Music Data Similarity (MDS) Browser Readme

Author: Andrew Silverman
Date: 15-June-2017

Table of Contents

1 Introduction

The MDS Browser is a sample application that loads files from Gracenote data exports and allows the user to interact with the data. It is primarily intended to serve as an example implementation of the similarity algorithm. The command-line application is written in C, and the source code and libraries can be used and modified by customers as needed. Although the application is primarily intended to help familiarize users with the data and the similarity algorithm, the libraries can also be used in other applications.

1.1 System Requirements

The source code can be compiled using a C compiler like Clang or gcc on any compatible operating system. The application itself does not take up much disk space, but the system running the application must have enough memory in which to load the data files. The configuration can be modified to load all files or only a subset depending on what functionality is needed. Please see below for configuration details.

1.2 Installation

To build the mdsbrowser application, you can simply use the Makefile that is included by running make:

```
$ make
cc -Wall -DBUILDDATE="\"Thu, Jun 15, 2017  3:51:14 PM\"" -
DBUILDHOST="\"ASILVERMAN-6065\"" -g -c gmd_json.c
cc -Wall -DBUILDDATE="\"Thu, Jun 15, 2017  3:51:14 PM\"" -
DBUILDHOST="\"ASILVERMAN-6065\"" -g -c jsmn.c
cc -Wall -DBUILDDATE="\"Thu, Jun 15, 2017  3:51:15 PM\"" -
DBUILDHOST="\"ASILVERMAN-6065\"" -g -c goet_tables.c
cc -Wall -DBUILDDATE="\"Thu, Jun 15, 2017  3:51:15 PM\"" -
DBUILDHOST="\"ASILVERMAN-6065\"" -g -c main.c
cc -Wall -DBUILDDATE="\"Thu, Jun 15, 2017  3:51:15 PM\"" -
DBUILDHOST="\"ASILVERMAN-6065\"" -g -o mdsbrowser goet_tables.o main.o
gmd_json.o jsmn.o
```

You can also build manually like so:

```
$ cc -Wall -g -c goet_tables.c
$ cc -Wall -g -c jsmn.c
$ cc -Wall -g -c gmd_json.c
$ cc -Wall -g -c main.c
$ cc -Wall -g -o mdsbrowser main.o gmd_json.o jsmn.o goet_tables.o
```

2 Configuration

The mdsbrowser requires a configuration file where you must specify the paths to files that the application reads into memory. The default config file is called mdsbrowser.config in the application's working directory. You can use the command line parameter -c followed by a file path to override where the program looks for the config file. This file can be used to set the format of the export files (tsv vs gmd) and also whether to load recordings and hierarchies. Note that the settings will be different depending on if you are using a GMD export vs a tsv-based export. Please see the example config files that are included with the source code for more details.

Example tsv-based data export config:

```
#
# Configuration file for MDS Browser
#

data_format           tsv
load_recordings       yes
load_hierarchies      yes

artist_file                   data/gracenote_na_radio_artist_20151104.tsv
artist_descriptor_file data/gracenote_na_radio_artist_descriptor_20151104.tsv
artist_popularity_file        data/gracenote_na_radio_artist_popularity_prod.tsv
#artist_popularity_file              data/pop/mids_lookups_356_0_uk.tsv

descriptor_file                     data/gracenote_na_radio_descriptor_20151104.tsv
descriptor_correlate_file     data/gracenote_na_radio_descriptor_correlate_20151104.tsv

recording_file                data/gracenote_na_radio_track_20151104.tsv
recording_descriptor_file     data/gracenote_na_radio_track_descriptor_20151104.tsv
recording_popularity_file     data/gracenote_na_radio_track_popularity_prod_20151104.tsv

hierarchy_type_file           data/gracenote_hierarchy_type_20151104.tsv
hierarchy_node_file           data/gracenote_hierarchy_node_20151104.tsv
hierarchy_node_language_file  data/gracenote_hierarchy_node_language_20151104.tsv
descriptor_to_hierarchy_node_file data/gracenote_descriptor_to_hierarchy_node_20151104.tsv

language_file                 data/gracenote_language_20151104.tsv
```

Example gmd style config:

```
#
# Configuration file for MDS Browser
#

data_format            gmd
load_recordings                yes

# only use one recording per song group, if yes will load song ids
one_recording_per_song yes

# hierarchies not supported from gmd at this time
load_hierarchies       no

artist_file                    data/GMD/20170527/gracenote_artist_20170527.json

album_master    data/GMD/20170527/gracenote_album_master_20170527.json
album_edition   data/GMD/20170527/gracenote_album_edition_20170527.json

descriptor_file                        data/GMD/20170527/gracenote_descriptor_20170527.json
descriptor_correlate_file      data/GMD/20170527/gracenote_correlate_20170527.json

recording_file                 data/GMD/20170527/gracenote_recording_20170527.json
```

Commenting out any of the lines containing a file path will skip loading that file at run time.

Note

Popularity files are not included as part of the standard data export. Please contact your Gracenote Engagement Manager to learn more about this feature.

3 Getting started

Open a terminal and run ./mdsbrowser to get started. The data files will be immediately loaded into memory. This may take up to 15 minutes depending on the settings in the mdsbrowser.config file.

```
$ ./mdsbrowser
Loading tables...
Loading artist GOET table 'data/gracenote_artist_descriptor_20140317.tsv'
Loaded 94118 artists, 501962 GOET values.
Loading artist names 'data/gracenote_artist_20140317.tsv'
Loaded 94118 artist names.
Won't load artist popularities.
Loading GOET table 'data/gracenote_descriptor_20140317.tsv'
Loaded 3560 GOETs.
Loading GOET correlate table 'data/gracenote_descriptor_correlate_20140317.tsv'
Loaded 864099 GOET correlates.
Max depth of correlate hash is 4.
Loading recording GOET table 'data/gracenote_recording_descriptor_20140317.tsv'
Loaded 28806868 recordings, 254634540 GOET values.
Loading recording metadata 'data/gracenote_recording_20140317.tsv'
Loaded metadata for 28806862 recordings.
Won't load recording popularities.
Adding recording links to artists.
Done adding recording links to artists.
Sorting artists and recordings.
Done sorting artists and recordings.
Loading hierarchies 'data/gracenote_hierarchy_type_20140317.tsv'
Loaded 20 hierarchies.
Loading hierarchy nodes 'data/gracenote_hierarchy_node_20140317.tsv'
Loaded 12679 hierarchy nodes.
Adding parent/child/sibling links to hierarchy nodes.
Done adding links to hierarchy nodes.
Loading descriptor to hierarchy node links
'data/gracenote_descriptor_to_hierarchy_node_20140317.tsv'
Loaded 36195 descriptor to hierarchy node links.
Loading languages 'data/gracenote_language_20140317.tsv'
Loaded 31 languages.
Loading hierarchy node localizations 'data/gracenote_hierarchy_node_language_20140317.tsv'
Loaded 393080 hierarchy node localizations.
>
```

Type "help" at any time to receive a list of available commands.

```
> help
   about -- Prints information on this program.
   search <name> -- Search artists by name.
   artist <Artist ID> -- Print artist's information.
   artistsim <Artist ID> <Artist ID> -- Print two artists' similarity score.
   searchsim <Artist ID> -- Search for artists similar to the one specified.
   goet <GOET ID> -- Print GOET's information.
   goetcorr <GOET ID> <GOET ID> -- Print two GOETs' correlation value.
   searchgoet <name> -- Search GOETs by name.
   goetart <GOET ID> -- Get artists for a GOET.
   rec <Recording ID> -- Print recording's information.
   searchsimrec <Recording ID> -- Search for recordings similar to the one specified.
   simthreshold [value] -- View or set similarity / GOET correlate threshold.
   popthreshold [value] -- View or set popularity threshold.
   maxres [value] -- View or set number of max search results displayed (0 = no limit).
   hiers -- View the list of available hierarchies.
   hier <HIER ID> [maxl] -- View the specified hierarchy. If specified, restrict output to maxl
levels.
   node <NODE ID> -- View the specified hierarchy node.
   langs -- View the list of supported languages.
   lang [LANG ID] -- View or set the display language for hierarchies.
   help -- This help text.
   quit or exit -- Exits the program.
```

The main entry point for exploration with the mdsbrowser is an artist. Use the search command to search with an artist string and the artist command to view a brief summary of all related artist data.

```
> search Jon Hopkins
Searching for Jon Hopkins...
    404140 Jon Hopkins
> artist 404140
Fetching artist 404140...
ID=404140
Name=Jon Hopkins
GOETs (10):
  1. Type=ATYPE, ID=23701, Weight=100, Name=Male Solo - Solo Instrumental
  2. Type=ERA, ID=2650, Weight=48, Name=2000's
  3. Type=ERA, ID=34808, Weight=26, Name=Mid 2010's
  4. Type=ERA, ID=34814, Weight=26, Name=Early 2010's
  5. Type=GENRE, ID=2861, Weight=29, Name=Ambient
  6. Type=GENRE, ID=2860, Weight=23, Name=Ambient Techno
  7. Type=GENRE, ID=2882, Weight=16, Name=IDM
  8. Type=GENRE, ID=2911, Weight=16, Name=Downtempo
  9. Type=GENRE, ID=3271, Weight=16, Name=Film Score
  10. Type=ORIGIN, ID=4579, Weight=100, Name=London
Recordings (195):
  1. Title=Second Sense, ID=2249467
  2. Title=Apparition, ID=13327440
  3. Title=The Low Places, ID=18954186
  4. Title=Contact Note, ID=1804680
  5. Title=Opalescent, ID=947019
```

Please see the data dictionary and diagrams for more information about the data models and hierarchies. Additionally, the Programmer's Guide to Gracenote Music Similarity document has much more detail about how the similarity commands are implemented.