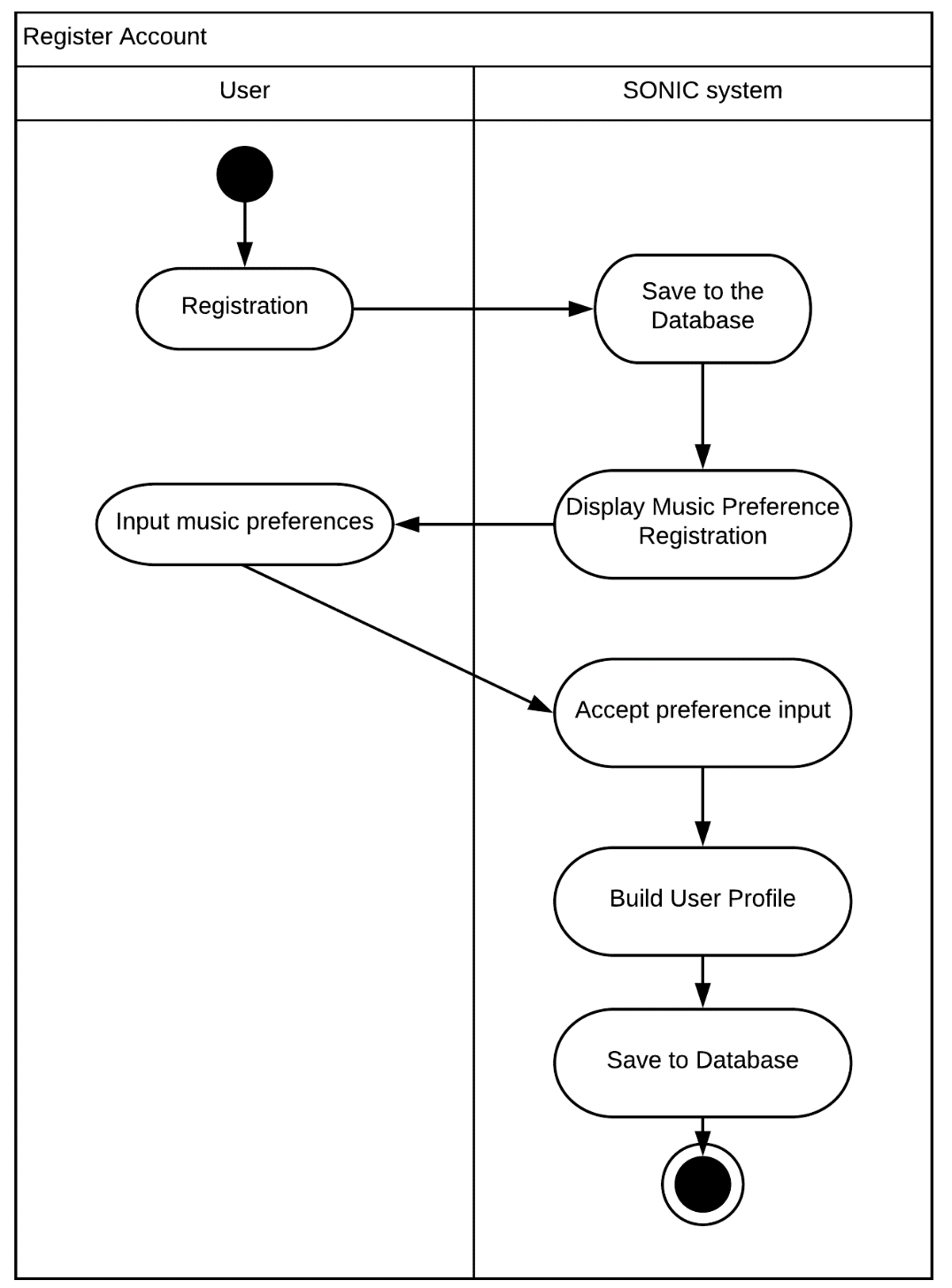
**Compute Sequence Association Rule Mining Recommendation**

|  |  |
| --- | --- |
| **Use Case** | Compute Sequence Association Rule Mining Recommendation |
| **Actors** | System |
| **Purpose** | To be able to recommend music videos based on music data and user’s watching history. |
| **Overview** | Determine top recommended music video based on several criteria’s |
| **Type** | Essential |
| **Pre-condition** | User watching history and music metadata must exist |
| **Post-condition** | The system must compute the weight of the music videos, filter it according to the set threshold and rank it according to the highest weight. |
| **Flow of Events** |  |
| **Actor Action** | **System Response** |
| 1. The user watches a video. | 1. The system will retrieve user’s watching history. 2. The system will implement sequence association rule mining. 3. The system retrieves suggested music videos that passed the set threshold from the top similar users. 4. The system will save the data. |

**Combine User-based Collaborative and Sequence Association Rule Mining Recommendation**

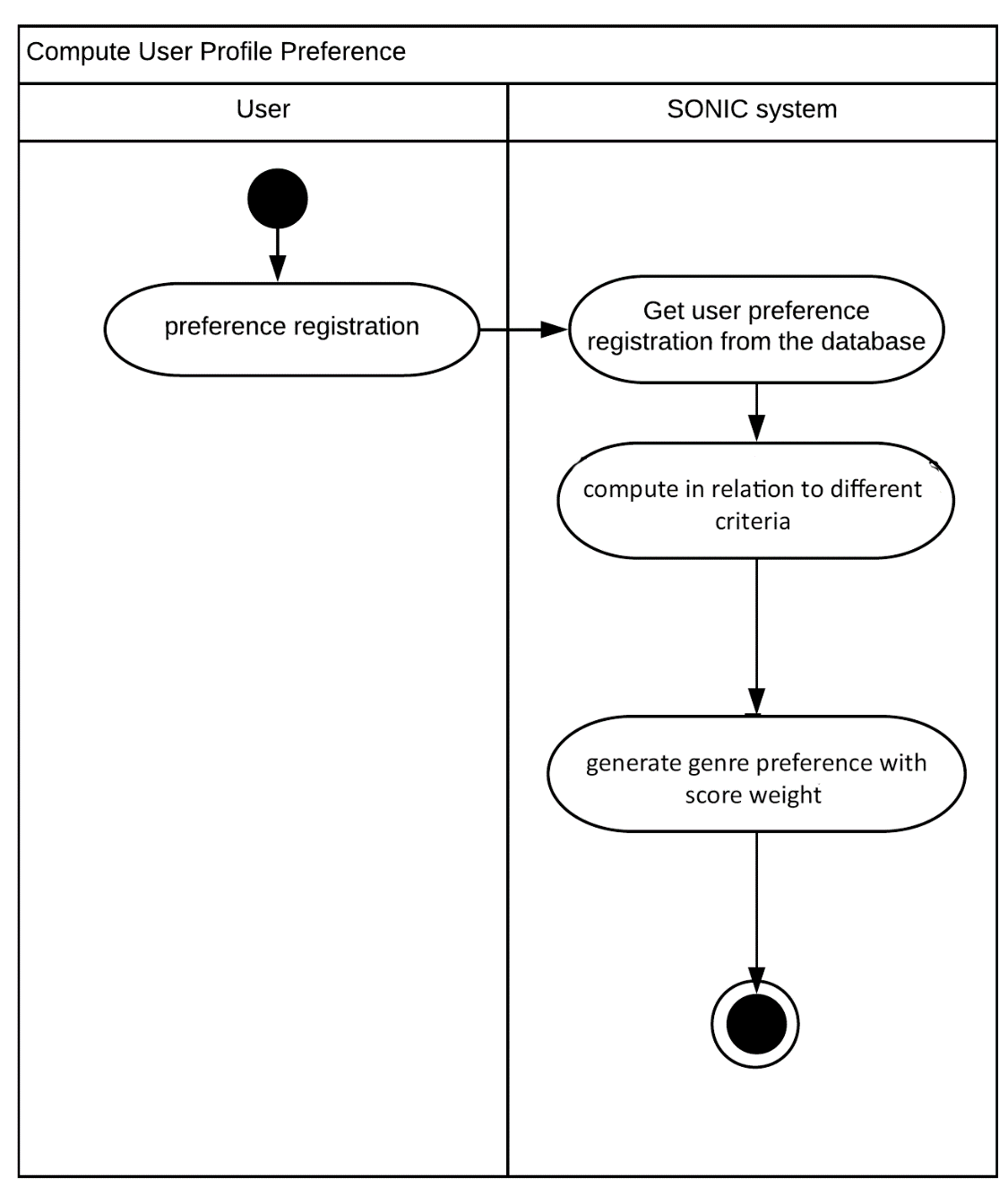
|  |  |
| --- | --- |
| **Use Case** | Combine User-based Collaborative and Sequence Association Rule Mining Recommendation |
| **Actors** | System |
| **Purpose** | To be able to recommend music videos based from the results of the User-based Collaborative and Sequence Association Rule Mining |
| **Overview** | Display’s top recommended music video from the two recommendation processing |
| **Type** | Essential |
| **Pre-condition** | Results from User-based Collaborative and Sequence Association Rule Mining |
| **Post-condition** | The system must compute the weight of the music videos, filter it according to the set threshold and rank it according to the highest weight. |
| **Flow of Events** |  |
| **Actor Action** | **System Response** |
|  | 1. The system will get the results from the two recommendation processing. 2. The system will evaluate the results for each videos. 3. The system will determine the top recommended music videos. |

**2.3 ACTIVITY DIAGRAM**

****

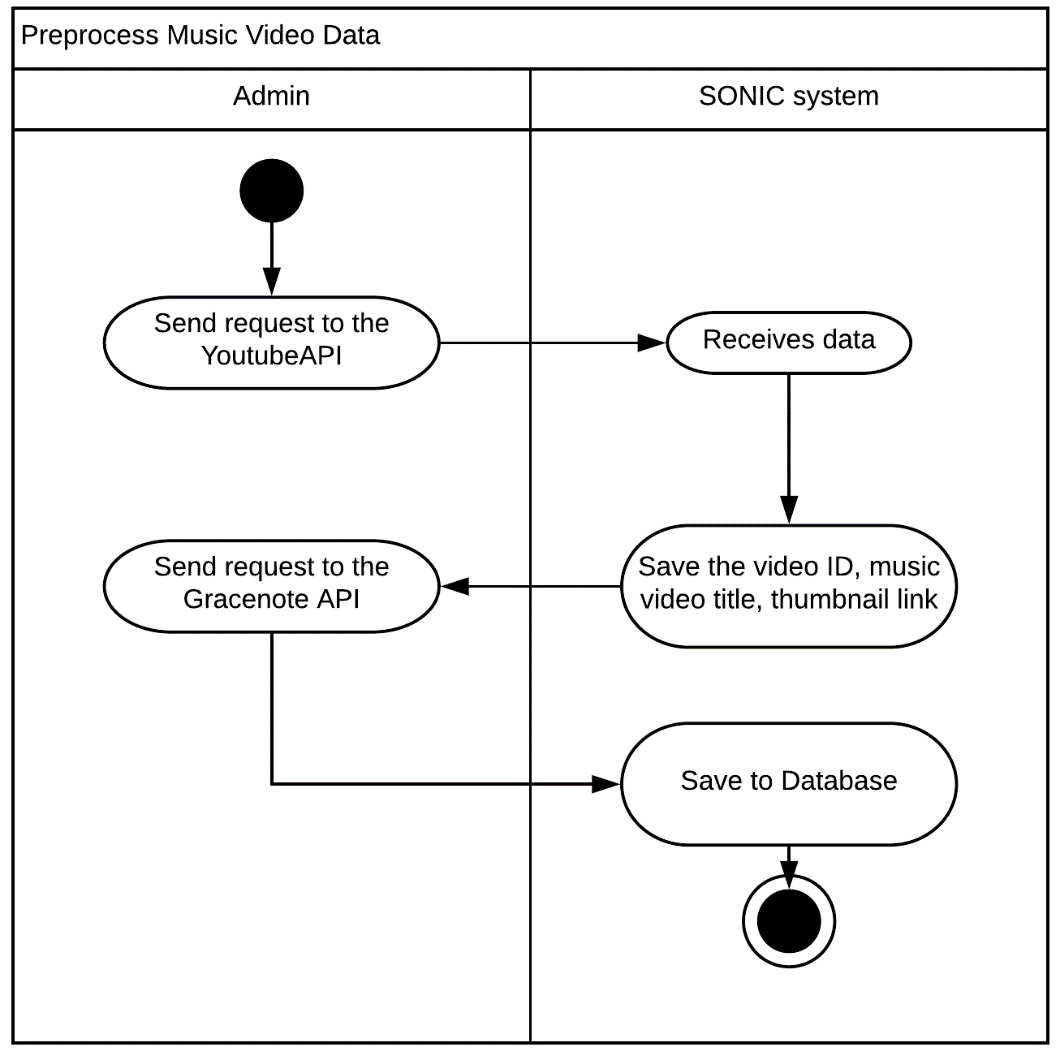
**Figure 7** Register Account

After the user registers his credentials, the user is asked to choose music genres that he prefers. Each genre has its own values to be computed in order to initialize a user profile with his genre weight points.



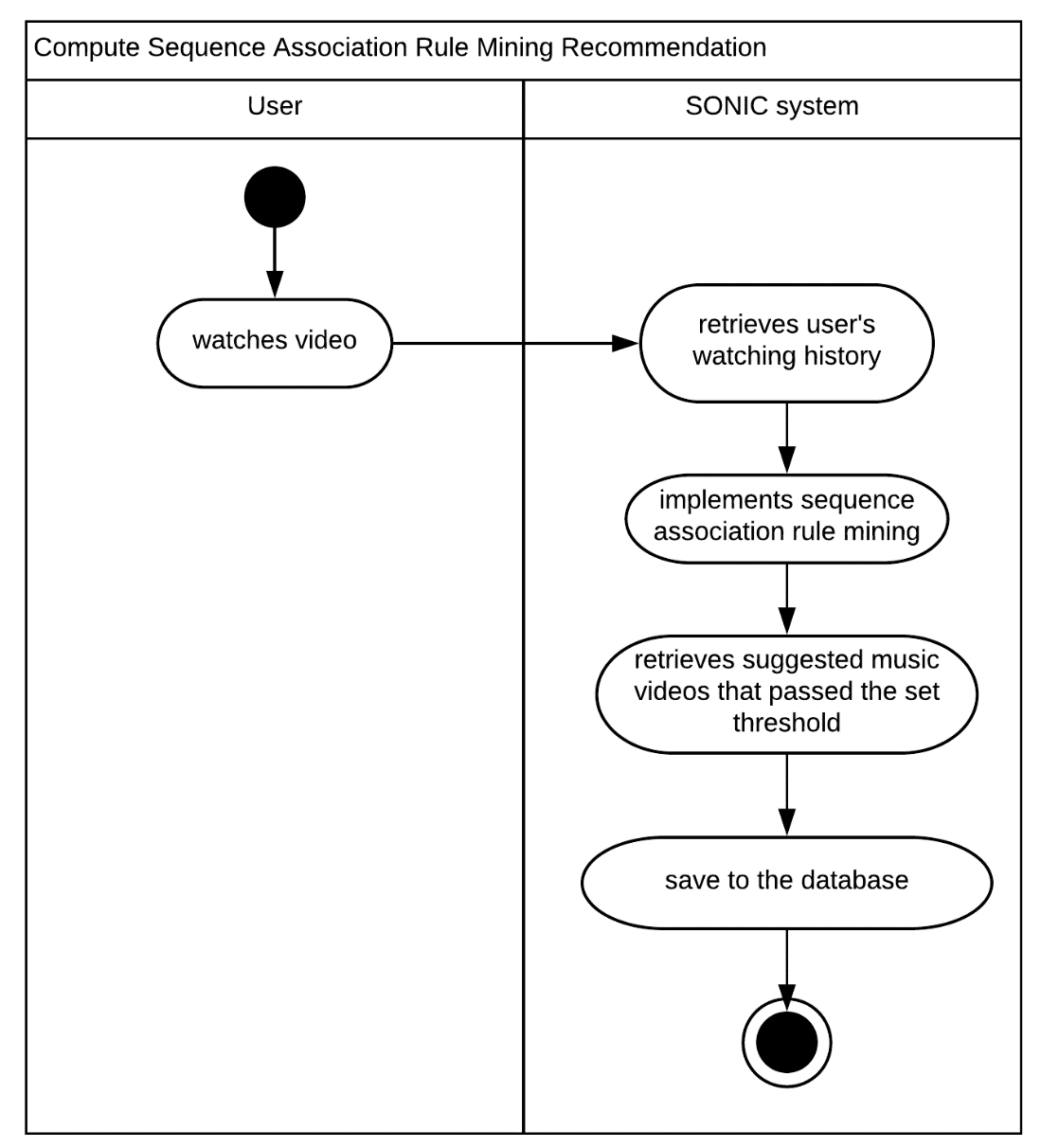
**Figure 8** Compute User Profile Preference

To know the user’s genre preference, it will be computed in relation to different criteria. This will help the system determine which videos are recommended to the user.

****

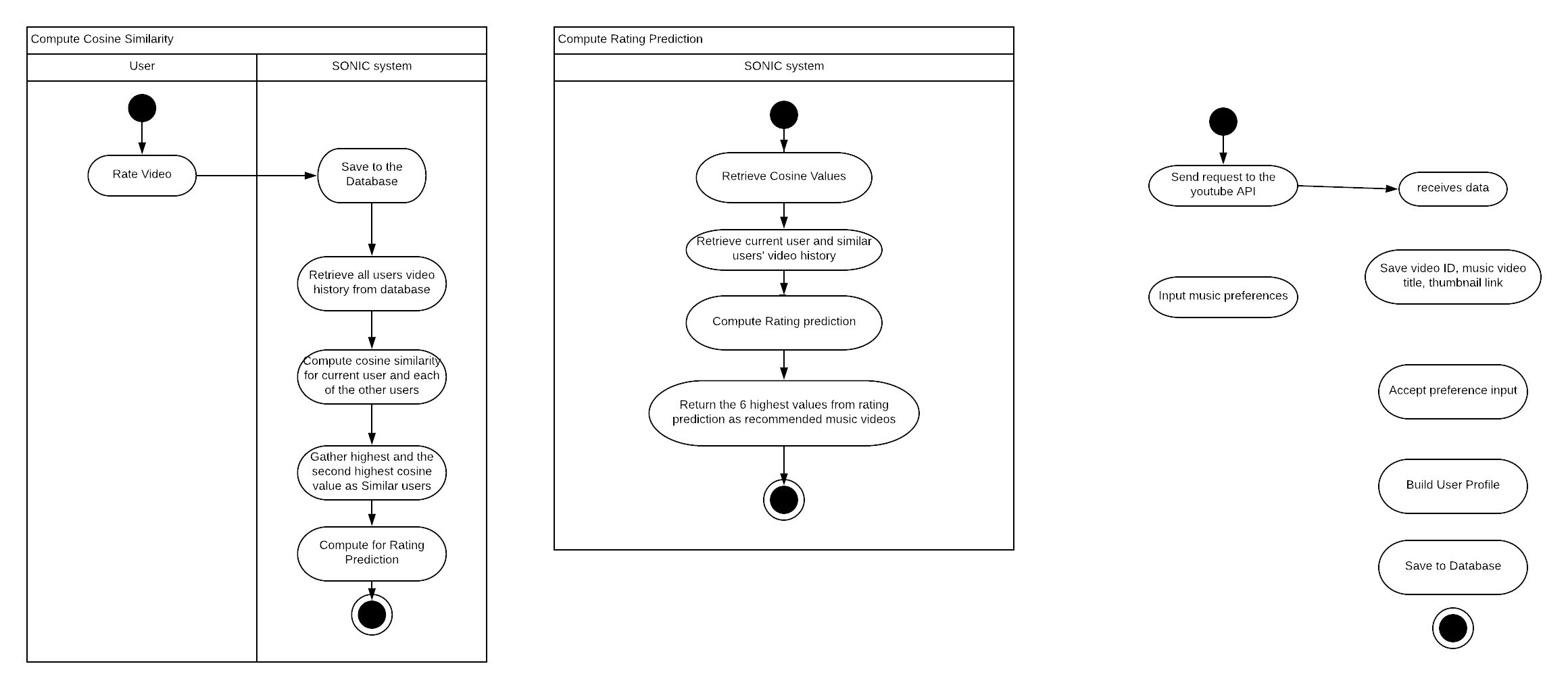
**Figure 9** Preprocess Music Video Data

The admin requests music videos by using YouTube API and then send it to the system. YouTube API doesn’t return much information so the system will save what the YouTube API returns and send request to the Gracenote API. The data returned by Gracenote API will be saved to the database.

****

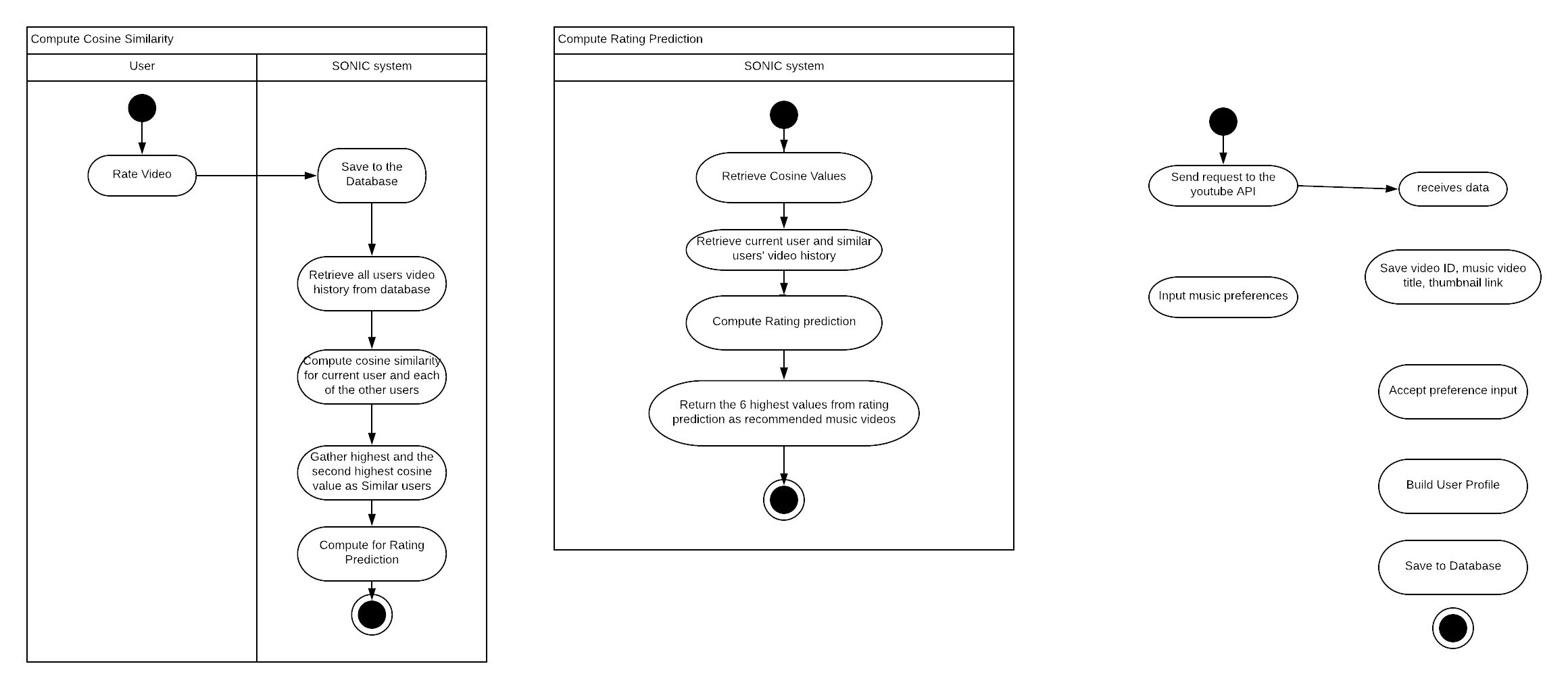
**Figure 11** Compute Sequence Association Rule Mining Recommendation

The system uses sequence association rule mining in finding the suggested music videos that passed the set threshold. It uses the user’s viewing history and evaluate it. The result of the process will be saved to the database.

****

**Figure 12** Compute Cosine Similarity

After the user rates the music video and then wishes to click on the other music video, the system will save the rating value of the previous music video to the database. The system then retrieves all the users’ video history including the rating given to the certain music video by the users and find the similar users to the current user by using the cosine similarity. Among all the cosine values, only 2 values starting with the highest value and then the following value will be used to compute for the rating prediction.



**Figure 13** Compute Rating Prediction

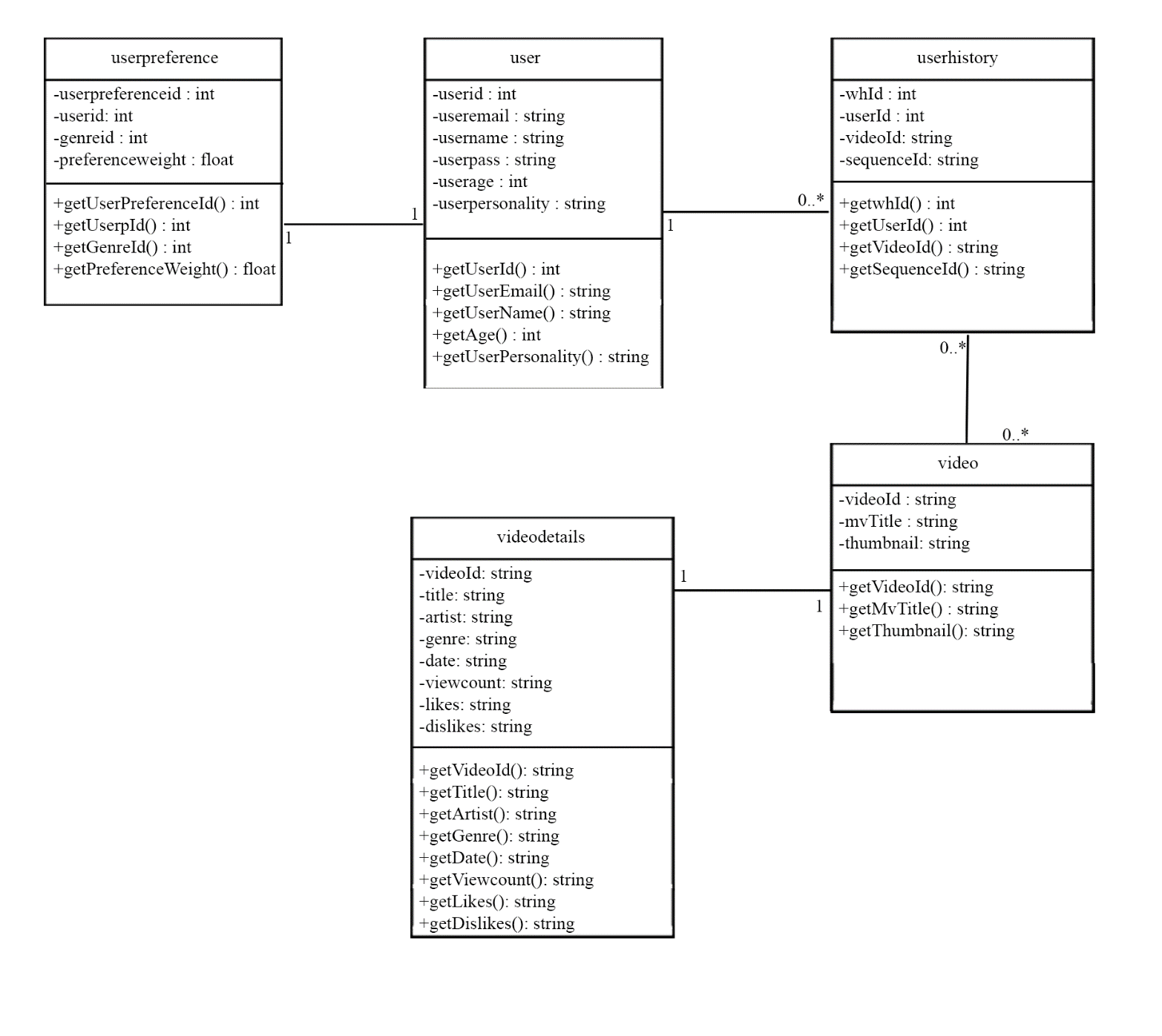
The system retrieves the cosine values of the similar users and retrieve the similar users’ watch history with the ratings of the music videos. Then the system will calculate for the rating prediction one similar user at a time. Only 6 highest values will be considered by the system, hence these values are the recommended music videos for the current user.

****

**Figure 14** Combine User-based Collaborative and Sequence Association Rule Mining Recommendation

The system will combine the results and then evaluate them. By using this process, the system can come up to more accurate results. After evaluating, the top results will be recommended to the user.

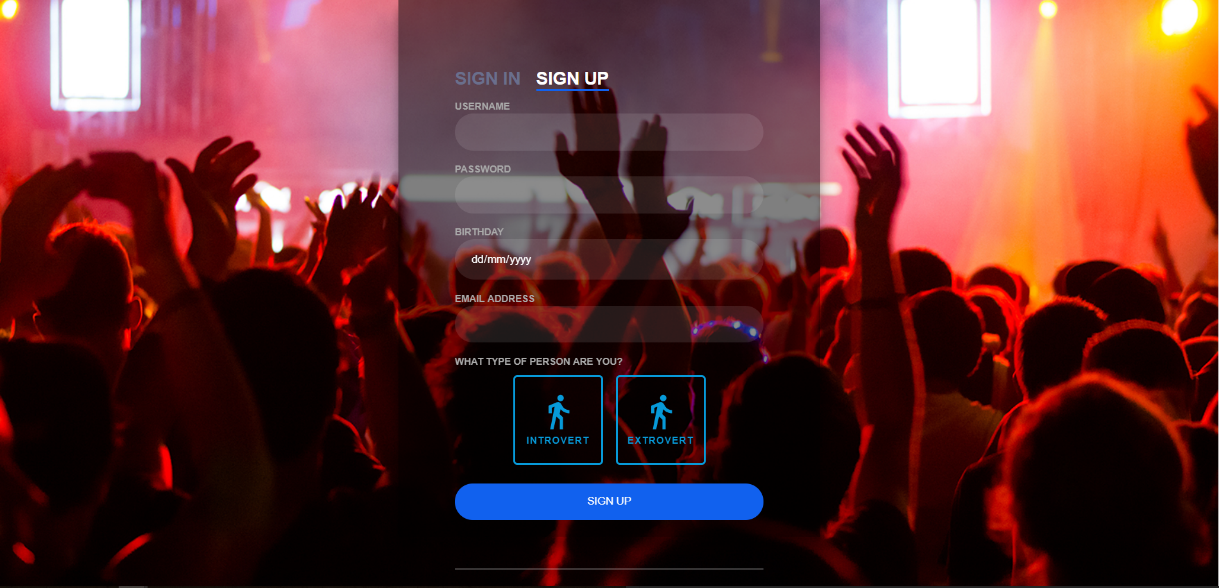
**2.4 CLASS DIAGRAM**



**Figure 14** Class Diagram

The diagram above shows the entire classes used in the system. Each class is composed of attributes needed for the class.

2.5 User Interface Desgin



**Figure n :** Sign Up Page

Figure n shows the Sign Up Page wherein user registers and input their basic information such as username, password, date of birth, email address, and personality type.



Figure n : Sign In Page

Figure n shows the Sign In Page where user logs in with the user of their registered username and password.



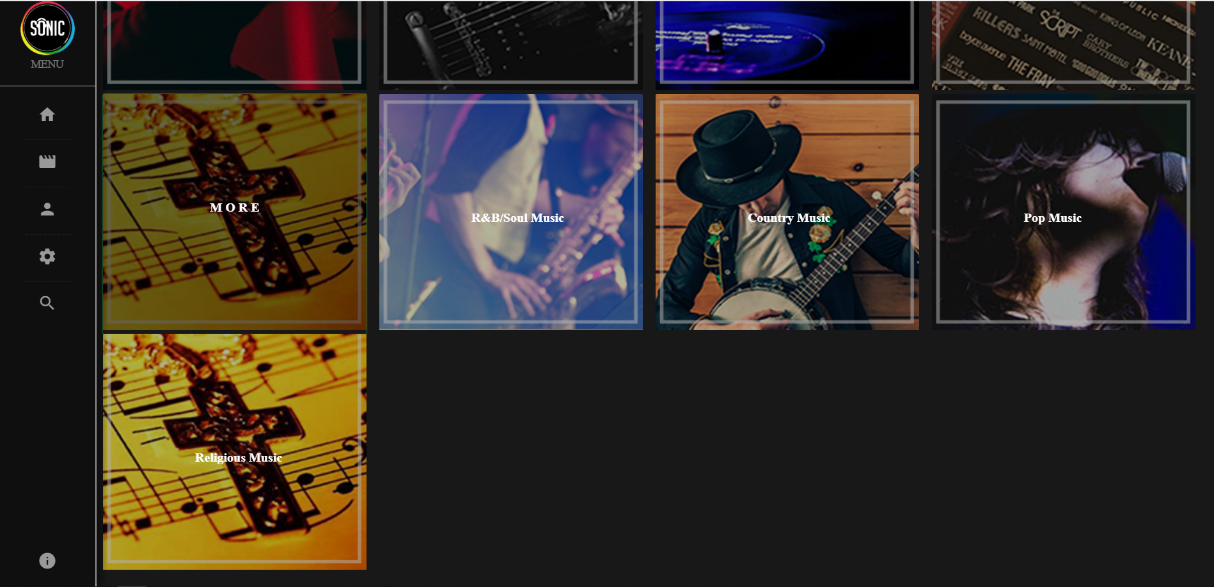
Figure n : Home Page

Figure n shows the Home Page wherein users can view several recommended videos. These videos are based on the result from the registration part of the user. The information gathered during the registration is processed through Multi-Criteria Based Recommendation.



Figure n : Explore Page

Figure n shows the Explore Page where users can view and discover more music videos. These music videos are categorized based on different genres.



**CHAPTER 3**

**SOFTWARE DEVELOPMENT AND TESTING**

This chapter includes the project development from development tools used, implementation of the different functionalities up to the testing processes made in the whole development process.

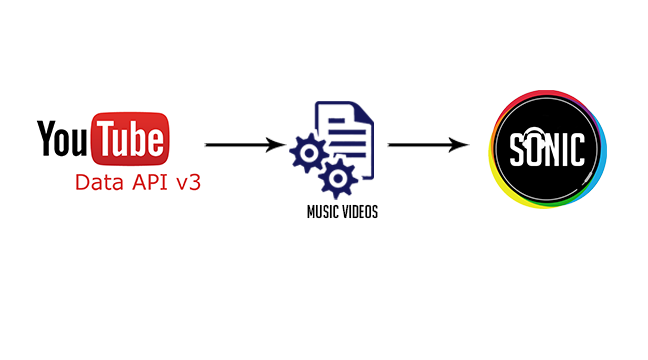
**3.1 DEVELOPMENT SOFTWARE PLATFORMS, DEVELOPMENT ENVIRONMENT AND TOOLS**

For the development process of the system, the researchers made use of the following languages, tools, and software:

* Java EE
* HTML, CSS, and Javascript
* Thymeleaf
* IntelliJ IDEA
* MySQL
* SQLyog
* XAMMP

**3.2 DEVELOPMENT PROCESS**

**3.2.1 YouTube API**

****

**Fig no**

Sonic requests Youtube API for music videos. Youtube API provides music videos depending on the query or the request being sent. In Sonic’s case, the music videos being provided includes the video ID, video title, and the video thumbnail link.

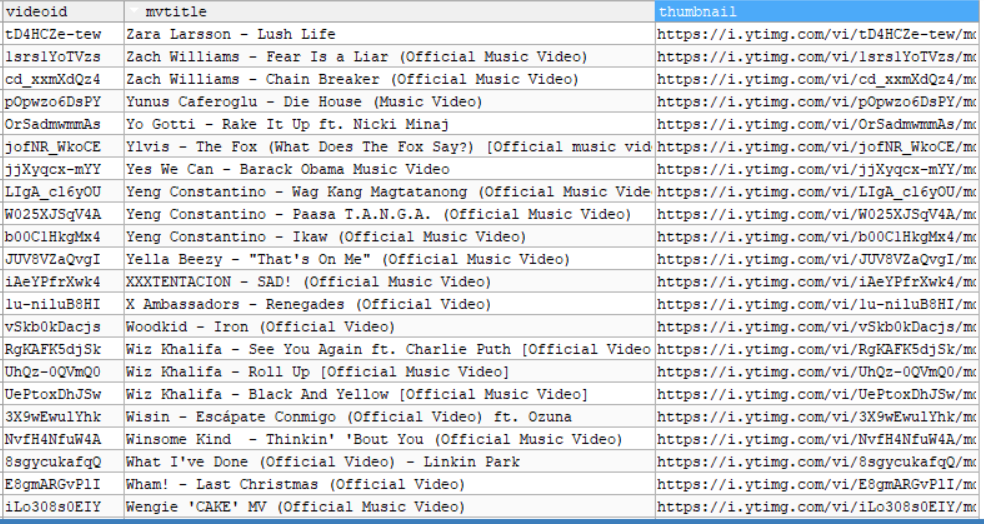


Figure n : Sample Data from YoutubeAPI

The figure above shows the sample data retrieved from the Youtube API. In order to fetch this data, the system will send a request to the API using the request function shown below.

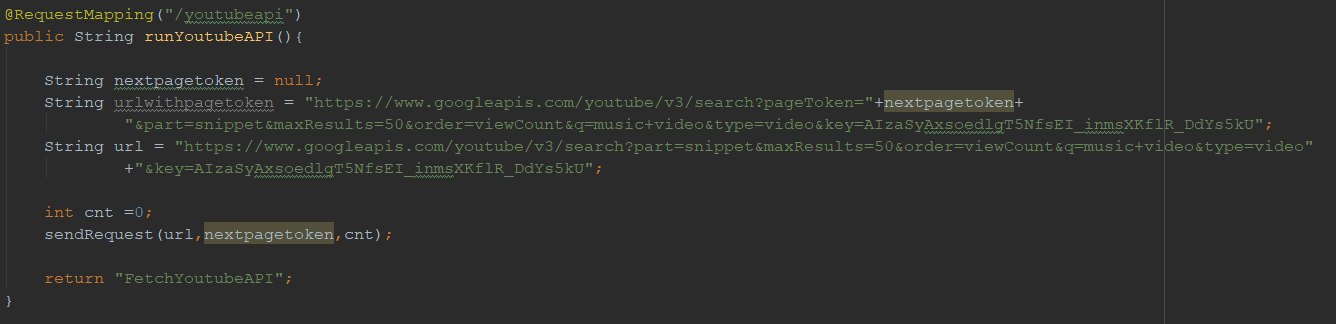


Figure n : YoutubeAPI Request Function

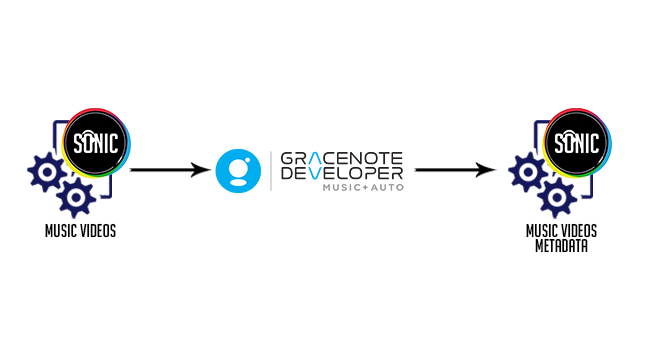
Figure n shows the request function for the Youtube API. This function includes the request link of the API that is sent to the sendRequest function shown below.



Fugure n : Send Request Function

Figure n shows the send request function that sends a request to Youtube API and extracts information from the JSON response.

**3.2.2 Gracenote API**

****

**Figure n :** Gracenote API

From the gathered music videos from YouTube API, Sonic sends request to Gracenote API for the metadata of the respective music. The data being used from the API are the artist name, genre, and date of release. These information are also combined with the data from YouTube which are the view count, likes, and dislikes. Afterwards, these data are being saved in one table in the database. Samples of data are shown in the figure below.



Figure n : Sample Music Metadata from Gracenote and YouTube

**//INSERT GRACENOTE CODE SNIPPET AND EXPLANATION**

**3.2.3 Multi-Criteria Based Recommendation**

In a Multi-Criteria Based Recommendation, multiple criteria are being considered in this process. These criteria depends on the admin configuration.

**//INSERT RUZ**

* + 1. **User Based Collaborative Filtering**

**//INSERT COLLABORATIVE FILTERING DEV PROCESS AND INFORMATION**

**//REGIL**

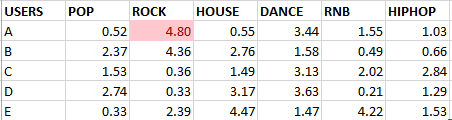
**// DATABASE, EXCEL SIMULATION**

**Bilbliography**

1. *Shah Khusro, Zafar Ali and Irfan Ullah , Recommender Systems: Issues, Challenges, and Research Opportunities*
2. *Resnick P and H.R. Varian, 1197. Recommender systems, Commun. ACM., 40: 56-58*
3. *Rashid AM, Albert I, Cosley D, Lam SK, McNee SM, Konstan JA et al. Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the international conference on intelligent user interfaces; 2002. p. 127–34.*
4. *Neighborhood-Based Collaborative Filtering – Springer, Reccomender Systems, The Textbook,Aggararwal, C.C. 2016*
5. *Desarkar, M. S., & Sarkar, S. (2012). Rating prediction using preference relations based matrix factorization. In UMAP Workshops.*
6. *Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., & Sartin, M. (1999). Combing content-based and collaborative filters in an online newspaper.*
7. *An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model , ) Yoshii, K; Goto, M; Komatani, K; Ogata, T; Okuno, HG, IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING (2008), 16(2): 435-447*
8. *Campagni, Renza & Merlini, D & Sprugnoli, Renzo. (2018). Sequential patterns analysis in a student database.*
9. *Algorithms for Big Data, Stefan Savev,* [*http://stefansavev.com/blog/cosine-similarity*](http://stefansavev.com/blog/cosine-similarity)

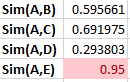
WAY LABOT PA

Every time the user watches music videos, it will then builds a user profile that is telling the genre preference of the user. The user may rate the music video in order to the system to update the user profile precisely. If the user doesn’t give rating to the music video, the system will use the rating prediction algorithm in order to predict the user’s rating to the unknown music videos.



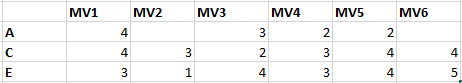
**Figure 2** Users with User Preference

The table above represents the user with its user preferences. In order to start the rating prediction algorithm, the system takes a user to measure their similarity using the Cosine Similarity. The highest value and the following 2nd highest value will be the basis for the rating prediction algorithm.



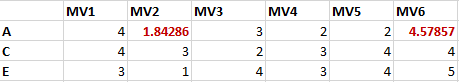
**Figure 3** Cosine Similarity results

As seen the diagram above, user E and C yields the highest and the 2nd highest result, respectively, which means they are the closest similar to user A. The system will look into the users’ video history focusing on the rating which the users give to the music video.



**Figure 4** Users histories with ratings

With the use of rating prediction algorithm, the music videos that are not rated by user A will now have corresponding values according to the formula being used.



**Figure 5** Rating Prediction results

The system will look for the highest value and the highest value will be recommended to the user.