AALBORG UNIVERSITY
COPENHAGEN

Semester: 4th

Aalborg University Copenhagen
A.C. Meyers Vænge 15
2450 København SV

Semester Coordinator: Henning Olesen

Secretary: Maiken Keller

Title: Addressing the Cold Start Problem in Music Recommendation Systems

Project Period: Fall 2015

Semester Theme: Master Thesis

Supervisor(s):

Niels Koefoed
Knud Erik Skouby

Members:
Arlind Thomas

Pages: 76
Finished: 11th of June, 2016

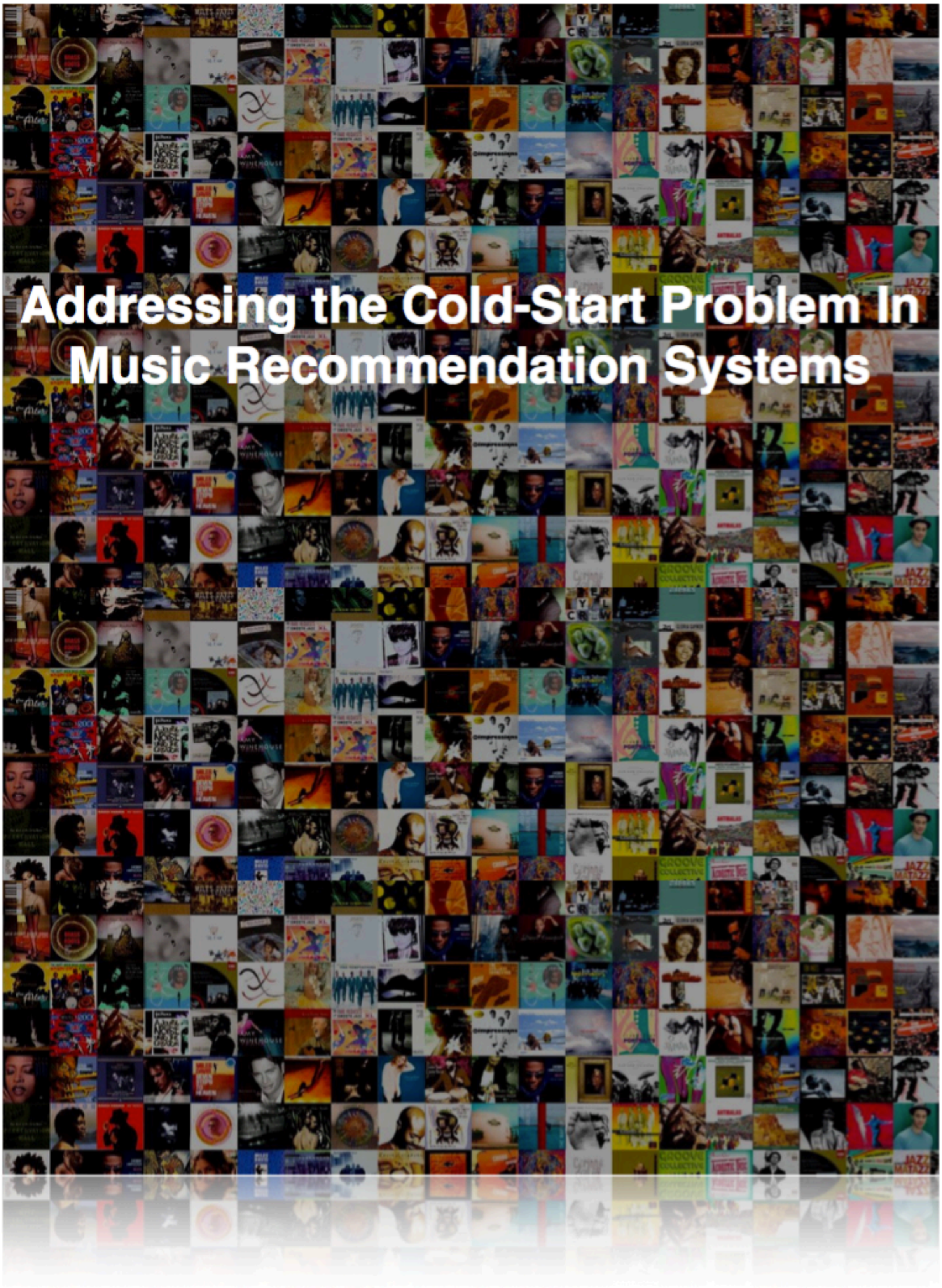
Abstract:

The digitization of music and the evolution to Web 2.0 led to the proliferation of several service providers that revolutionized music industry. These music platforms have utilized recommendation systems (MRS) in order to collect data from users' preferences in order to recommend new music they might want to listen to.

MRS's however suffer from the so called Cold-Start problem which refers to the inability of the system to recommend new tracks and/or artists because there is not enough information, such as user ratings associated with them.

The goal of the thesis is to address the Cold-Start problem by proposing a solution, which will be using SoundCloud API to recommend brand new tracks and/or artists.

This solution uses users' preferences shaped by already existing platforms (i.e. Spotify) and retrieves information only for new tracks and/or artists based on factors such as likes, shares, as well as comment, listening and favoriting counts.


Addressing the Cold-Start Problem In
Music Recommendation Systems

Contents

List of Figures

List of Tables

Acknowledgements

This master program has been a wonderful experience, as well as challenging from time to time. There are some people in particular who supported and guided me throughout this journey.

I would like to thank my supervisor, Niels Koefoed for giving me the chance to work with him on this interesting topic during the thesis period.

Second, I would like to thank Antonela & Stavris for helping, supporting and always pushing me forward when I was considering giving up.

I would also like to thank my family for all of their support and sacrifices they have made to help me be where I am now.

1 Introduction

The last decade, music consumption has changed dramatically. The digitization of music has had a huge impact in the music industry. The Internet evolution contributed also to this change because it was the source and distribution channel of digital music. As a result, a very large amount of music was accessible to users.

With the digitalization of music, another phenomenon aroused; music piracy and as a result, peer-to-peer file sharing. Peer-to-peer services such as Napster, changed people's behavior and interaction with music. Napster was a free music-sharing website that offered users the option to search, upload and download music files in a compressed digital format. This meant that almost anyone could download free music from the service and use it on their computers and portable devices. Over two years, Napster became extremely popular [1]. However, in 2000, multiple lawsuits were filed against Napster for copyright infringement, which was the reason that the U.S. Court of Appeals decided to shut it down [2].

With the rise of online piracy and declining of physical albums, Apple's iTunes were considered as the music industry's saving grace. Users could now purchase digital content legally. However, after iTunes had digitized the industry music, sales of downloaded music started to decrease from 2013 until at this day. The main reason for this decline were streaming music platforms, where people could easily stream entire music catalogue for a very cheap price without the need to download. These services range from modest systems which provide several playlists of media (prostopleer.com) to others that also offer recommendation of other tracks to the user (Spotify) and other more complex collaborative systems which gives the user the chance to interact with the system (Pandora, Last.fm). These platforms revolutionized music again from something that someone owns to something that someone can access from the cloud [3].

Music Recommendation Systems were part of this evolution and were developed in parallel with the Web 2.0. They play a vital role in the music industry and the academic community has had an extensive focus on them [3].

1.1 Background & Motivation

Recommendation Systems´ role is to provide users with the most relevant items based on their preferences or by their past evaluations and interactions. More specifically, *Recommendation Systems (RSs) are" software tools and techniques providing suggestions for items to be of use to a user. The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read"* [4]. Recommendation Systems work by collecting information on the preferences of its users for a set of items. This information can be acquired explicitly or implicitly [3].

- ***Implicit information*** is gathered from the user interaction with the service. For example, by collecting items related information such as viewing, playing times or users' ratings.
- ***Explicit information*** is when information is provided by the users themselves such as giving an opinion, rating or liking an item.

MRSs are an integral part of music service providers, which are constantly improving their recommendation techniques in order to generate as accurate recommendations as possible. A successful MRS should be able to automatically detect users' preferences and suggests music or generates playlists of tracks, they would probably like in the future [3]. However, there is still a lot that needs to be done in order for an MRS to provide the best recommendations to the users.

Music is a powerful communication and self-expression tool. It has a big impact on people's lives since it is an activity they engaged in frequently. Furthermore, according to a previous research [5], participants listen to music more often than engaging in other activities such as watching TV, movies or reading books.

Apart of being an enthusiast of both music and innovative technologies (in this case MRSs), my motivation is to provide a solution where users could get better and more accurate recommendations based on their music preferences.

This thesis was based on all the knowledge, skills and competencies that I have acquired from the previous semesters. Scientific methods and theories in a specific problem have been applied and discussed; as well as results were evaluated in an academic perspective. This thesis will also fit my chosen specialization, which is Service Development. Furthermore, this master thesis has been based in most of the modules including the lectures, classroom instructions, project works, workshops, exercises (individually and in groups) as well as teacher feedbacks and literature.

1.2 Research Question

According to Brian Whitman, which is the co-founder of music intelligence platform The Echo Nest, the problems with the MRSs is that they are not generally optimized to help the listener have a better musical experience and do not help new independent artists be discoverable to the public [6]. They are statistically optimizing to sell more items and consequently make more money. This means that new added music or new musicians are competing for an audience among millions of others trying just as hard.

"If there was something "intelligent" that could predict a song or artist to a person, both sides (musician and listener) win, music is amazing, there is a ton of data, and it's very far from solved" [6].

1.2.1 The Cold Start Problem

The most used methods in MRSs are Collaborative filtering and Content-based filtering techniques.

Collaborative filtering is the process where the system analyzes a user's preferences from his historical usage data. The system then recommends tracks based on other users that share the same music preferences with him. On the other hand, Content-based technique uses a song´s content and metadata such as artists, album and genre to recommend music [3]. CF usually performs better recommendations than CBF [7]. This is true only when we assume that there are already usage data such as previous tracks ratings. Otherwise, CF cannot work accurately and as a result, it will suffer from the well-known, Cold Start problem.

More specifically, Cold Start problem is divided into two categories: new items and new users. The first category refers to the issue caused by new items that are supposed to be recommended, but there is not enough information (such as ratings) associated with them. On the other hand, the new user problem happens when a new user joins a system, but little is known about him. As a result, the recommendation system cannot make personalized recommendation for him, unless he starts rating different items [8].

On the other hand, CBF is generally less sensitive to Cold Start problem, because it can still recommend items, even though it lacks ratings [9]. However, CBF suffers from other issues. Perhaps the biggest issue that occurs in CBF technique is that the system will recommend music that the user is already familiar [10].

The goal of this report is to propose a solution in order to improve the Cold Start problem in MRSs and offer users new music recommendations, which they have not heard before. This thesis will <u>only</u> focus on the Cold-start problem related to new items. More particularly, it will recommend new tracks and artists that the user has not listened to, based on their preferences using the already existing MRSs as well as information from other sources.

Therefore, this report will answer the below research questions:

1. *Which are the factors that can be used in MRSs in order to tackle the Cold Start problem related to the new items?*
2. *How could we utilize these factors in order to improve the overall recommendation experience in regards to the Cold-start problem?*

1.2.2 Project Delimitation

The purpose of this master thesis is not to create a new recommendation technique, but rather propose a solution that would take advantage of the current collaborative filtering technique used in MRSs and add extra information from other sources to help them improve their recommendations when it comes particularly to fresh music.

Also, the thesis is not intended to solve the Cold Start problem in its core. It will however improve the recommendations issues derived from The Cold Start problem such as the ability of the MRS to recommend new tracks and artist to the users.

Moreover, due to man power and time limited resources, I will implement only a small part of the system. The implementation will be a proof of concept in order to demonstrate that the potential service can be successful.

2 Methodology

This sector describes the process that was followed in order to answer the research question. After the problem formulation was defined and narrowed down to a specific research question, there was a need for a research approach to be designed in order to finally answer it.

As a first step, I started looking what was available in the literature on the concepts of music recommendation systems (MRSs), commercial music platforms and services, qualities and taxonomy of RSs. Second, I have conducted some interviews with individuals that are already familiar with MRS. Finally, the knowledge obtained from the interviews and the literature study will help me explain and argue about the assumptions, choices, and the decisions made during this work process.

2.1 Primary Research: Qualitative Study

This master thesis is relied in a qualitative research method. Qalitative research is used to have a better understanding of underlying reasons, opinions, and motivations. It provides insights into the problem and it dives deeper into it. In general, qualitative research is linked with a focus on processes and identified with detailed generation data methods such as semi structured and/or in depth interviews, among others [11]. Interviews are used here as a qualitative research tool because they generate rich data which can be analyzed in different ways [12].

During this thesis, four semi-structured interviews with users of MRSs were conducted. The chosen interviewees are four people from 25 to 37 years old and are relatively frequent users of MRS. All of them have a master education, and three of them in similar fields of ICT and Computer Science and one of them had also conducted a research in RSs. As such, the interviews were very insightful and valuable feedback were provided.

An interview is usually called “semi-structured” because it is closer to observation in comparison to the “structured” one that uses “closed” questions similar to a questionnaire. As such, questions were formed in a way that allowed helping me understand better the already existing problems of several music recommended systems. For a review on the actual questions used in the interviews, see the Appendix chapter.

Apart from interviews, in this thesis were also used personas and scenarios. Two personas were used, in order to gain an understanding of the users’ requirements and consequently improve the software design [13]. Respectively, two scenarios were also used in order to extract user requirements and to get a better understanding of the domain problem.

The feedback acquired from the interviews, scenarios and personas were a stepping-stone for implementing my proposed solution.

2.2 Secondary Research: Literature Study

Secondary sources are important resources for collecting data related to the topic of the research question. I used different sources for the research study such as the tech-blogs in the Internet, books and various research papers.

During the search process, I used different key words, such as music data, music recommendation systems, Cold-start problem, Collaborative filtering, Spotify etc. These keywords were used in different combinations to find relevant literature for explaining and understanding better various MRSs such as Spotify, Last.fm, iTunes & iTunes Apple Music and Pandora in terms of different factors they use to shape users’ preferences and the way they may be used to address the Cold-start problem.

2.3 Analysis

The next step is to analyze the data gathered from the primary and secondary research including interviews, personas and scenarios in a combination with the literature review. Therefore, I have to make sense of the data that I have collected by exploring and interpreting them. The analysis played an important role in understanding MRS problems and as such developing my new solution that would address the Cold-start problem.

2.4 Software Development Method

The selected method for the implementation of my solution is the Waterfall method. The Waterfall method consists of several phases where each of them should be fully finalized in order for the next phase to initiate. This method was used in my thesis because it is quite simple to understand and it can be easily managed. In this model, phases are processed and completed one at a time [42].

3 State-of-the-Art

3.1 Recommendation Techniques

Music recommendation systems are based on several filtering techniques. However, only the below approaches will be covered in this chapter: Collaborative, Content-based, Demographic as well as Hybrid Filtering Techniques.

The second part of this chapter will describe the commercial music services that have employed recommendation systems into their platform.

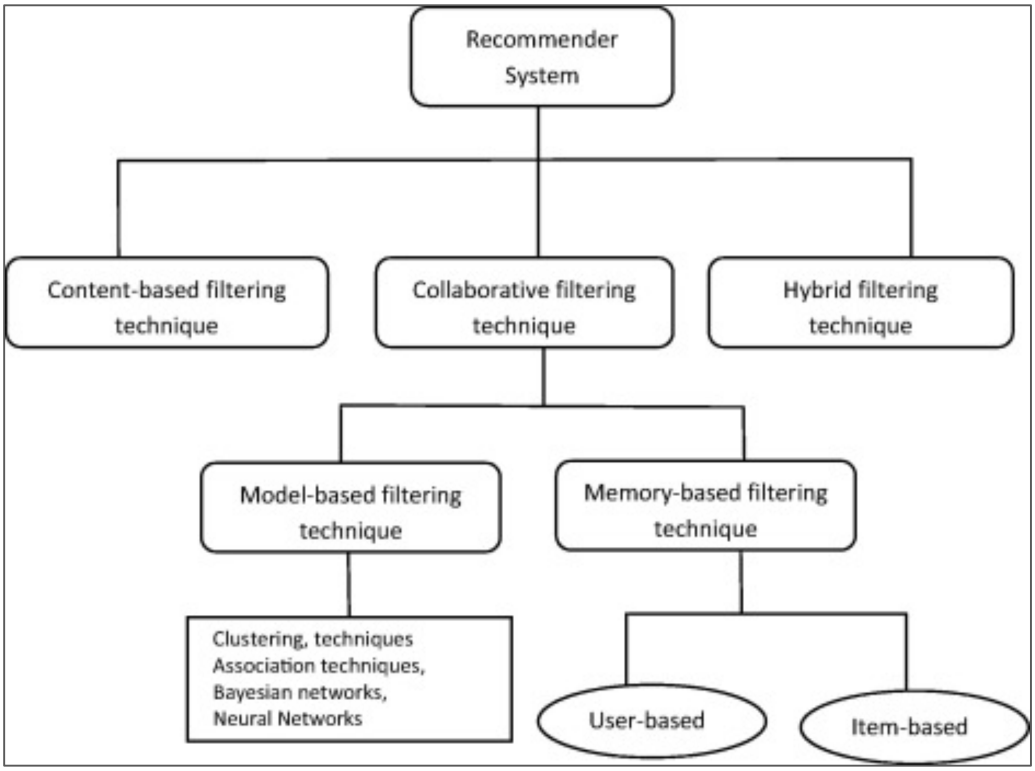


Figure 1: Recommendation Techniques [14]

3.1.1 Collaborative Filtering Technique

This technique is considered to be one of the most significant and widely used in recommendation systems. According to Dunning and Friedman [15], *"a collaborative filtering algorithm usually works by searching a large group of people and finding a smaller set with tastes similar to yours."* It is predicting what users will like based on their similarity to other users. Therefore, for example, if user X and Y rate "m" items similarly, they would probably rate other items similarly as well. Collaborative Technique is divided by two categories: Memory-based and model-based.

Memory-based CF can be categorized in two ways: user-based and item-based techniques.

- *User-based* collaborative filtering works by finding other users whose ratings behavior is similar to the current users and use their ratings on other items to predict what the user will like in the future.
- *Item-based* collaborative filtering on the other hand calculate predictions using the similarity between items and not users. If for example, two items tend to have the preferences from the same users, then they are similar and users are expected to have similar preferences for similar items.

Model-Based CF make recommendations by first employing a model consisting of user ratings. This model building process is done by using data mining as well as machine learning algorithms such as Bayesian network, clustering, and rule-based approaches. This technique works by building a model based on a dataset, which consist of ratings. Then information is extracted from the dataset, which is used as a model to make recommendations without using the complete dataset each time [3].

3.1.1.1 Limitations

Collaborative filtering approaches are content-agnostic [16]. This means that any kind of information about the items that are being recommended is not used at all since they rely only in usage data. This creates a number of problem listed below:

- Data sparsity is a term use to describe the above-mentioned problem when it comes to large databases. This problem occurs in cases when the recommendation generates weak recommendations because users rate a small number of items in the database. This means that the bigger the number of users or items is to a database, the more-likely less rated items by the users are, especially when it comes to User-Based CF.
- The cold-start problem refers to the fact that some users or items have no data associated with them. As a result, the recommendation cannot make relevant predictions. This problem affects particularly the profile of new users and items. This is a big disadvantage when it comes to MRSs meaning that the recommendation system recommends mostly popular music rather than music from artists that they have never heard before.
- Another drawback is a term known as the Gray Sheep. This problem refers to the users that have taste different then the majority of the other users and thus will have poor recommendations [17].

3.1.2 Content-based Filtering Technique

CB Filtering is also a widely used technique. The system works by recommending items that are similar to the ones that the user has positively rated in the past [4]. In a content-based recommendation system, keywords or attributes are used to describe an item. Items are ranked by how closely they match the user attribute profile, and the best matches are recommended. For example, if a user has positively rated a track that belongs to the rock genre, then the system recommends other tracks from this genre. There are many different kinds of information associated with music that could aid recommendation: tags, artist and album information, lyrics, text mined from the web (reviews, interviews etc.), and the audio signal itself. To make their recommendation better, Spotify acquired the music intelligence company called The Echo Nest. A track consists of different kinds of information such as artist, track and album title, genre, release year, lyrics as well as its audio. The perfect example of content-based RS is the popular streaming service named Pandora Radio, which plays music according to the user feedback [3].

3.1.2.1 Limitation

The main advantage of Content-based compared to the Collaborative Filtering is that the system can still recommend items, even though there is no information taken by other users in the database. However, this technique has several disadvantages.

- Perhaps the biggest problem is the cold-start problem associated with new users. The system cannot recognize user's preferences, so as a result cannot recommend music to them.
- CB Recommendation are also affected by the Gray-sheep problem according to his collection or is biased to a particular genre.
- The novelty factor is also another issue. Since the user receive recommendations, which are very similar to his profile, this could decrease his satisfaction. The reason is that the user could probably know those items and want other recommendation, which are not familiar to them instead. Another term related to novelty, is serendipity or otherwise unexpected or fortunate discoveries.

- Another limitation is that subjectivity or personal opinions is not taken into account when recommending items to users, because the recommendation is focused only on finding items that are similar by only their features [17].

3.1.3 Demographic-based

Demographic filtering is a technique used to recommend items based on user's demographic profile. The system works by identifying what kind of users like a certain item. In demographic filtering, it is assumed that the users with common demographics such as age, gender, country etc. will also have same tastes and preferences [18].

3.1.3.1 Limitation

- Since this technique recommends the same items to people with same demographics, recommendations can be too general and not specific for a particular user.
- Building the profile is also another issue. Some approaches try to get (unstructured) information from user's webpages, weblogs, etc. In this case, text classification techniques are used to create the clusters, and classify the users [19].

3.1.4 Hybrid-based

The hybrid method is a combination of the previous approaches with the purpose of avoiding the above-mentioned limitations and delivering better results. Most commonly, collaborative filtering is combined with some other technique in an attempt to avoid the ramp-up problem. Some of the methods as introduced by Burke on 2002 [20] are:

- *Weighted hybrid:* The results from different techniques are combined together to generate a single better recommendation.
- *Switching hybrid:* The system selects one recommendation technique among others depending on the current situation.
- *Mixed hybrid:* The system uses multiple techniques to present multiple recommendations.
- *Feature combination hybrid:* A single recommendation technique uses features from other combined recommendation data sources.
- *Cascade hybrid:* The system uses multiple recommendation techniques, which have different priorities.
- *Feature augmentation hybrid:* The system uses a recommendation technique to compute a set of features, which are later used as an input for another one.
- *Meta-level hybrid:* A model produced by a recommendation technique, is used as an input for another one.

3.1.5 Community-based

This type of system recommends items based on the preferences of the users' friends. This quote is often used to describe this technique: "Tell me who your friends are, and I will tell you who you are". [4] Today social networking sites are very popular and people can share similar interests or may rate content similarly, so Recommendation Systems can take advantage of it. The current recommendation system used at Spotify is based solely on artist data. Using data about related artists from All Music Guide and combining this with history of played tracks, other artists are recommended.

3.2 Commercial Music Platforms

There are several music platforms on the internet that help users discover music based on their music preferences. Most of them use complex recommendation algorithms incorporated into their platform. The most popular ones are Pandora, Last.fm, Spotify as well as iTunes and Apple Music. This section will describe the way these platforms use their recommendation techniques to offer recommendations to its users.

3.2.1 Last FM

Last.fm is one of the most popular online music and social community platform that provide users with personalized music recommendations. Last.fm is based on collaborative filtering technique. It recommends music by observing users' listening habits and comparing those against the listening behavior of other users. As Last.fm music does not contain its own catalogue, it pulls in music from multiple sources such as Spotify or YouTube.

In order to use the platform, users first have to sign up. Last.fm next step is to gradually build the user's profile by providing music recommendations based on user's listening behavior. In order to accomplish this, Last.fm uses a tool referred as scrobbling. Scrobblers are simply metadata for every track that the user listens to such as track name, artist name, album name, as well as timestamp (the time the track starts playing) [21]. Below is an example of my first scrobbler:

- *Artist: David Bowie*
- *Track: Changes*
- *Album: Hunky Dory*
- *Timestamp: 3 May, 3:03pm*

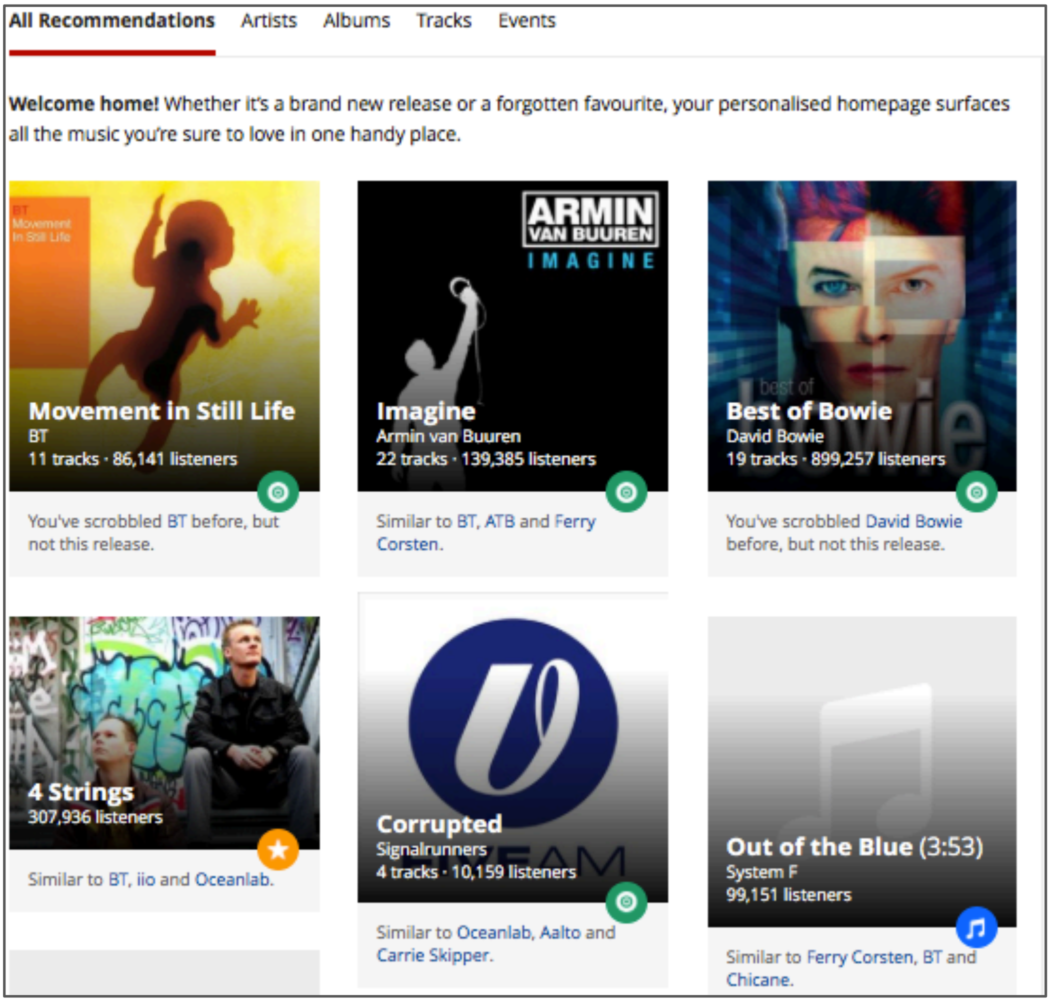


Figure 2: Last.fm Recommendations Screenshot

Scrobbling is the action of sending those data to Last.fm's database, which later are being displayed on the user's profile page. These data help Last.fm create recommendations to help the user discover new music. Scrobbling tool is built into various music services such as Spotify, Deezer and Rdio. Last.fm has also built software that is compatible with Mac OSX, Windows, iOS as well as Android. All these data from million users combined are fundamental in order to help Last.fm increase accurate recommendation accuracy [22].

As seen in the Figure 2, Last.fm offers recommendation based on Artists Albums and Tracks. When for example, the user clicks in particular artist, other options are displayed. Some of them are Top tracks and albums, similar artists, tags, and biography.

Last.fm Library has a dozen other features. Explorer feature that provide the user with a list of music that they had listened in the past. Moreover, Last.fm "Live" feature uses innovative interactive widgets that let users visualize in real-time the listening habits and trends of the global Last.fm community. Users can either listen to music that other users listening at the moment, music that is trending as well as now music not heard before [23].

3.2.2 Pandora Radio

Pandora is a popular Internet radio and automated music recommendation service on the Internet that bases its recommendations on the "Music Genome Project", where highly-trained musicians employed by Pandora work with the purpose of *"capturing the complex musical DNA of songs"* [24].

Pandora Radio uses the properties of a track or artist, which consist of as many as 400 attributes in order to generate a station that plays music with similar properties. This attributes are provided from the Genome Project. More specifically, the staff at the Genome Project analyze every new track that comes out, examine a list of possible attributes associated with it and finally assigns the track a numerical rating [25]. The purpose of the Music Genome Project is to make predictions as accurate possible about the kind of music, a user would like next. Additionally, Pandora has a thumb up and down button when users can like or dislike a particular track. This way the user can emphasize or deemphasize certain attributes of a track, which as a result makes its recommendations even better [26].

Pandora Radio is an example of a content-based recommendation system, which plays music that is similar to the ones that the user provided as an initial seed or positively rated. There quite a few features that make Pandora a great music recommendation system.

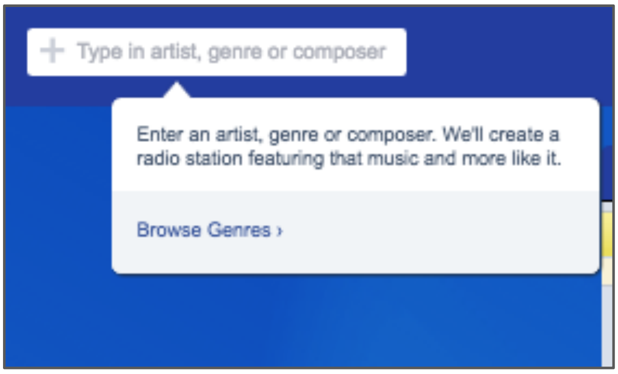


Figure 3: Pandora's Way of Creating a Radio Station Based on User's Input

The first time the user accesses Pandora website, a small window with a search bar is being displayed, where the user is being asked to enter either an artist, composer or a genre. Pandora then creates a music station that matches the user's input. This is done right away, even though the user has not registered yet. However, for better recommendation a user has to register for the system to offer better recommendations.

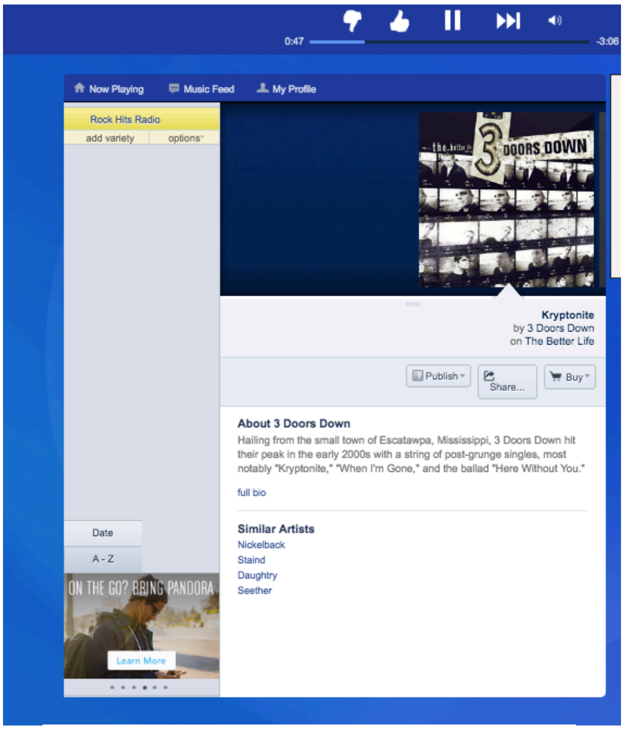


Figure 4: Pandora Radio Station

"Create a Station" search bar feature is always located in the top left corner of the website. A user can create radio stations based on the artist, track as well as composer. In this case, I've searched for Moby and the search box shows me results about artists or tracks where I can choose from.

The user has the option to explore artists or tracss by clicking on the "Search for". He can also choose among a big list of genre categories to create stations that matches that genre category. Moreover, Thumbprint Radio feature can play tracks that are positively rated (thumbs up) from all your stations, as well as other tracks related to those thumbs [27].

3.2.3 Music Recommendation System Used in iTunes & Apple Music

iTunes is a popular media player developed by Apple where users can play, organize and buy digital music and videos. iTunes has been using Genius, which is considered one of the most popular music recommendation systems.

Genius recommendation system uses Gracenote MusicID, a music recognition technology for audio identification and fingerprinting, which provides descriptive metadata that *"defines the detailed musical characteristics of songs and artists including Genres, Mood, Era, Origin, Tempo and Artist Type" [28].* These metadata are used to identify tracks in the Genius database. When enabled, Genius starts to analyze the user's music library. Genius then uses collaborative filtering technique to compare the tracks΄ metadata from the user's music library to the huge database of music sales, play history and ratings data gathered from iTunes users in order to generate playlists recommendations [29]

In 2015, Apple launched Apple Music, a subscription-based music streaming service where users can stream entire music catalogue from Apple. Apple Music has three main features: [30] Beats 1, an internet radio station, which broadcasts live music worldwide; Connect, a social networking feature for artists to share content with fans such as photos, videos as well as snippets of new tracks; Curated Playlists, playlists recommendations based on users΄ preferences which according to Apple, they are not built by algorithms, but by real people [31].

More specifically, Apple Music is based on Genius Recommendation which iTunes has already been using the last years. However, Apple Music is built in with more features in order to better understand the users' listening habit and preferences

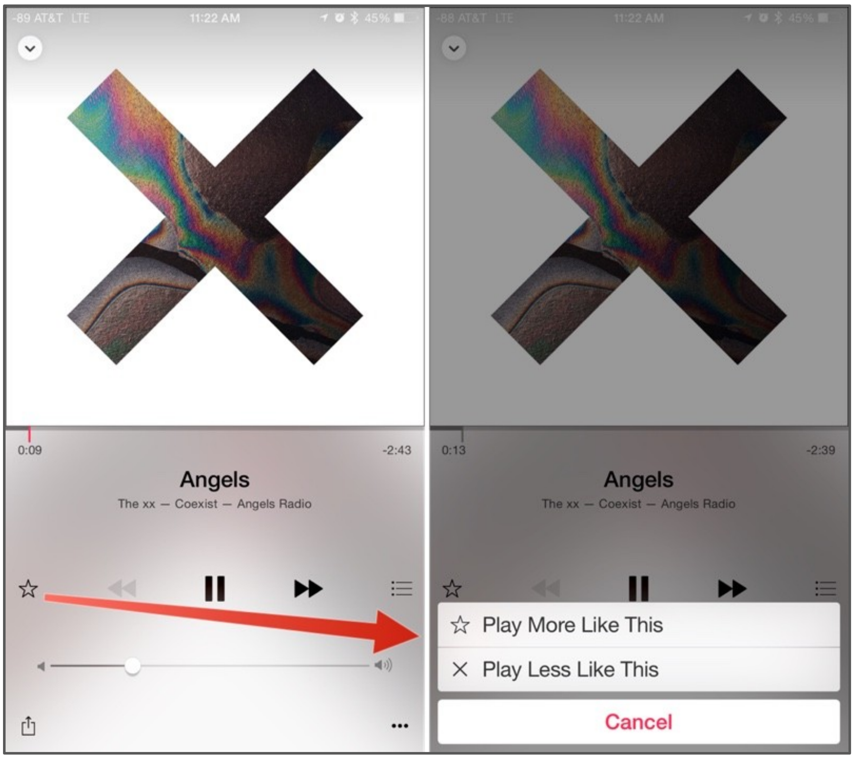


Figure 5: Apple's Music Rating Feature 1

and provide seamless recommendations. It features a "For You" section which is dedicated to recommendations. Therefore, the recommendation takes advantage of the following factors: Hearts. - Anything that is playing at a particular moment in Apple music can be liked by tapping the heart button. By tapping the heart icon regularly, helps the content on the "For You" section have better understanding of someone's music taste.

A user has also an option to dislike an album. He can achieve this by tapping and holding on the album cover on the "For You" section, and selecting the option: "I Don't Like This Suggestion.

Furthermore, if a user is interested in creating a radio station from individual tracks, when tapping "Start Station,", the playlist displays a start icon instead of a heart. The idea behind this feature is that the users, by taping the

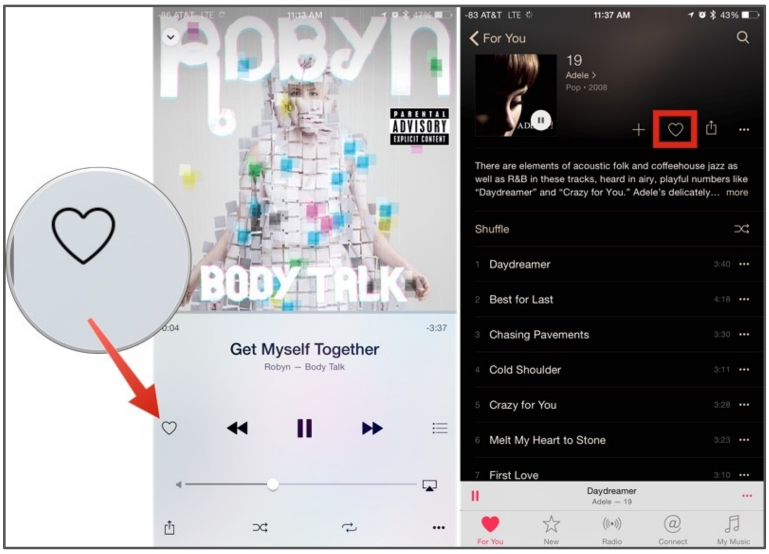


Figure 6: Apple Music Rating Feature 2

stars, can chose Play More Like This or Play Less Like This option to like or dislike a particular station.

Moreover, by pressing on an album cover or playlist recommendation gives the possibility to choose the "I Don't Like This Suggestion." option to customize even more the “For You Recommendations” [32].

Last, Apple Music provide also a tab called “New” which displays the latest album releases, trending tracks and artists.

3.2.4 Spotify

Spotify is a music streaming service that provides music to the users through its application. After creating an account, users can access all the Spotify music catalogue for free, legally. However, the user can, in return, not access the music file, but can thus only listen to it within Spotify’s application. In the free version, the user will also get audio advertisements between the tracks they are listening. This can however be avoided by paying a small subscription fee.

Spotify’s main recommendation feature is Discover which can be found under the Browse feature. Discover consist of a number of playlists, artist and albums which Spotify think the user would like to hear. Discover is based on the listening habits of the user and it is frequently updated depending on how often the user use the streaming service. The Discover’s criteria to recommend music are *[33]*:

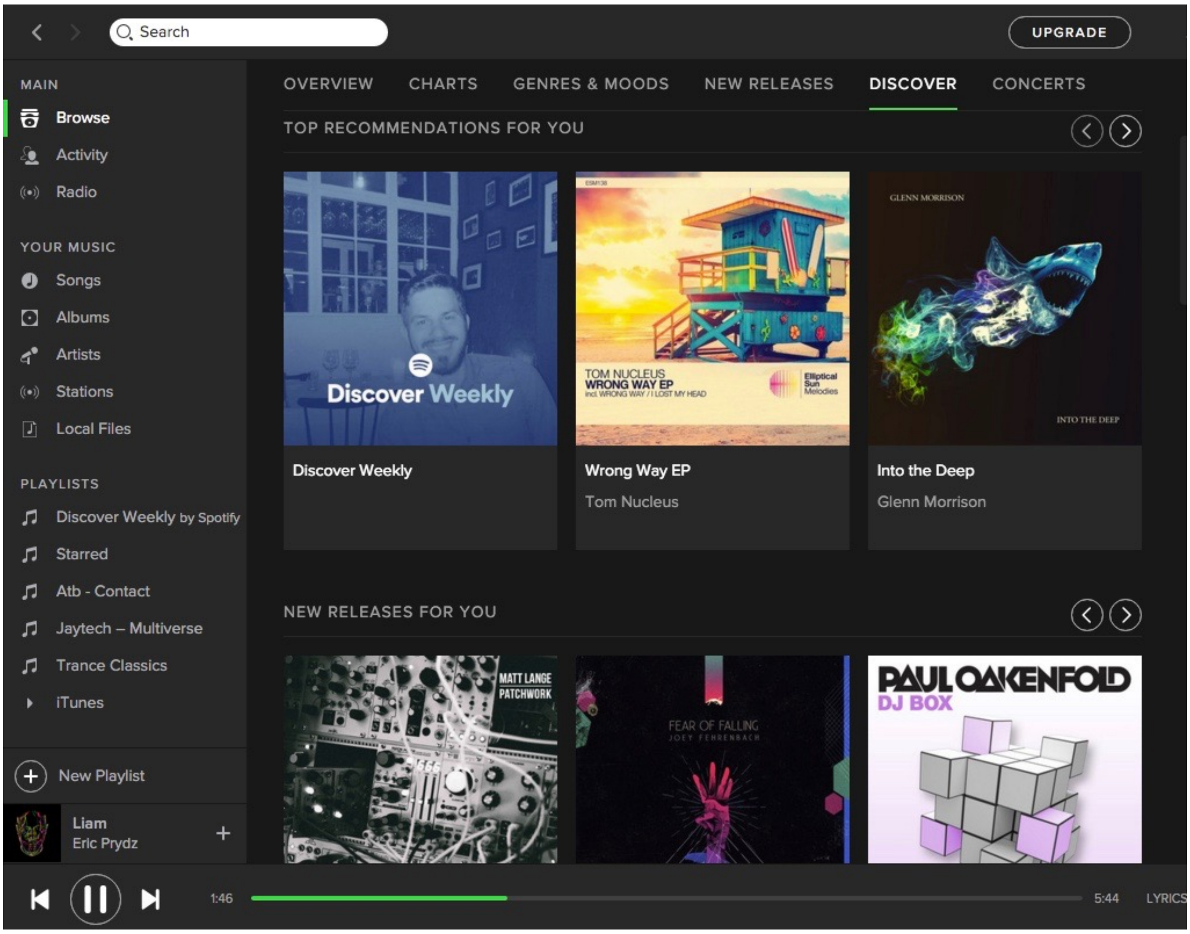


Figure 7: Spotify Recommendation Features

- *Top Recommendations*
- *Similar Artists*
- *New Releases*
- *Similar to…*
- *Because You Listened to…*

Spotify most famous feature is Discover Weekly by Spotify. It is a weekly-renewed playlist of tracks according to the listening history of the use and other users with similar music taste. The user must use Spotify at least two weeks in order to propose a playlist [34].

The main criteria in Discover Weekly is other people. Spotify looks at all of its users' playlist, which is a representation of music taste. Playlist is the core of Spotify ecosystem because it explains listening habits between people with similar music preferences. Spotify creates individual profile of users through clustering of artists and micro-genres. For example, a user's taste is described not only as "pop" and "rap", but with finely distinctive sub-genres such as "synth-pop" and "indie-r'n'b". This is created with a help of a music intelligence and analytics company Echo Nest acquired by Spotify in 2014 [35].

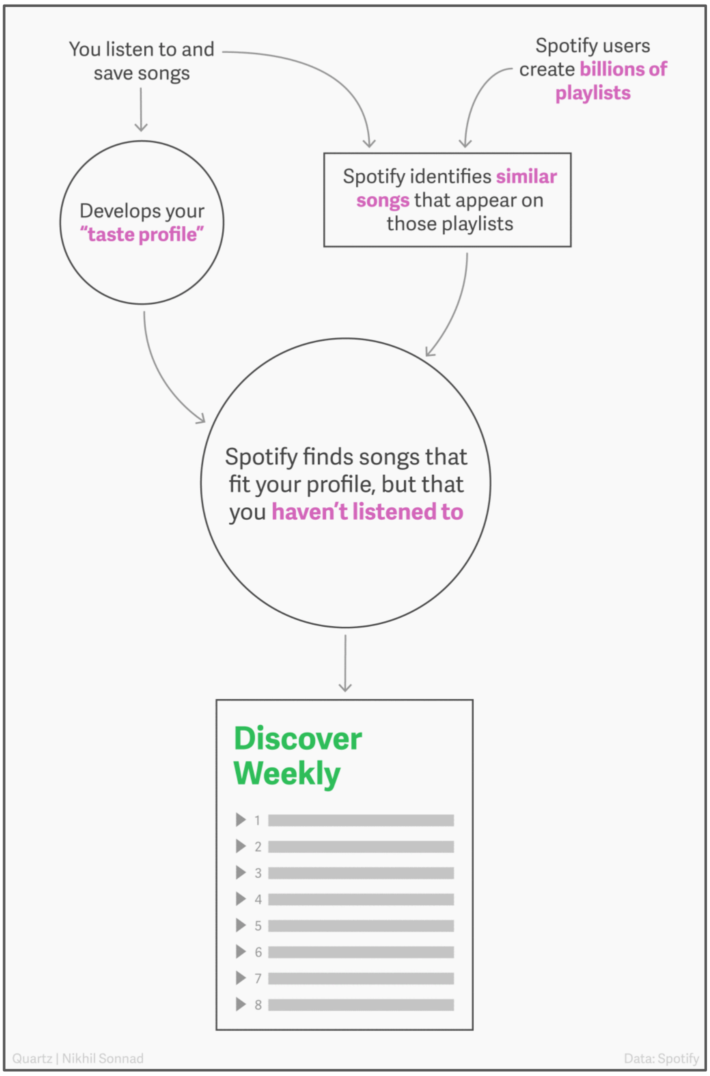


Figure 8: The approach to Spotify's Discover Weekly

All these features are incorporated by algorithmic techniques such as collaborative filtering and natural language processing. From the software perspective, the open source software Kafka is used to process the information real-time [36].

Other recommendation features are Fresh Finds as well as Spotify Radio.

Fresh Finds is a feature that recommends new undiscovered music that is being updated every week and the user can save their favorite track [37]. Fresh Finds creates playlist through an automated curation to new breakthrough songs through five genres. Spotify combines the online buzz about brand new music across blogs and music sites and a big number of listening hours on Spotify in order to capture anonymous listening patterns of fans that tend to listen to upcoming music [38]

On the other hand, Spotify Radio's feature proposes two kinds of radio-alike playlists based on the artist criteria and genre criteria [39].

4 Analysis

4.1 Interview Analysis

The purposes of this chapter was to include interviews that will add to the understanding of what is the relationship between MRSs and users and the problems that they encounter. The interviews helped us understand the users' needs, and based on their propositions I will address the problems discussed above in the problem formulation.

In the same time, the State of the Art chapter included an analysis of the different popular platform and technologies that make use of the MRS. So, this chapter will be also used here in order to reflect on the users 'experience with the above mentioned platforms.

The first interviewer stated that he has been mostly using the iTunes platform the last ten years, tried Spotify, and a little bit Pandora Radio. According to him, iTunes helps him find similar artist or music by using Genius Music recommendation system. He believes that the particular MRS is quite useful. Most of the time it recommends music that he has heard before, but also some new music, even though it is less likely.

However, during this interview an interesting fact came up about the lack in depth in its recommendation because it recommends only the most popular tracks or artist. He specifically says "… *most of the time the recommendation matches his preferences, but many times it doesn't recommend new music and as a result I have to search for them manually."* This confirms an already and well-known problem in the field of MRS. According to the literature (add reference) MRS platform, do not have enough data associated with the new artist, or track in order to make recommendations on them. According to him, Spotify is one of the best platform that has the best recommendation system among other platform like iTunes or Pandora Radio. He believes that the recommendations are more accurate. He believes that MRS are useful and productive to find new music. When he searches for example a particular artist or genre, the system can sometimes offer good recommendation. He thinks that MR has been more user-centered, and user should write his preferences such as the artists, genres, area etc. as an input for the system to provide better recommendations.

In the past, he has also used Pandora Radio. According to him, Pandora radio has the unique characteristic that it does not recommend music listened to by other users with similar tastes, but rather using only the personal preferences of the listener. The user's interaction with the application or website is utilized to prompt recommendations that are more accurate. The more he used the service, the more enjoyable the station's choices are and therefore the most successful the MRS is, according to his opinion.

The second interviewer has also used iTunes, Apple Music, Spotify as well as SoundCloud. He has not used Pandora and Last.fm because the services are not available in the country he lives now (Greece). According to him, Apple Music does not have many categories for electronic music in particular (a genre that he likes). He thinks that Apple Music have done so little to accomplish in order to provide better recommendation. The system lacks the ability to find new tracks, but matches the results according to the genre that makes it too general for him. So he prefers his recommendations to be more specific, "*…they have a lot of room for improvement".* On the other hand, he believes that Spotify has a better recommendation system. "*When you listen to a track, it pops up other tracks which matches the artist mostly and the genre in particular*."

In his experience, Spotify is much better than Apple Music when it comes to music recommendations. His favorite feature is the Radio feature on Spotify, which plays music based on the artist, but the recommendations are better and more personalized than Apple Music.

The interview became quite interesting when the conversation was shifted towards the SoundCloud music platform. According to him: *"SoundCloud has the best MRS from all the other services*". He elaborated more with an example. He was listening to a podcast from an electronic music artist called Mat Zo and while he was listening to the podcast, other tracks were popped that he loved very much. He further stated that Soundcloud has also a follow button, which helps the listeners follow artists that pop up. After following these artists, the results from the MRS became better. *"The recommendations were right to the point."* It is important to mention that even though Soundcloud lacks some features like the ones in Spotify, it still recommended new music. Moreover, Soundcloud even showed recommended him unreleased tracks, which he finds it great.

As a music enthusiast, he wishes from a MRS to recommend new music, especially new artist from his preferred genre, electronic music. For him the ideal MRS would be the system to analyze the sound itself such as the melody, BPM, rhythm rather than recommend music according to the metadata. SoundCloud *"has done some improvements, but believes it has potential to do much more".*

Lastly, he believes that one big disadvantage of MRS that some of its recommendations are advertisements. They are used in order to make more money, which means they compromise on the quality of the recommendations.

The third interviewer has used platforms such as You Tube and Spotify. The closest she knows from MRSs are the Related Video feature and the Discover feature in Spotify. It is worth mentioning that her background has no relation with ICT or Computer Science, as such it seems that she has a lack of understanding on how MRSs work. Although, she believes that, for example Spotify, knows her well and can recommend tracks and artists that she likes (through the Playlist feature), it fails to suggest new and unpopular artists and tracks. She states *"…I really like covers of songs so I would love to be recommended new covers…".* Furthermore, she is a bit sceptic about the recommendations in Spotify. *"I believe that somehow recommendations are advertised-driven and that's a bit sad".* Furthermore, the recommendations she gets many times do not match here preferences, for example, she was recommended some Danish tracks (and she generally does not listen to Danish music) and pop artists that she is not fond of such as Drake and David Guetta.

The fourth interviewer has a professional and educational background in ICT and it seems like he has a deep understanding on how MRSs work as he has conducted a similar research in the past. This means that he viewed the questions from another perspective, more from the specialist view and perhaps less from a user perspective. He also suggested some interesting points.

He has used iTunes and Spotify and says that in general these platforms manage to recommend good music to him. However, he would expect more serendipity on his overall experience with MRSs. This means that he would prefer new and fresh material to be suggested. He addressed this problem as the "cold start problem" and he says that, it happens due to the lack of metadata from the platform. "*I recently read an article about that problem and that Pandora is trying to do something about it…*" also "…a *solution would be two platforms collaborating together in order to use this information from SoundCloud…*"

4.2 Personas & Scenarios

According to Alan Cooper, personas are user profiles that are defined by their behavior, goals and motives. Personas are not real users, but they accurately represent characteristics of real world users. Personas allows us to focus on a few memorable characters rather than thousands of individuals. Therefore, we have a better understanding of the users' requirements that represent a specific target group. In this way personas, help us understand the users and consequently improve a software or product design [13]

Personas are the main characters that comes into play through Scenarios
"*like actors reading a script, to test the validity of our design*" [13].

He also refers to scenarios as *"a concise description of a persona using a software-based product to achieve a goal"* [13]. Scenarios are often created from material gathered during the research phase. This fact also applies in my case. More specifically, the interviews from the real participants helped me have a better understanding the users' needs and as a result facilitated the process of constructing a scenario. Below I present two personas and scenarios:

4.2.1 Persona 1

- *Name: Tobias Hansen*
- *Age: 28*
- *Occupation: Software Developer*
- *Education: Master MSc.*

Tobias was born in Aarhus, but lives in Copenhagen for the last 9 years. He works as a software developer in Nordea Bank. He is a music enthusiast and owns a huge catalogue of vinyl, CDs as well as digital music in his computer and iPod. He loves to listen specially to jazz and blues. He loves to discover new tracks and artists and he spent quite some time searching in places like digital stores, streaming platforms, social networking sites as well as forums.

4.2.2 Scenario 1

4.2.2.1 As Is Scenario

After returning home from an exhausted day at work, Tobias decided to relax by listening to some music. He is just tired of the same tracks and albums, so he decided to try Pandora. He likes the fact that Pandora provides accurate recommendations, but feels that they it does not necessarily recommend new tracks and there is a lack of serendipity.

More specifically, he expects from MRS to find new music from new artists with the genre he loves: in this case jazz and blues. Additionally, MRS should know him better and provide a more personalized and serendipitous recommendation.

4.2.2.2 To Be Scenario

After being disappointed by the recommendations from different music platforms, Tobias opens Spotify and tries MY New Releases recommendation playlist. The MRS starts to play tracks from the genre jazz and blues, from artists and tracks he has not heard in the past. Since Tobias taste is very meticulous, he is skeptical in the beginning. However, after some time, he is positively surprised with a few recommendation results. He like the fact that even nowadays there still new artist that tries to keep his favorite genres alive and make quality music.
He will definitely use the feature again.

4.2.3 Persona 2

- *Name: Pete Tong*
- *Age: 44*
- *Occupation: Radio Host*
- *Education: High School*

Pete was born in Great Britain. He is a DJ who works in a British radio station and hosts a famous weekly radio show named "Essential Mix". "Essential Mix" broadcasts many styles of electronic dance music.

4.2.4 Scenario 2

4.2.4.1 As Is Scenario

Pete´s show is largely focusing on finding new acts of the electronic and dance music scene. His show is largely famous among the dance community, mainly because of his discovering and airing of new tracks and artists. He has worked hard to maintain the title of one of the most prestigious shows in the world and he is still trying to in order to keep up with this reputation. Therefore, Pete is constantly trying to find new talents in order for his show to maintain the popularity and prestige. Pete is a famous host so he receives a lot of new material every day in his e-mail box, but it impossible and a time-consuming process for him to listen to every one of them. For this purpose, he uses commercial platforms such as Spotify, Apple Music, and Pandora in order to discover new talents. However, those platforms have MRS features that do not provide him with the desired outcome, which is to find music that matches his personal preferences as well as find new artists and tracks that would match his radio show´s music style.

4.2.4.2 To Be Scenario

Pete constantly looks for new music in order to feature in his show, for that reason he uses MRSs. When Pete uses Spotify, he uses my new solution (MY New Releases) in order to find new releases that match his preferences, which is particularly a mix of different styles of the electronic music genre such as house, deep house, progressive house and minimal-techno.
This means that my solution provides him with a more customized recommendation because the system uses his preferences based on his interaction with the MRS, as well as his listening history through a collaborative filtering technique.

5 Problem Solution

As discussed in the State of the Art chapter, one of the most recent development (which were presented nearly one month ago) in addressing the Cold-start problem is the Fresh Finds in Spotify. This is the newest recommendation that recommends new unknown trending music to its users. According to the founder of Echo Nest, Whitman, the feature creates playlists by crawling several blogs; new websites as well as reviews to capture anonymous listening patterns of fans that tend to listen to upcoming music. Finally, a playlist of brand new tracks that could have the potential to be hits soon, is produced. This feature has as a vision that less commercial artist will have the chance to become more visible and the listeners discover brand new tracks from artists that they have not heard before. However, Fresh Finds does not take into consideration the users 'preferences and as a result do not recommend fresh tracks that identify with the listening taste of the user. This means that this feature may propose fresh music but does not improve the overall experience of the user with the MRS.

My solution has a different approach. The system will use the existing collaborative filtering technique which is used by several music platforms (such as Spotify), and will retrieve information from the SoundCloud platform to recommend new unheard tracks and artists from the users based on their music preferences. For this reason, I chose SoundCloud based on some specific criteria.

SoundCloud is a platform where artists can upload and share tracks. It is also considered a perfect tool for an artist to promote their work. All tracks are uploaded as a waveform and have a distinctive URL. Users can post comments on specific parts of this waveform. The comment is being shown in the waveform every time the specific part of the the track is being played [40]. As a result, this creates the perfect environments from artists to promote their work and interact with their fans and for the users to discover new music, even unreleased ones [41].

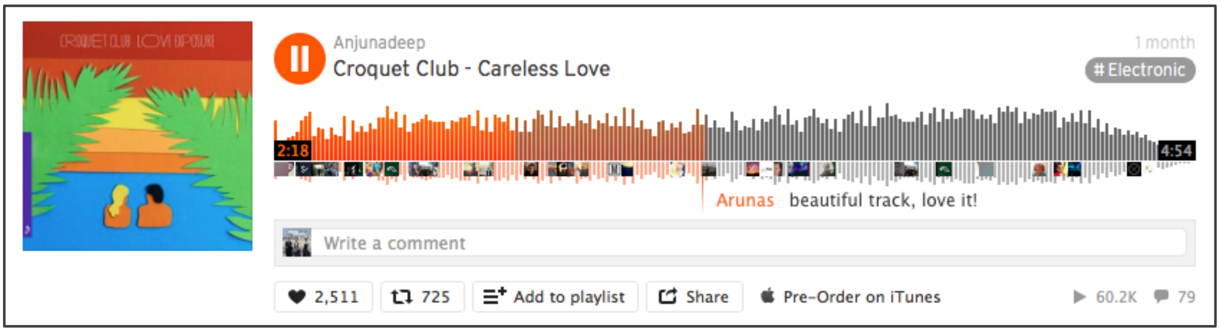


Figure 9: SoundCloud Waveform Track

This fact is also validated by my interviews where two participants expressed their satisfaction using the platform, particularly when it comes to discover new tracks and artists. SoundCloud has also great sharing options because they are embedded easily in any social networking site such as Facebook or Twitter. This allows artists to distribute their music with a broader audience. SoundCloud furthermore offers an API, which allows other applications to use its feature resources like tracks, playlist as well as users. I will also take advantage of some of these for my proposed solution.

5.1 Design

This section demonstrates the design process followed in my proposed solution. More particularly, this solution is called “My new releases” and can be part of any of the commercial music platforms described in the State of the art chapter. However, in the design process I used Spotify as an example. The reason why I used Spotify is because of the feedback acquired by the interviews that Spotify was one of the most popular and frequent used platform. Furthermore, the interviewees stated that Spotify made the best recommendations among other platforms.

5.1.1 Existing Design

In the actual design, Spotify's recommendations are shown in one of "Browse" features, called "*Discover"*. Recommendations are categorized in different sections like "*Top recommendations for you"* which include "*Discover Weekly*", the weekly playlist that Spotify suggest to the user based on his/her preferences; "*New releases for you",* which recommends new released tracks and albums from artists that were played recently; *"Suggested for you based on* (an artist played recently)", "*Because you listened to* (another artist played recently)", "*Similar to* (a third artist played recently)", "Because you listened to (a recent album)" etc.

As mentioned, "New releases for you", recommends only new material from already played artists, that Spotify already know from the listening history of the user. However, in these sections no recommendations are suggested regarding brand new artists, that the user could be interested to hear. Even the "Discover weekly" playlists does not include new undiscovered tracks but is limited only to popular tracks and artists. This is what the next subchapter presents. A new and improved design to recommend brand new tracks and artists that the users have not listened to, based on their preferences, using the already existing MRSs as well as extra information from SoundCloud music platform.

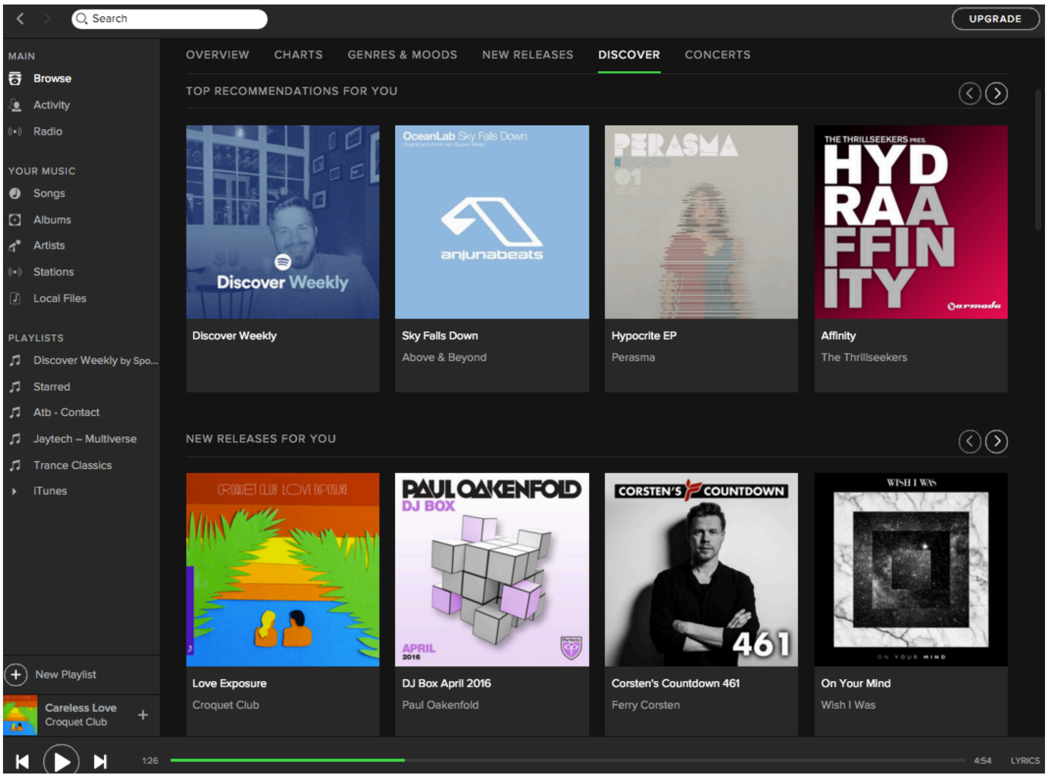


Figure 10: Spotify's Current Design

5.1.2 New Design

The new design is set up in the same environment as the existing design, but in this case, we have an added feature. Again, we go to "Browse" and then "Discover" and then we scroll down to find the new added feature called "My new releases" which has two subcategories called "My new releases - songs recommendations" and "My new releases - artists' recommendations".

As you can see in the Figure 11 below, the first box in red is meant to recommend brand new songs, while the second box is meant to recommend brand new artists.

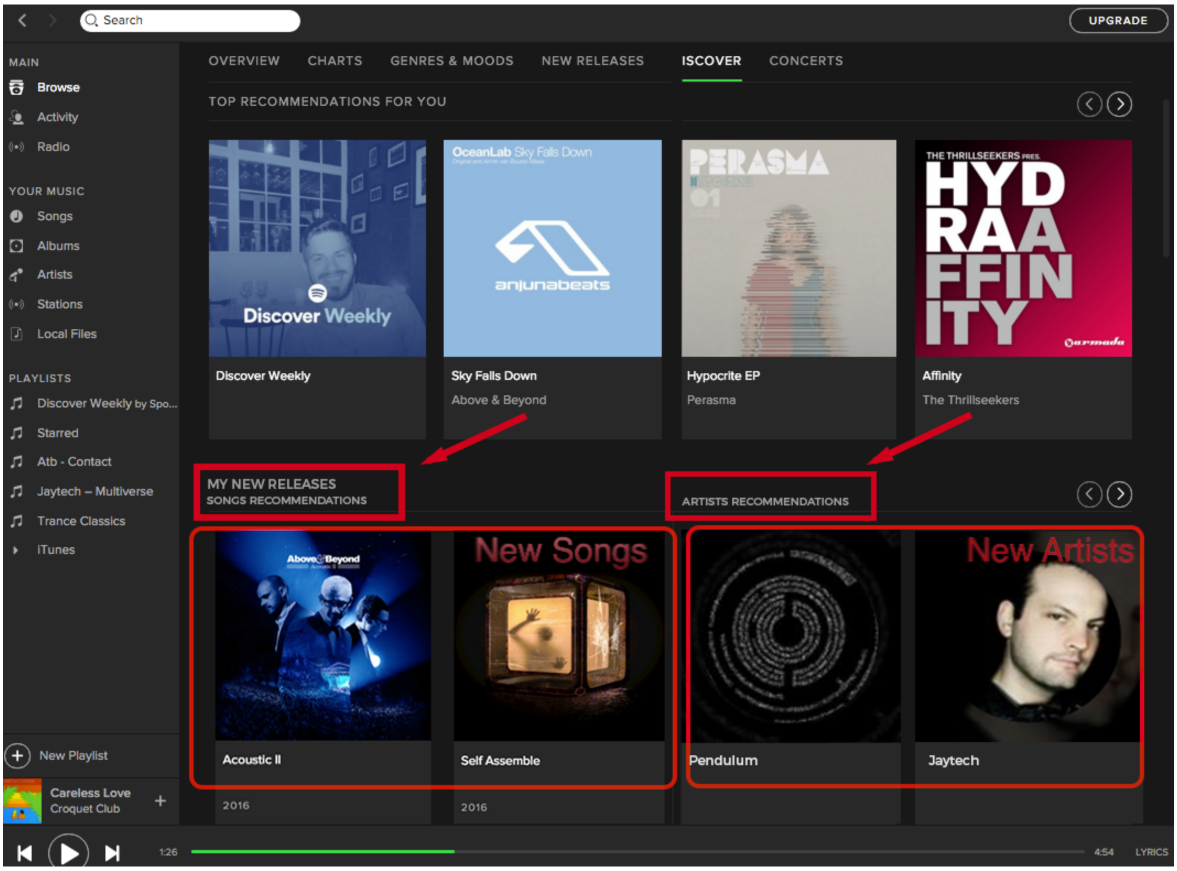


Figure 11: Proposed Design

5.2 UML

UML stands for Unified Modelling Language. UML is a method of visualizing a system by using a collection of several types of behavioural diagrams. [47] Below, I present the context diagram, use cases as well as class diagram.

5.2.1 Context Diagram

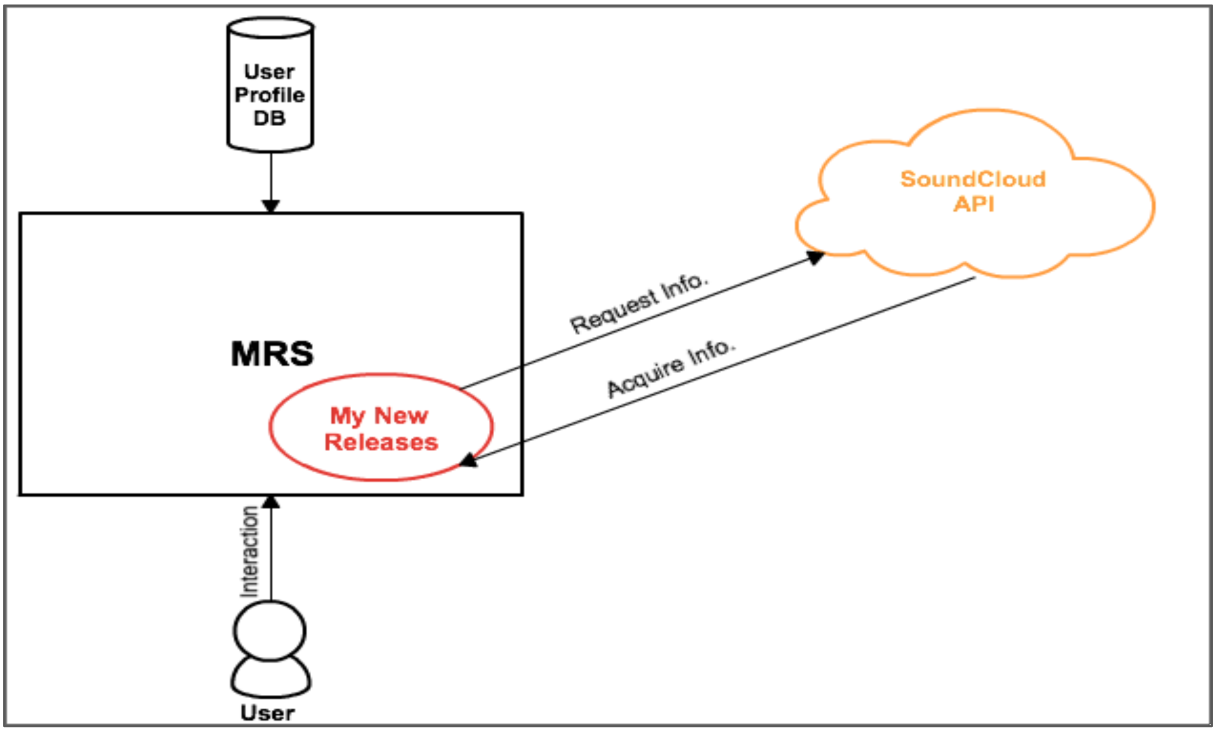


Figure 12: Context Diagram

The Figure 12 shows the Context Diagram of a proposed music recommendation system, which is located in the small rectangle called MRS. The MRS interacts with the user by taking all the data of the user's listening habits. All this information is sent to the database circle form called User Info Database. User Info Database collects the entire user's preferences data. Inside the MRS rectangle is located a small circle in Blue, called My New Releases Feature. This feature is a recommendation feature part of MRS that is linked with SoundCloud API in order to retrieve some particular data. This data is then sent to My New Releases feature to generate specific recommendations. A detailed presentation of it is included in the next chapter.

5.2.2 Class Diagram

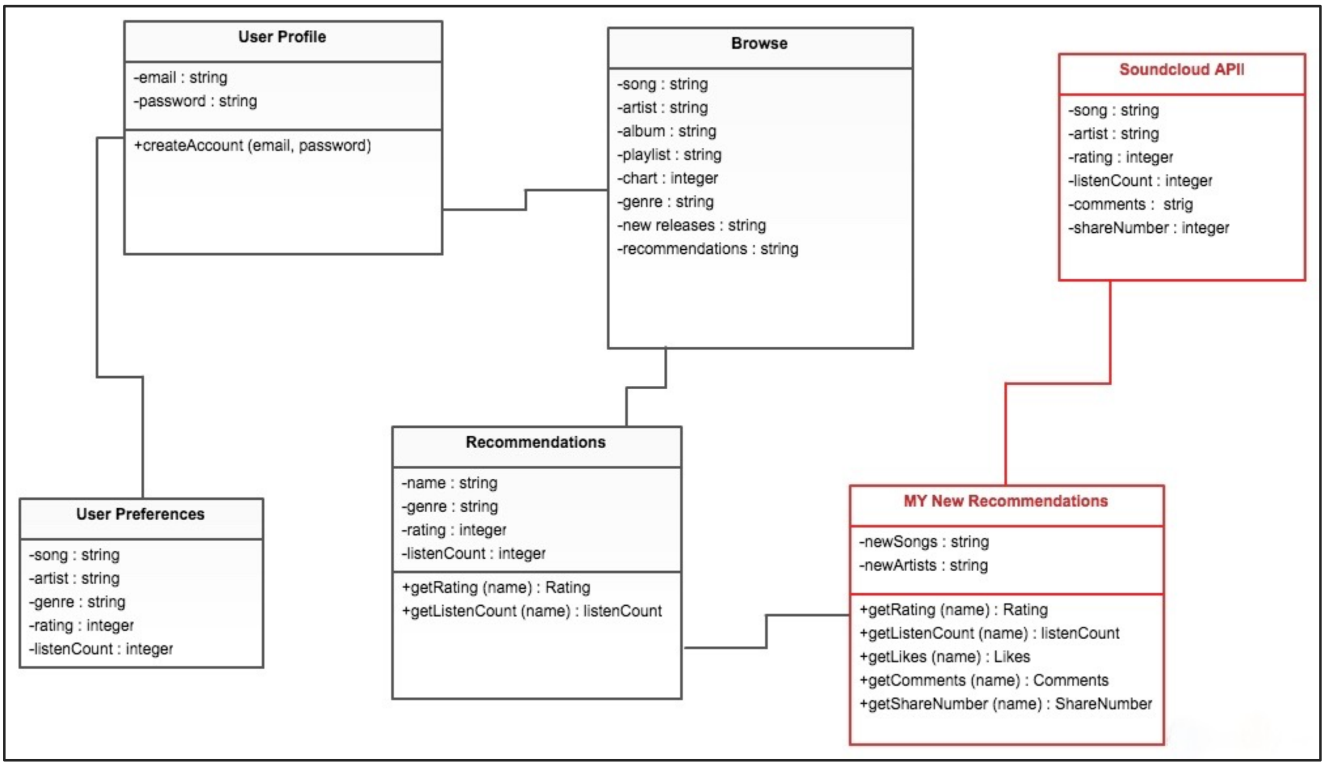


Figure 13: Class Diagram

The Figure 12 shows the Context Diagram of a proposed music recommendation system, which is located in the small rectangle called MRS. The MRS interacts with the user by taking all the data of the user's listening habits. All this information is sent to the database circle form called User Info Database. User Info Database collects all the user's preferences data. Inside the MRS rectangle is located a small circle in Blue, called MY New Releases Feature. This feature is a recommendation feature part of MRS that is linked with SoundCloud API in order to retrieve some particular data. This data is then sent to My New Releases feature to generate specific recommendations. The class diagram shows an MRS with my proposed solution.

5.2.3 Use Cases

The figure below shows various possible use cases, which the user is required to do in order to accomplish a specific task. The big rectangle shape represents Spotify platform. Inside the rectangle we can see the uses cases in circle shapes and arrows named either <include> or <extend>.

<include> means that the use case functionality can be duplicated into another use cases, while <extend> mean the the use case delivers additional functionality to another use case [45].

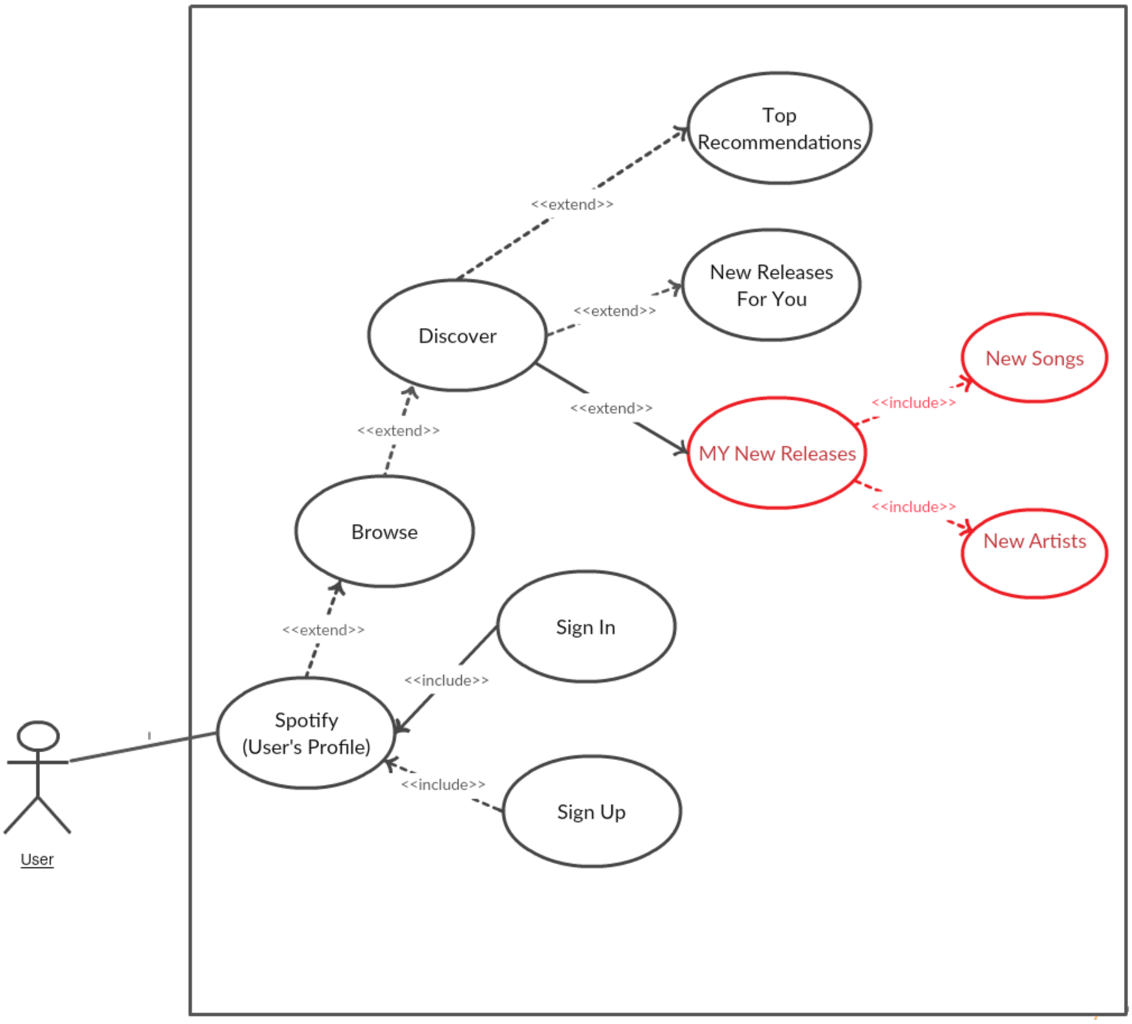


Figure 14: Use Cases Diagram

Below I have created a table with four uses cases that explained in detail what the user has to do to accomplish a particular task. The table is divided into 5 parts:

- *Use Case*, the part that represents the title of the use case.
- Second row is *Description*, which explains what the user need to do in each case.
- Third row is *The Pre-condition* that describes the state of the system before the particular case is able to start.
- The *Post-condition* on the other hand explains the state that the system after the particular case is executed [43].

Lastly, Basic Flow explain the steps that the user need to take to accomplish a specific task.

User Case	Description	Pre-condition	Post-condition	Basic Flow
Use Case 1: Access MRS	In order to access the MRS, the user has to sign in or sign up to Spotify.	The user cannot access his profile on Spotify, if he has not signed up.	The user is logged into Spotify's home page.	The user has to provide his email and password in order to access his profile. Alternatively, he can log in via Facebook.
Use Case 2: Browse Music	In order to browse songs, artists, genres, songs according to mood as well as discover music, the user has to go to the Browse section of the platform.	The user cannot browse music if he is not logged in into his platform.	The user can is able to see a list of options such as charts, new releases, recommendations etc.	The user has to click on the Browse tab in order to access the recommendation.

Use Case 3: Discover Music	In order to discover music based on user' preferences such as top recommendations, new releases etc., the user has to to go to the Discover section of the platform.	The user cannot browse music without accessing Browse Section.	The user can is able to see a list of recommended music base on his preferences.	The user has to click on the Discovery tab to access all the Discovery features.
Use Case: 4 Discover New Music	In order to discover only new songs and artists based on a user's preferences, he has to access the My New Releases feature, which is part of the Discovery section.	The user cannot access the My New Releases feature without accessing the Discover section.	The users would be able to see a list of 10 recommended songs and artist, based on his preferences.	After clicking the Discovery tab, the user has to scroll down to see the My New releases feature on Spotify's platform.

Table 1: Use Cases

5.3 Software Requirements Specifications (SRS)

SRS is a description of a specific software product, program or group of programs that performs a set of particular functions in a given environment (46). Some of the issues that SRS should address are the purpose of the software program, the relation of the user with the software and the hardware, the performance of the software in terms of speed, response time, availability etc.; design-related limitations in the implementation of the software, etc. (46). According to (46), a good way to sort SRSs is the one based on the degree of necessity and importance. Not all requirements are equally important for my solution. Thus, the SRSs below are as categorized in essential, conditional and optional.

Finally, the requirements presented in the table below are based on the analysis of the user needs deriving from the semi-structured interviews, personas and scenarios descriptions and the UML diagrams.

Requirement ID	Requirement Specifications	Priority
R1	Users expect recommendations from brand new artists.	Essential
R2	Users expect recommendation on brand new tracks.	Essential
R3	Users expect exclusively brand new recommendations based on their music preferences (listening history, ratings)	Essential
R4	Users expect more personalised recommendations.	Essential
R5	Users expect from a MRS to prioritize the new recommendations.	Optional
R6	Users expect from this feature to be more visible in the music platform.	Optional
R7	Users expects recommendations from unreleased tracks (Demo).	Optional
R8	The system requires additional information from SoundCloud API.	Essential
R9	The system requires data retrieval from other sources such as music blogs.	Optional

Table 2: Software Requirements Specifications

5.4 System Architecture

The final solution of the system architecture is developed based on the system and the user requirements. This is described in details below (figure 14). According to this figure, the system is comprised of hardware and software functional elements, as well as the human interaction between them [44]. The infrastructure of the architecture is relying in the music recommended platform and a new module on top of it named "My new releases".

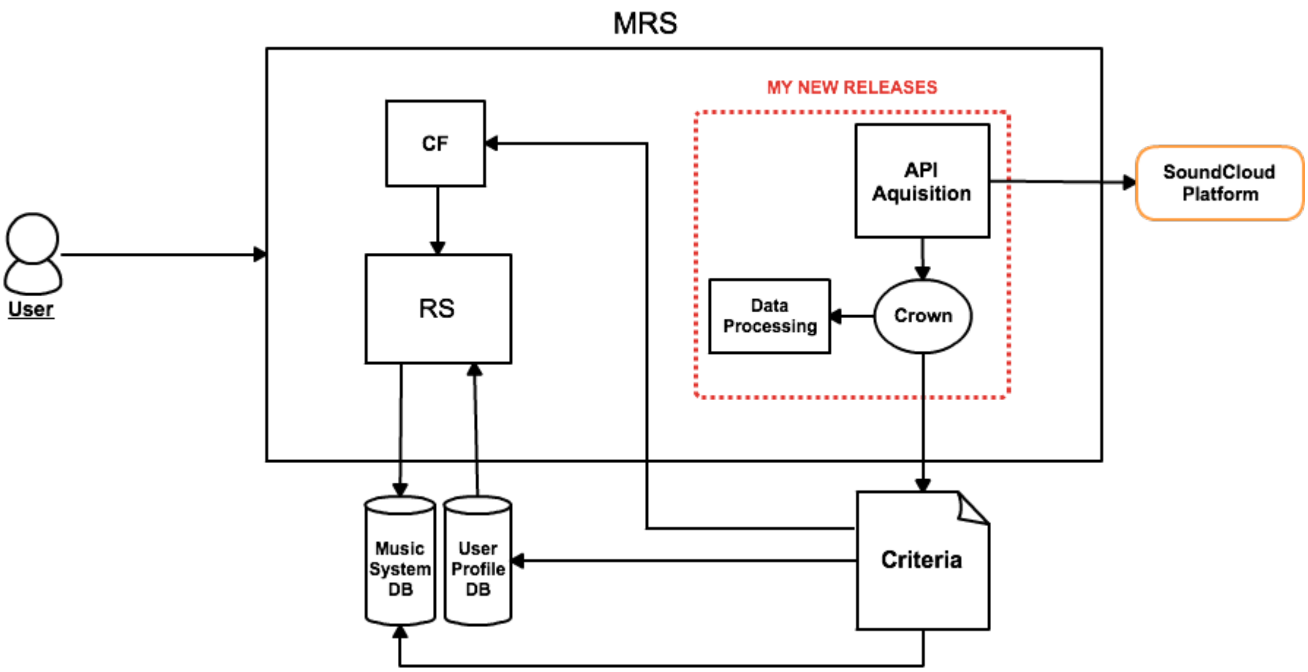


Figure 15: System Architecture Diagram

Below I will describe each of these elements and their functionalities:

The MRS consists of the following parts:

1. The database where all the available tracks of the platform are stored.
2. The user profiles, which is the information regarding the user preferences, such as ratings and listening counts or history and likes.
3. The collaborative filtering method, which make possible recommendations, based on users' listening behavior.

The "My new releases" module consists of:

- ***The Acquisition Method functionality***, which is to connect to the Soundcloud API and retrieve the necessary information that will be used in the proposed solution. This information consists of the different factors such as the release year, title, genre, comments counts, playback counts as well as favorite counts for every song uploaded in the system. The table below presents a summary of the information retrieved from SoundCloud and how this information is going to be used.

Acquisition Info.	Description of the Factor	Example Value
comments_counts	This factor represents the number of comments made on a particular timestamp of a track. This information is useful because comments indicate that the users enjoy the track.	32
release_day	This factor represents the day a particular track is released. Since my proposed solution is based on new tracks and artist recommendations, this factor is critical as it indicates the freshness of the track.	2016
title	This factor refers to the title of the track. This factor is essential since a playlist of recommendation tracks needs to be displayed on the web page.	"Another Chance"
genre	This factor represents a particular genre of a track. This factor is also crucial because the system should recommend new music and artist based on the user's genre preference.	Rock
playback_counts	The number which a track has been played. This factor is also important because the more the play counts, the more the possibility that the users might enjoy the track.	6857
favoritings_counts	The number which a track has been favorited. In this particular case, we are certain that the users have enjoyed a track, for the fact that they rated it as one of their favorites.	23

Table 3: Information Retrieved from SoundCloud

The above information will be used during the development, in order to address the research problem.

- ***The Crown Method***; makes possible to call the Acquisition Method in specific time intervals. This means that the SoundCloud API will be accessed once per day due to the API access limitation. This method will ensure that the information retrieved is up-to-date. It is also used for security reasons due to the fact that "My New Releases" feature would not be so depended on the SoundCloud API.

- ***The Processing Method*** is in charge for processing the retrieved information from the above-mentioned API. This information will be processed according to specific algorithms in order to be used later on in the main recommendation system. The results from the processing method would be a list of popular tracks and artists according to the above information table. E.g., the list proposed is based on a combination of the highest number of likes, shares and listening counts.

- ***The Criteria Method*** evaluates the above two lists based on specific criteria such as Freshness and number of likes. Every time that a list is produced the Criteria Method will check the music database recommendation to make sure that the above criteria are fulfilled. If these criteria are satisfied then the MRS will recommend these lists to the user.

5.5 System Implementation

The implementation of my proposed solution was based on the requirements that resulted from the analysis and the UML diagrams. The implementation will be a proof of concept in order to demonstrate that the potential service can be successful.
The method used for developing this solution is the Waterfall Method.

The following section presents how the above methods will be implemented:

API Acquisition

The code below, which is written in JavaScript, demonstrates a function that allows us to connect to the SoundCloud API. The SoundCloud API provides tools that makes possible for us to take advantage of several SoundCloud features and use them for our purpose. More specifically, The API is used in our case to extract the required information that will be stored into the new database and as a result generate a playlist of tracks and artists. In order to get access to all these data, I have created a SoundCloud developer account.

```
<script src="//connect.soundcloud.com/sdk.js"></script>
SC.initialize({
client_id: "9f2993e05a13357e4e3dac5f7034d044",
redirect_uri: "http://example.com/callback.html",
});
```

The code below represents an example of the extracted information from SoundCloud API for some tracks before filtering. As seen in the figure below this information is somehow unstructured, but still we can distinguish some parameters such as track ID, the date which it was created, duration, size, genre, audio format and many more.

{"kind":"track","id":17352827,"created_at":"2011/06/17 22:19:32 +0000","user_id":3845411,"duration":191080,"commentable":true,"state":"finished","original_content_size":5348935,"last_modified":"2012/05/20 01:03:40 +0000","sharing":"public","tag_list":"","permalink":"genres","streamable":true,"embeddable_by":"all","downloadable":true,"purchase_url":null,"label_id":null,"purchase_title":null,"genre":"HIP HOP, BOOM BAP, SOUL","title":"GENRES","description":"","label_name":"","release":"","track_type":"","key_signature":"","isrc":"","video_url":null,"bpm":85.0,"release_year":null,"release_month":null,"release_day":null,"original_format":"mp3","license":"all-rights-reserved","uri":"https://api.soundcloud.com/tracks/17352827"

As described in the Table 3, I am not interested in all this information, but only in specific ones such as favoritings counts, playback counts, genre, title, release day as well as comments counts. The section below demonstrates the way that specific parameters of a particular track are being filtered from the extracted information acquired previously from the SoundCloud API. As a result, a list of every track from the year 2016 described by title, genre as well as comment, play and favoriting counts is generated. We used the function below in order to achieve this.

```
var genres = Math.floor(Math.random() * 100000);
  SC.get("/tracks", {
   genre: genres,
   bpm: { from: 120 },
   release_year: '2016',
   track_type : 'original', 'remix',
   limit: 100
  }, function(tracks) {
  var tmp = '';
  for (var i = 0; i < tracks.length; i++) {
```

```
tmp = '<a href="' + tracks[i].permalink_url + '">' +tracks[i].release_year + '-' +
tracks[i].title + '-' + tracks[i].genre + ' - ' + tracks[i].comment_count + '-' +
tracks[i].playback_count + '-' +tracks[i].favoritings_count  + '</a>';
 $("<tr/>").html(tmp).appendTo("#track-list");
}
});
```

However, apart from normal tracks, SoundCloud Platform provides other digital media such as radio shows, podcasts, demo as well as unreleased tracks which we are not interested in. So My New Releases feature should only retrieve information based on some criteria. These criteria are:

- Released tracks
- Fresh tracks
- Original and remixed tracks.
- Genre

As a result, this function helped us bypass the overflow of information that the SoundCloud API provides for a specific track.

After gathering all these information, they need to be stored in our database. The purpose for this action is that in this way, the data will be independent from the SoundCloud API. Once the data are being stored, then this information can be still used for the list of recommended tracks, even if for some reason the Soundcloud API encounters a problem. Below is the information stored through a PHP business logic to my database.

```
<?php
$con = mysql_connect("mysqlDB","cis_id","Aris2016");
if (!$con) {
die('Could not connect: ' . mysql_error()); }
<form name="myform" action="<?php echo $_SERVER['$PHP_SELF']; ?>"
method="POST">
        <input type="hidden" name="tmp" id="tmp" />
</form>
mysql_select_db("User_id", $con);
 for($i = $start; $i < $end; $i+=$step)
{ $sql="INSERT INTO NewRelaeses table (start.CommentsCount, start.Title,
start.ReleaseYear, start.Genre, start.PlaybackCount, start.FavoritingsCount)}
echo "records added";  ?>
```

The Crown Method

After we acquire the relevant information form SoundCloud API, we should make possible that these information is being called in specific time intervals due to the API limitations. In our particular case The Crown method is used to perform calls once per day and afterwards store the information extracted into our database. As a result, the Information will be up-to-date.

The Processing Method

After the relevant information is acquired from SoundCloud API, the MYSQL query below shows the way how the top 10 tracks are extracted from the database. The tracks that make the list have the highest value in terms of playback, favoritings as well comments count.

```
SELECT
  *
FROM
  NewReleases
WHERE
  id = (
      SELECT   id
      FROM     NewReleases AS Lookup
      WHERE    Lookup.ReleaseYear = NewReleases.ReleaseYear
      ORDER BY CommentsCount DESC,  PlaybackCount DESC,
FavoritingsCount DESC
      LIMIT    10  );
```

The Criteria Method

The Criteria Method Acquires the users' preferences from MRS and compare them with users 'preferences in our database, based on the the genre, title and freshness criteria. If the criteria of the user from the MRS matches with the ones retrieved from SoundCloud API, then the function will query the database to get a playlist of 10 tracks. This function should also calculate tracks that have the highest number of playback, favoritings as well as comment counts. An example on how the data will be displayed is presented below:

Release Data	Title	Genre	Comments Count	Playback Count	Favoritings Count
2016	Fractures	House	11	112	122
2016	DJ RAP	Electro House	15	2	212
2016	SG1.1	Electro House	0	6	1
2016	STORM	Techno	11	56	11
2016	Scoley 1	Dubstep	2	10	25

2016	Deep & Soul	Deep & Soul	0	33	6
2016	H O U R S	Techno	0	2	1
2016	Trap City	Hip Hop	12	24	50
2016	Two Friends	Deep House	0	2	1

Table 4: A Playlist of Tracks

Finally, I have created a particular playlist such as the one represented in Table 4 is show, this show is displayed in our proof of concept implementation:

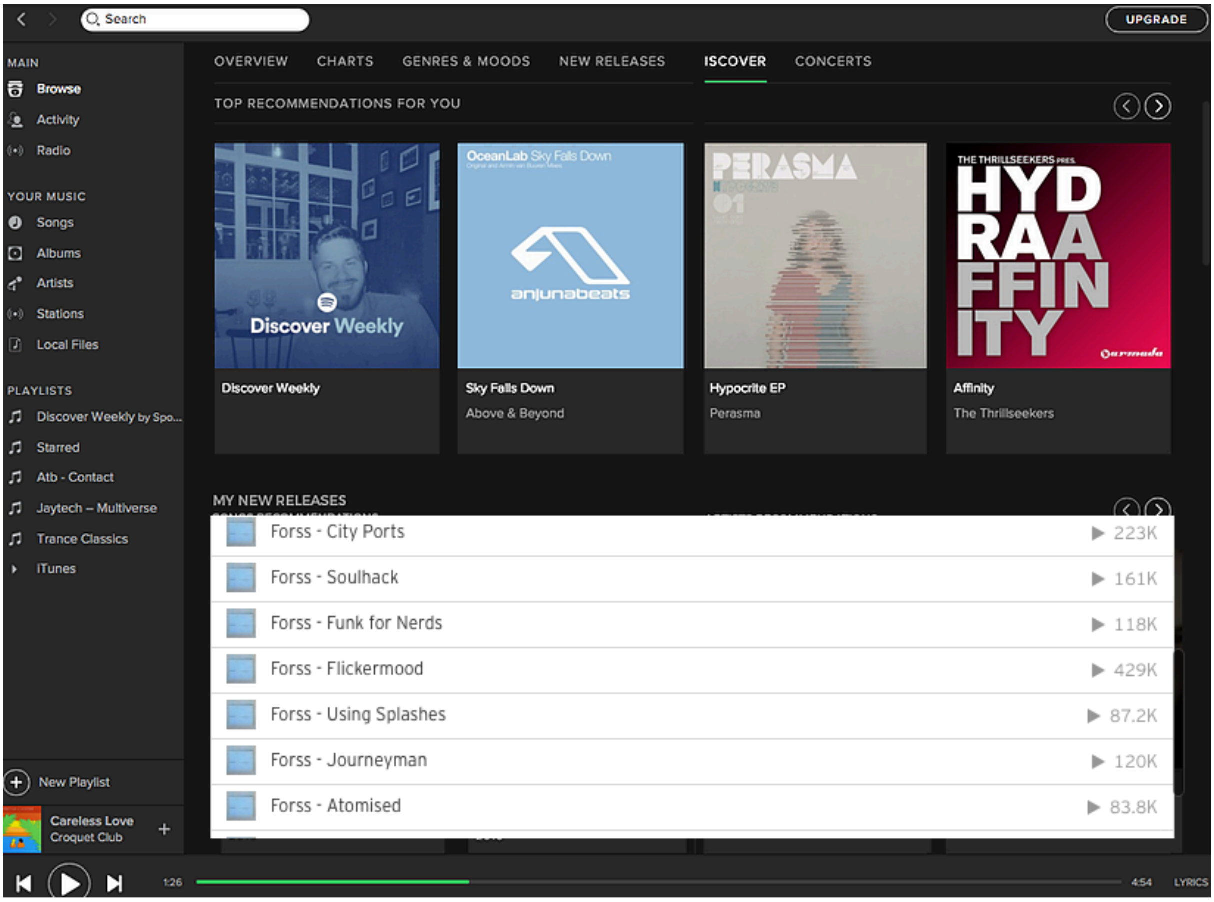


Figure 16: A list of Recommended Tracks

As a result, in the figure above, a playlist of 10 is generated based on the highest value of playback counts.

6 Conclusion

The goal of this master thesis was to propose a solution in order to address the Cold Start problem in Music Recommendation Systems (MRSs). The proposed solution offers users better recommendations in term of new music which they haven't heard before. More particularly, it would recommend new tracks and artists based on their music preferences using collaborative filtering technique and extra information acquired from SoundCloud music platform.

First, after the problem formulation was defined and narrowed down to a specific research question, I explored the problem domain by using both primary and secondary research. More specifically, I searched for literature sources on the concepts of MRSs. I also conducted four interviews with individuals that were already familiar with MRSs. The knowledge obtained from the qualitative research and the literature study helped me understand the relationship between users and MRSs and the problems that they encountered.

Next, I explored several recommendation techniques used in MRSs such as Collaborative, Content-based, Demographic as well as Hybrid Filtering Techniques. Additionally, I analysed the commercial music services that use their own version of MRSs in order to offer users accurate recommendations, such as Pandora, Last.FM, Spotify as well as iTunes and Apple Music.

The analysis process derived from the semi structured interviews, personas as well as scenarios descriptions, helped me understand the users' needs.

Finally, based on the users' needs as well as the UML diagrams, I was able to create the Software Requirement Specifications, the System Architecture diagram as wells as System Implementation.

The solution proposed in this thesis is a recommendation feature called "My New Releases" that can be part of a current music platform (such as Spotify). The feature more specifically uses SoundCloud API to retrieve specific information of tracks. After

this information is acquired and processed, the feature generates a list of 100 tracks and artists. The playlist consist only of fresh tracks and is based on a combination of the highest number of rating factors such as comments, playback as well as favoriting counts. This playlist is compared with the users´ preferences on the existing platform database, based on the genre, title and freshness criteria. The result of this comparison will be playlist of 10 tracks. The idea behind this solution is to offer users new tracks and artist they haven't heard before which would match with their preferences.
So, this solution addresses the issues derived from Cold Start Problem related to new items, in our particular case to new tracks and artists.

Furthermore "My New Releases" has utilized the factors such as comment, playback as well as favoriting counts, in order to improve the overall experience of a user in an MRS.

6.1 Future Work

In the section above we said that the playlist that is generated from the SoundCloud API is based on a combination of the highest number of rating criteria such as comments, playback as well as favoriting counts. This criteria however could be enriched by using comments and annotations made by users on a specific track on SoundCloud.

More particularly, this information could be stored into a database, analysed and provide valuable insights about users 'preferences. This way we can see if users enjoy a particular track or not. This criteria would complement our feature, and as a result improve our proposed solution.

7 References

[1] "How Napster Changed the Music Industry | Sessions X", Sessionsx.com, 2016. [Online]. Available: http://sessionsx.com/history/looking-back/how-napster-changed-the-music-industry. [Accessed: 04- Apr- 2016].

[2] M. Einhorn, "Copyright, Prevention, and Rational Governance: File-Sharing and Napster", Columbia Journal of Law & the Arts, Forthcoming, SSRN Electronic Journal.

[3] C. Angulo, "Music Recommender System", Master of Science in Computer Engineering, Politecnico di Milano, Facoltà di Ingegneria dell'Informazione, 2011.

[4] F. Ricci, Recommender systems handbook. New York: Springer, 2011.

[5] P. Rentfrow and S. Gosling, "The do re mi's of everyday life: The structure and personality correlates of music preferences.", Journal of Personality and Social Psychology, vol. 84, no. 6, pp. 1236-1256, 2003.

[6] "How music recommendation works – and doesn't work," Brian Whitman. [Online]. Available: http://notes.variogr.am/post/37675885491/how-music-recommendation-works-and-doesnt-work. [Accessed: 07-Apr-2016].

[7] M. Slaney, "Web-Scale Multimedia Analysis: Does Content Matter?", IEEE Multimedia, vol. 18, no. 2, pp. 12-15, 2011.

[8] J. Bobadilla, F. Ortega, A. Hernando and J. Bernal, "A collaborative filtering approach to mitigate the new user cold start problem", Knowledge-Based Systems, vol. 26, pp. 225-238, 2012.

[9] V. Rohani, Z. Kasirun, S. Kumar and S. Shamshirband, "An Effective Recommender Algorithm for Cold-Start Problem in Academic Social Networks", Mathematical Problems in Engineering, vol. 2014, pp. 1-11, 2014.

[10] A. van den Oord, S. Dieleman, B. Schrauwen, ”Deep Content-based Music Recommendation”, Advances in Neural Information Processing Systems 26, 2013.

[11] J. Ritchie and J. Lewis, "Qualitative Research Practice: A Guide for Social Science Students and Researchers", Choice Reviews Online, vol. 41, no. 03, pp. 41-1319-41-1319, 2003.

[12] B. Leech, "Asking Questions: Techniques for Semistructured Interviews", Political Science & Politics, vol. 35, no. 04, pp. 665-668, 2002.

[13] A. Cooper, The Inmates Are Running the Asylum. Indianapolis, IN: Sams, 1999.

[14] F. Isinkaye, Y. Folajimi and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation", Egyptian Informatics Journal, vol. 16, no. 3, pp. 261-273, 2015.

[15] T. Dunning and B. Friedman, Practical machine learning. Sebastopol, CA: O'Reilly Media, 2014.

[16] A. DIELEMAN, "Recommending music on Spotify with deep learning", 2016. [Online]. Available: http://benanne.github.io/2014/08/05/spotify-cnns.html. [Accessed: 23- Apr- 2016].

[17] O. Herrada, "Music Recommendation and Discovery In The Long Tail", Ph.D, Universitat Pompeu Fabra in Barcelona, Spain., 2008.

[18] Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez, "Recommender systems survey", Knowledge-Based Systems, vol. 46, pp. 109-132, 2013.

[19] M. Pazzani, ”A Framework for Collaborative , Content - Based and Demographic Filtering”, Artificial Intelligence Review vol. 13 (5-6), pp. 393-408,1999.

[20] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, ”Recommender Systems: An Introduction, vol 5, pp 128-139, 2011

[21] "Maddieman", Last.fm, 2011. [Online]. Available: http://www.last.fm/user/Maddieman/journal/2011/09/21/4oqdps_beginner's_guide_to_last.fmfor. [Accessed: 05-Mar-2016].

[22] J. Benjamin W. McGruder and J. Kingma, "Remember Last.fm? A Fresh Look at the Redesigned Music Service", MakeUseOf, 2015. [Online]. Available: http://www.makeuseof.com/tag/remember-last-fm-look-redesigned-music-service/. [Accessed: 13-Apr-2016].

[23] "What's new — Last.fm", Last.fm, 2016. [Online]. Available: http://www.last.fm/about/whatsnew. [Accessed: 30-May-2016].

[24] "What is Pandora?", Pandora, 2016. [Online]. Available: http://help.pandora.com/customer/en/portal/articles/182180-what-is-pandora-. [Accessed: 07-Jun-2016].

[25] L. Grossman, "How Computers Know What We Want — Before We Do", TIME.com, 2010. [Online]. Available: http://content.time.com/time/magazine/article/0,9171,1992403-1,00.html. [Accessed: 02-May-2016].

[26] "What is Pandora?", Pandora, 2016. [Online]. Available: http://help.pandora.com/customer/en/portal/articles/182180-what-is-pandora-. [Accessed: 11-Mar-2016].

[27] "Create a Station", Pandora, 2016. [Online]. Available: http://help.pandora.com/customer/en/portal/articles/84831-create-a-station. [Accessed: 01-Jun-2016].

[28] "Apple iTunes - Gracenote", Gracenote, 2016. [Online]. Available: http://www.gracenote.com/project/itunes/. [Accessed: 22-Apr- 2016].

[29] L. Barrington, R. Oda, and G. Lanckriet, Cognitive Science, October, Number Ismir," Smarter Than Genius? Human Evaluation of Music Recommender Systems", p.357–362, 2009.

[30] L. Yamshon and L. Yamshon, "Apple Music FAQ: The ins and outs of Apple's new streaming music service", Macworld, 2016. [Online]. Available: http://www.macworld.com/article/2934744/software-music/apple-music-faq-the-ins-and-outs-of-apples-new-streaming-music-service.html. [Accessed: 07- Jun- 2016].

[31] "Music - Discover", Apple, 2016. [Online]. Available: http://www.apple.com/music/discover/. [Accessed: 01-May-2016].

[32] J. Dalrymple, "How to properly use “Likes” in Apple Music", Loopinsight.com, 2015. [Online]. Available: http://www.loopinsight.com/2015/07/02/how-to-properly-use-likes-in-apple-music/. [Accessed: 27- Apr-2016].

[33] "Discover - Spotify", Spotify, 2016. [Online]. Available: https://support.spotify.com/dk/using_spotify/discover_music/discover/. [Accessed: 13-Apr- 2016].

[34] "Discover Weekly - Spotify", Spotify, 2016. [Online]. Available: https://support.spotify.com/dk/using_spotify/playlists/discover-weekly/. [Accessed: 01-Jun- 2016].

[35] D. Etherington, "Spotify Acquires The Echo Nest, Gaining Control Of The Music DNA Company That Powers Its Rivals", TechCrunch, 2016. [Online]. Available: http://techcrunch.com/2014/03/06/spotify-acquires-the-echo-nest/. [Accessed: 29-Mar-2016].

[36] A. Pasick, "The magic that makes Spotify’s Discover Weekly playlists so damn good", Quartz, 2016. [Online]. Available: http://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good/. [Accessed: 28-May- 2016].

[37] "Fresh Finds - Spotify", Spotify, 2016. [Online]. Available: https://support.spotify.com/dk/using_spotify/for_artists/fresh-finds/. [Accessed: 25-May- 2016].

[38] T. Team, "Introducing Fresh Finds", News, 2016. [Online]. Available: https://news.spotify.com/us/2016/03/02/introducing-fresh-finds/. [Accessed: 20- May-2016].

[39] "Spotify Radio - Spotify", Spotify, 2016. [Online]. Available: https://support.spotify.com/dk/using_spotify/search_play/spotify-radio/. [Accessed: 25- Apr-2016].

[40] "SoundCloud | What is SoundCloud?", Help.soundcloud.com, 2016. [Online]. Available: http://help.soundcloud.com/customer/en/portal/topics/869303-what-is-soundcloud-/articles?b_id=10674. [Accessed: 22-Apr-2016].

[41] E. Buskirk, "SoundCloud Threatens MySpace as Music Destination for Twitter Era", WIRED, 2016. [Online]. Available: http://www.wired.com/2009/07/soundcloud-threatens-myspace-as-music-destination-for-twitter-era/. [Accessed: 07- Jun- 2016].

[42] A. Munassar, A. Govardhan and N. Mohammed, "A Comparison Between Five Models Of Software Engineering", Ph.D, Principal JNTUH of Engineering College, Jagityal, Karimnagar (Dt), A.P., India, 2010.

[43] "Guideline: Detail Use Cases and Scenarios", Epf.eclipse.org, 2016. [Online]. Available: http://epf.eclipse.org/wikis/openup/practice.tech.use_case_driven_dev.base/guidances/guidelines/detail_ucs_and_scenarios_6BC56BB7.html. [Accessed: 10- Jun- 2016].

[44] "Defining Software Architecture", *Software Engineering Institute, Carnegie Mellon University*, 2016. [Online]. Available:

http://www.sei.cmu.edu/architecture/. [Accessed: 03- Jun- 2016].

[45]"Reuse in Use-Case Models: <<extend>>, <<include>>, and Inheritance", *Agilemodeling.com*, 2016. [Online]. Available: http://agilemodeling.com/essays/useCaseReuse.htm. [Accessed: 10- May- 2016].

[46] J. Doe, "Recommended Practice for Software Requirements Specifications", 1998.

[47] K. Siau and Q. Cao, "Unified Modeling Language", *Journal of Database Management*, vol. 12, no. 1, pp. 26-34, 2001.

8 Appendix

8.1 Semi-structured Interview Questionnaire

1. Briefly say something about yourself (e.g. age, your background).
2. Have you ever used MRS?

   a. Yes: Mention some of the them.

   b. No: What about iTunes, Spotify, YouTube or something similar to them.
3. What do you expect from a MRS?
4. Tell me something about your experience and interaction. (If you are satisfied with the recommendation results).
5. How often do you choose the recommendation offered to you?
6. Do you find music recommendation useful for you?
7. Do the recommendations always match your music preferences?
8. Can you tell me an example?
9. What type of recommendation results do you usually get? (Popular, demographic, artist-based, song-based, genre-based).
10. New How fair do you think the recommendation results are? Do you think that they are advertisement – based?
11. If you use more than one MRSs, which one you think is more effective?
12. Which your favorite feature in RS?
13. Do you think RS help you discover new music?
14. Do you think that MRSs lack the ability to recommend new coming music and artists?
15. How can an MRS be improved in order to recommend new songs or artists?

8.2 Interviews

8.2.1 Summary of the First Interview

First interview was with Giannis Moustakas. He is studying computer science in Greece and he is familiar with music recommendation systems.

Giannis has been mostly used iTunes platform the last ten years, and also tried Spotify, Pandora Radio as well as Shazam. iTunes has been the main platform which he uses quite often. He likes the search engine where he can search artists, songs in order to see most of their works. According to him iTunes helps him find similar artist or music by using Genius Music recommendation system. He believes that the the particular RS is quite useful. Most of the time it recommends music that he has heard before, but also some new music, even though it is less likely. However, many times he thinks that the RS lacks some depth in its recommendation because it recommends only the most popular songs or artist. He thinks that most of the time the recommendation matches his preferences, but the many times it doesn't recommend new music and as a result he has to search for them manually.

According to him Spotify is one of the best platform that has the best recommendation system among other platform like iTunes, Pandora Radio as well as Shazam. He believes that the recommendations are more accurate. He believes that MRS are useful and productive to find new music. When he searches for example a particular artist or genre, the system can sometimes offer good recommendation. He thinks that MR has be more user-centered, and user should write his preferences such as the artists, genres, area etc. as an input for the system to provide better RS.

Giannis has use Pandora before. According to him Pandora radio has the very unique characteristic that it doesn't recommend any music listened to by other users with similar tastes. It uses only the personal preferences of the listener to recommend music that is tagged and categorized from Pandora employees only. The user's interaction with the application or website is utilized to prompt more accurate recommendations. The more he used the service, the more enjoyable the station's choices are and therefore the most successful the RS is in his opinion.

He has also used Shazam a couple of years ago. According to added some options, aside its primary use, which included music related to what he searched in the first place for. It is a somehow useful and nice feature, but it seems lacking a personality based on the user, and recommends only music that is related to one search at a time. He found out that it is very easy to use, mostly because it's shown in every search Shazam returns with a music track.

8.2.2 Summary of the Second Interview

Second interview is with Giorgos Moustakas. He is student in the University of Aegean and studying in the department of information and communication engendering. He has a big passion for computers and has chosen Computer Science and Engendering as a subject for his future career. He is also a developer and has developed apps for iOS devices. Moreover, he is having an enthusiasm about security and system administration.

He has used MRS such in the past such as iTunes (Apple Music), SoundCloud as well as Spotify. He has not used Pandora and Last.fm because the services are not available in Greece.

According to him, Apple Music does not have many categories for electronic music in particular (which he likes). He thinks that Apple music have done so little to accomplish in order to provide better recommendation. The system lacks the ability to find new songs, but matches the results according to the genre which makes it too general for him. So he prefers his recommendations to be more specific. They have a lot of room for improvement.

Spotify on the other has a better recommendation system. When you listen to a track, it pops up other songs which matches the artist mostly and the genre in particular. In his experience Spotify is way better than Apple Music when it comes to MR. His favorite feature is the Radio feature on Spotify, which plays music based based on the artist, but the recommendations are better and more personal than Apple Music.

He also has used SoundCloud. He thinks has that SoundCloud has the best RS from all the other services. Recently he was listening to a podcast from an electronic music artist called Mat Zo and while he was listening to the podcast, other songs were popped that he loved very much. SoundCloud has also a follow button which helps the listeners follow artists that pop up. After following these artist, the results from the RS became better. So he thinks the the recommendations were right to the point. Even though SoundCloud lacks some features like the ones in Spotify, it still recommended new music. Moreover, SoundCloud even showed recommended him unreleased songs, which he finds it great.

As a music enthusiast he wishes from a MRS to recommend new music, especially new artist from his preferred genre, electronic music. For him the ideal MRs would be the system to analyze the sound itself such as the melody, BPM, rhythm rather than recommend music according to the metadata. SoundCloud has done some improvements, but believes it has potential to do much more.

He thinks that one big disadvantage of MRS are its recommendations are basically advertisements. They are used in order to make more money, which means they compromise on the quality of the recommendations.

8.2.3 Summary of the Third Interview

Third Interview was from a Greek girl, age 32 years old. She lives in CPH for 3 years and works in Nordea in the Capital Market Services department.

She has used in the past Spotify and YouTube. She listens to music on YouTube very often. The only recommendation feature that she has encountered is Related Videos, which is located on the right side of YouTube's which are recommendation that YouTube recommends her to see.

She also uses Spotify to listen to music as well. She occasionally uses the feature called Discovery which is a part of the Spotify platform that shows recommendations such as new songs and artists. This is the relationship she has with MRS.

She doesn't use the recommendation feature very often. She goes there when she has free time and she want to listen to music and discover for example a new track from her existed favorite artist as well as songs from new artists.

She really likes music covers in particular, covers of her favorite songs and artist. Spotify has helped her to find new covers. Spotify has a big database and sometimes the recommendation has helped her find an interesting track.

She expects from a MRS to find good music. She does not necessarily expect from it to recommend the most popular and famous songs. She believes that MRSs are not innocent, because they can probably recommend something that might have an agreement with the artist, something that suits her preferences, but it also has payed Spotify in order to be advertised, but she does not like that. She wishes to listen to new music and discover songs, artists as well.

In general, she thinks that a MRS in Spotify matches her preferences. For example, Spotify has a particular playlist in the discover section. When she plays this playlist that they suggest to her, it matches her preferences and sometimes she is positively surprised. She thinks that Spotify knows her sometimes and recommends to her new songs, which she likes. Therefore, she believes that MRSs are fair in regards to her preferences, but they can be better because they sometimes recommend songs and artist that she doesn't necessarily like. So she thinks that there are some gaps there.

Her favorite recommendation feature in Spotify is Your music feature that tells her what she heard before… history of her previous songs. Also, the discover. When for example she goes to Browse, then Discovers, she sees Top Recommendation for her. And she sees recommendation similar to the artists that she has listened before. She also checks on sometimes the new releases. But in the new releases there are a lot of ads, and doesn't mate her preferences. There are some Danish artists which she doesn't listen to, i.e. Drake which she also doesn't like and David Guetta which again she's not crazy about.

She believes that the MRS in Spotify doesn't represent her preferences. If she for example goes to Discover, she can see that they have rec. based on her previous preferences, but the new releases are not based on her preferences at all. So new releases feature is just rather a feature showing music in general, music that is advertised, local artists which she's not fan of. So, she doesn't think that new released feature recommends new artist.

In the question, what would you like for a MRS to do and how can be improved, she states that she hates the commercialization of platform. She would like to have more personal and more customized use. She would really like for it to know her better maybe. And also she would like some new rec. like new releases feature, but to have new releases matching my preferences, rather than random ones.

8.2.4 Summary of the Fourth Interview

The fourth interview is with a Greek man that works as a developer. He has finished the studies in Aalborg University CPH and is familiar with RS because part of his master thesis was using a method, more specifically a collaborative filtering method to develop a specific system.

He has used a few music platforms such as iTunes and Spotify. He believes that those platforms helped him find music that he likes mostly, based on his profile or similar taste with other users, and they are doing a great job on it in general.

He expects the recommendation to be really close to his taste of music. He also expects not to do a lot of procedure, fill up his profile, give the least information to them and recommends him relevant music according to his preferences.

He follows recommendation pretty often, especially in iTunes. He like listening to music, so he goes to iTunes in order to find different new artists and songs. Most of the time MRS matches his preferences. But he's looking for new music, he expects more serendipity, because he believes there is a lack of serendipity in general or that the system lacks the ability to find new music. For example, one of the issues is when someone has some specific preferences for a specific category of music and he want to discover new artist and songs from this specific category. He refers to the new artist, not the popular ones. Most of The RS have problems dealing with recommending new artist because not enough information or metadata are available for the RS in order to recommend a specific artist or song. There are not enough data because let's say in collaborative filtering information is needed from the other users, but since it's a new artist there is nothing, not rating, so it's difficult for the RS to add this artist to the database and recommend it to you. This problem is called the cold start problem.

There is much more to do in this direction, of course there are RS that try to address this problem. In an article he read, Pandora is doing something similar. They are trying somehow to gather info about new artist and songs and introduce them to its database.

In the question, how can recommendation systems be improved, he adds that there is a lot of information and new technologies are coming up. For a platform to have a good recommendation it should collaborate with other platforms. There are some platforms for example specifically for new artists and they can get feedback from users like SoundCloud. They have a really nice feature. For example, when an artist uploads his new song there, users can go to the streaming song and put their comments and say that they like a particular part of music. So there is information from the users for that specific songs.

He believes that the solution would be two platform collaborating together to somehow use this information (information based on user feedback) from SoundCloud in order to make better recommendation for new upcoming artist and songs.