Using Text-Based Information Retrieval Methods with Audio Data

D. Scott McCaulay
smccaula (at) indiana.edu

Traditionally the focus of Information Retrieval has been on text documents. As the online availability of non-text data such as music and video continues to grow, there is a recognized need for automated tools to organize and assist in the retrieval of this data. This study looks at the novel concept of using established text IR methods for the classification and retrieval of digital audio data. Specifically it proposes an automated content analysis process which could assist in the classification of audio files as a complement or alternative to the current common practice of manually adding metadata to organize digital audio collections. In this proposed automated process, features extracted from the audio files, for example the occurrence of a specific frequency at a specific amplitude, take the place of words in a text document as terms for classification purposes. Similarity between audio files is calculated based on the term frequency of these extracted features within a given audio file and the inverse document frequency of the same term in the population of audio files being evaluated.

This study analyzes a collection of 600 MP3 audio files representing 20 songs each by 30 selected artists, and uses feature extraction, TF/IDF and cosine similarity to calculate a measure of similarity between each of the files. The artists were selected to cover a range of time periods and musical genres. The expectation that audio files by the same artist will show a significantly higher level of similarity than files by different artists is supported by the results of this experiment. The experiment takes a fusion approach to feature extraction, extracting four sets of features representing different aspects of frequency and dynamics of the audio and calculating similarity on a weighted combination of the results from the different feature sets. The application envisioned for this methodology is the generation of a playlist of audio files classified as "similar" to a user-supplied seed audio file. In other words, a QBE application where the query is a selected audio file, and the result set is a ranked list of audio files retrieved from a library based on a calculation of similarity to the query file.

References:

Tzanetakis, G. and Cook, P. (2002). "Musical Genre Classification of Audio Signals." IEEE Transactions on Speech and Audio Processing 10(5): 293-302.

Downie, J. S. and Nelson, M. (2000). "Evaluation of a Simple and Effective Music Information Retrieval Method." Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval: 73-80.