Comparing Modern Music Searching to the Text Retrieval Processes

25 October 2020
Bob Manasco
manasco2@illinois.edu

Internet text search engines have existed for as long as the internet itself. The first modern recognizable text search engine (Archie) was created in 1990 by a student at McGill University in Montreal (Wall, n.d.). At Archie's time, the web was so small, all web pages could be searched manually, but with the explosive growth in popularity over the early-to-mid 1990s, several other search engines followed that utilized more advanced algorithms: "bag of words" concepts, inverted indexing, and Vector Space Model processing. As technology and pattern recognition concepts advanced, new types of cutting-edge searching has been developed, including audio and music searching. Instead of searching text for specific words, music search engines can search sounds to find similar or matching sounds, useful for identifying songs, for instance. While differences do exist, there are many similarities to text search engines. The process of "fingerprinting" an audio file is a method of compressing data, not dissimilar to treating documents as "bags of words" and indexing. Some music search engines use audio fingerprints and the Vector Space Model, similar to text retrieval methods. There are even engines that now combine these audio searching methods with more-traditional text retrieval to create a hybrid engine, capable of searching for musical and lyrical styles.

Modern search engines treat queries and documents as "bags of words" and then process these with inverted indexing algorithms to compress the amount of data needed to search. In audio searching, a similar principle is applied, but is called fingerprinting. This is a compression technique that drastically reduces the size required to search. Using a Discrete Fourier Translation, the signal is given a Fourier analysis, which "converts a finite list of equally spaced samples of a function into the list of coefficients of a finite combination of complex sinusoids, ordered by their frequencies, by considering if those sinusoids had been sampled at the same rate" (Jovanovic, 2015). Using a sliding window of time, and filtering for only the frequencies important to the listener (for example, approximate ranges for bass guitar, voice, and other instruments), the song is reduced to a collection of highest-magnitude frequencies. This is sometimes referred to as a "bag of audio words" (Riley et al., 2008), very similar to the inverted indexing used by text search engines.

The next challenge facing an audio search engine is how to match a query with an audio file. Unsurprisingly, the solution is also fairly similar to how many text retrieval engines handle the problem: using established similarity calculations from the Vector Space Model (Riley et al., 2008). One possibility is using the approximate nearest neighbor to determine similarity. Due to high dimensionality, recognizing the pattern by finding the actual nearest neighbor to a sample can be problematic. This is caused by a circumstance sometimes referred to as the "dimensionality curse phenomenon," which describes the situation where distances between points very near to each other and points very far

away from each other become almost equal when considering many dimensions. Therefore, nearest neighbor calculations have difficulty when attempting to discriminate candidate points. Instead, there are several efficient algorithms for finding *approximate* nearest neighbor (Miller et al., 2005). By finding approximate matches, this has the additional advantage of not only bypassing the high-dimensionality nearest neighbor problem, but also allowing matches with songs that are very similar, although not exactly the same, such as matching a live recording with the studio version or matching a remixed rendition of a song with the original (Riley et al., 2008).

There are even modern search engines that now combine this type of audio searching with more traditional text retrieval. Qwant is a French search engine that recently released Qwant Music, a dedicated search engine specifically for music searching. It uses artificial intelligence and machine learning to retrieve many pertinent secondary types of information (artist discography, current news, touring schedule, photos, videos, etc.) when performing a music search (Su, 2018). In addition to fingerprinting the audio file, as described above, Qwant also has the ability to automatically create a "mood fingerprint" based on a song's tempo, complexity, or percussivity, which allows users to, for example, find "… rock songs that are complex and upbeat but with love lyrics to create a truly personalized playlist" (Su, 2018). Even internet search giant Google is experimenting with audio music searching. They have recently released the capability to search based on a user's hum or whistle of a tune, called "Hum to Search" (Adams, 2020). This is a different type of searching from just listening for an exact replica of a song, and it uses machine learning to retrieve potential matches based on the hummed audio. Google has developed and trained models to pinpoint matches on multiple sources, including original recordings but also recordings of other users singing, whistling, or humming popular tunes. "Hum to Search" is available not only in Google search widget, but it also directly interfaces with the popular voice-activated Google Assistant. After searching for specific music, Google will provide a series of most likely options based on the hummed sample. Then, users can choose to play these closest matches as well as peruse information related to the performing artists, tracks, albums, and more, similar to Qwant. Using a user's own hum, whistle, or singing to search for music has been around since as early as 2009 (for instance, in the SoundHound app) (Conner, 2020), but now Google combines this with the ability for traditional text searching of lyrics.

In conclusion, we see that while audio music search engines cannot behave exactly the same as text search engines, due to the differing source data, there are some striking similarities. Both have a need to compress files before searching, to improve performance. Both utilize a type of inverted indexing, one based on "bag of words," the other based on "bag of audio words." Both even utilize rough similarity matching, utilizing approximate nearest neighbors in the Vector Space Model. Finally, it is exciting to see the two technologies working in coordination to allow users to search in ways not possible before now. It seems that audio and music retrieval will be experiencing massive growth in the next few years, like text search engines in the mid-to-late 1990s, and machine learning will be enabling the new discoveries and opportunities as they grow.

References

Adams, R. D. (2020, October 16). Google announces "hum to search" machine learning music search feature. Retrieved November 02, 2020, from https://www.techrepublic.com/article/google-announces-hum-to-search-machine-learning-music-search-feature/

Conner, K. (2020, October 21). Google has a new hum-to-search feature for your phone. Here's how it works. Retrieved November 02, 2020, from https://www.cnet.com/how-to/google-has-a-new-hum-to-search-feature-for-your-phone-heres-how-it-works/

Jovanovic, J. (2015, February 02). *How does Shazam work? Music Recognition Algorithms, Fingerprinting, and Processing*. Retrieved October 17, 2020, from https://www.toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition

Miller, M. L. (1), Rodriguez, M. A. (2), & Cox, I. J. (3). (2005, November). Audio fingerprinting: Nearest neighbor search in high dimensional binary spaces. *Proceedings of 2002 IEEE Workshop on Multimedia Signal Processing, MMSP 2002*, 182–185. https://doi-org.proxy2.library.illinois.edu/10.1109/MMSP.2002.1203277

Riley, M., Heinen, E., & Ghosh, J. (2008). *A text retrieval approach to content-based audio retrieval*. 295–300.

Su, J. (2018, March 14). *SXSW: Qwant Music Is World's First Music Search Engine Matching Songs Based On Lyrics And Melody*. Forbes. Retrieved October 25, 2020, from https://www.forbes.com/sites/jeanbaptiste/2018/03/14/sxsw-qwant-music-is-worlds-first-music-search-engine-matching-songs-based-on-lyrics-and-melody/

Wall, A. (n.d.). History of Search Engines: From 1945 to Google Today. Retrieved October 25, 2020, from http://www.searchenginehistory.com/