

SORTMYTUNES – FAQ (APRIL 2008)

WHAT IS SORTMYTUNES?

SortMyTunes is a **fully customizable system for Personal Music Classification (PMC)**. PMC is defined as the storage, recollection and organization of any isolated collection of music. It can be an online or an offline collection, as long as the collection is relatively static and completely accessible.

WHY IS PERSONAL MUSIC CLASSIFICATION (PMC) IMPORTANT?

Most collections of music can be considered a Personal Music System. Anything from a personal hard drive full of music listened to via a computer MP3 player, to the vast databases of music on sites like last.fm or Pandora. Additionally, PMC plays a very important role in academic work, such as the MIREX contests (where PMCs are used as ground truth). The ability to organize, classify and recall these databases quickly, efficiently and in a customizable fashion is a fundamental problem in gaining a further understanding of the musical composition of the collection.

HOW DOES SORTMYTUNES WORK?

SortMyTunes allows the user to classify her system into groups, known as "pods". These pods can be infinitely variable in size (tracks within each pod) and quantity (number of total pods). **The user first manually classifies a limited amount of music into the pods, and then SortMyTunes takes a percentage of the database, looks at the metadata associated with each incipit track, and classifies it into the most appropriate pod.** The user makes any manual modifications to the resulting classification (such as adding/removal of pods, re-classification of tracks due to error), and repeats the process on another portion of the database. This iterative process continues until the database is fully classified.

WHERE DOES THE METADATA COME FROM?

Metadata can be harvested from sources like **last.fm or musicbrainz**, or it can be manually created in some circumstances.

WHAT ARE "TAGS"?

Tags are the fundamental building block of the SortMyTunes metadata. **The metadata associated with each track is simply a list of tags that describe the track.** This gives the algorithm a basis for comparison among tracks in the database.

IS THE ACTUAL AUDIO DATA REALLY NECESSARY?

No. This actually streamlines the entire process, allowing the bulky audio data to be handled separately during the cumbersome sorting/searching/inserting/deleting that takes place during the classification process.

HOW LARGE A COLLECTION CAN SORTMYTUNES HANDLE?

There is no size limit, except for the standard practical limitations (HD space, processor speed, bandwidth, etc).

WHAT IS THE EFFICIENCY OF SORTMYTUNES?

Using the *k-means* classification algorithm, SortMyTunes currently operates at a **Big O efficiency of n^2** . Research is currently underway to get this efficiency down by an order of logarithmic magnitude.

WHERE CAN I GET AN ONLINE COPY OF THESE NOTES?

<http://www.martinmccrory.com>