The Data

Our dataset is a Spotify dataset containing over 160,000 songs from 1921 to 2020. [Reference]

For each song in the dataset there are nineteen different features, split into four groups by the type of value they contain. Four of the features are categorical, for example the name of the song and the artist. Two of the features are dummies, for example whether the song has explicit content. With this feature a song has a zero here if it does not contain explicit content and a one if it does contain explicit content. The third type contains only one feature that does not fit anywhere else, partly because this feature is how all the other features are recognised as belonging to the same song within Spotify’s database. This feature is the Spotify song ID. The final type is where the majority of the features are, these are all the features that take numerical inputs. There are a range of features in here, from track length measured in milliseconds to danceability. As are a few other features, the danceability of a sing is defined by Spotify as a number between zero and one, where zero is no danceability and one is lots of danceability. This number is made by considering a range of features in the music including the tempo (speed) of the music, the strength of the beat and how regular the song is. However, we do not have access to these features so will just be using their combination found under danceability.

Although this dataset contains nineteen features we will not be using all nineteen within our project. [Enter here some of the features we don’t use]

This data is fairly accurate. It has been collated from data available from Spotify developers, which is officially connected with the Spotify brand. Although it should be noted that due to the lengthy request process to see this data we cannot confirm that the data is the same as that take from Spotify developers. Another issue is that of Spotify itself, as our songs range from 1921 to 2020, however Spotify only started in 2008, which increases the chances of any songs from before 2008 having an incorrect year as there were not uploaded as they were released.

Project Aim

The overall aim of our project is to create a query answering system that allows users to search for Spotify songs within certain parameters that they input, such as how “dancey” they want the music to be or what genre it is in.

Objective 1:

The first objective is to allow users to find songs within their chosen range for a chosen feature. For example, the user may ask for songs that are more “dancey” and ask for songs that have a “dance” rating between 7 and 10. The program will then output all songs that fall in that range. It will also be possible to build on this, so it will be possible for the user to search like this for as many features as they would like from the available list.

Objective 2:

Our second objective is to allow users to find songs similar to one they input. For example, the user may ask for songs similar to “Danny Boy”, the program will output songs that have similar characteristics to “Danny Boy” in order of how similar they are. It will be possible to additionally place restrictions as in the above objective, for example explicit songs can be removed.

Objective 3:

Our final objective is for users to be refine their list of songs by genre. For example, if the user has a list of songs similar to “Danny Boy” they would be able to further refine this by asking for only songs that come under the “Pop” genre. As genre is not one of our features in the given dataset this will be done by defining what properties songs in each genre have and separating them manually.

**System Design**

Architecture

100 words describing what happens to the data. Use a diagram.

Algorithms

4~ bullet points with the key coding functions of the system. 1-2 sentences describing each one.

**Conclusion**

Achievements (200 words)

Our program successfully…..

Limitations (100 words)

This project was limited by the dataset and…

Future Work (100 words)

In future work we would like to gain a wider range of features on each song and…