

Gordon Music Collection Database Management System

Douglas Eck
Université de
Montréal

Jorge Orpinel
New York
University

Ron J. Weiss
New York
University

Juan Pablo Bello
New York
University

We present Gordon, a multi-platform database tool for managing very large music collections, targeted toward music information retrieval research. Gordon contains a unified framework for managing audio files in any format, matched with corresponding artist, album, and track metadata which can be resolved against the MusicBrainz service. Also supported are per-track annotations, useful for storing e.g. reference chord transcriptions alongside the corresponding audio. Currently in development is a framework for feature extraction and caching.

Gordon can be used as a central repository for MIR datasets within a research group and greatly simplifies many common tasks in configuring and running experiments. Using Gordon, there is no need to worry about managing a haphazardly organized collection of files, audio I/O, or maintaining consistent feature extraction. Evaluating an algorithm on a different data set can be as simple as changing a single line of code once the data have been imported into the database.

The package is written in Python and includes a database backend based on SQLAlchemy, a simple web application for interacting with the database, and a more powerful Python API. The source code is licensed under the GNU GPL and is available on the bitbucket open source repository at <http://bitbucket.org/ronw/gordon>.