Million Song Dataset

| Home | Getting the dataset | Code | Tutorial | Tasks / Demos | More data | Forum | FAQ | Contact / Cite | Blog |

Home » Frequently Asked Questions

Field list

Submitted by millionsong on Fri, 01/14/2011 – 14:37

Below are a list of all fields available in the files of the dataset. The same list with data from a specific song is available here. Another reference is the code: display_song.py: if a field is displayed, the field exists and there should be a getter for it (if we forgot some in matlab or java, please let us know).

For the analysis fields, we suggest you first read The Echo Nest analyze documentation. The main audio features are 'segments_pitches' and 'segments_timbre'.

Field name	Type	Description	Link
analysis sample rate	float	sample rate of the audio used	url
artist 7digitalid	int	ID from 7digital.com or −1	url
artist familiarity	float	algorithmic estimation	url
artist hotttnesss	float	algorithmic estimation	url
artist id	string	Echo Nest ID	url
artist latitude	float	latitude	
artist location	string	location name	
artist longitude	float	longitude	
artist mbid	string	ID from musicbrainz.org	url
artist mbtags	array string	tags from musicbrainz.org	url
artist mbtags count	array int	tag counts for musicbrainz tags	url
artist name	string	artist name	url
artist playmeid	int	ID from playme.com, or −1	url
artist terms	array string	Echo Nest tags	url
artist terms freq	array float	Echo Nest tags freqs	url
artist terms weight	array float	Echo Nest tags weight	url
audio md5	string	audio hash code	
bars confidence	array float	confidence measure	url
bars start	array float	beginning of bars, usually on a beat	url
beats confidence	array float	confidence measure	url
beats start	array float	result of beat tracking	url
danceability	float	algorithmic estimation	
duration	float	in seconds	
end of fade in	float	seconds at the beginning of the song	url
energy	float	energy from listener point of view	
key	int	key the song is in	url
key confidence	float	confidence measure	url
loudness	float	overall loudness in dB	url
mode	int	major or minor	url
mode confidence	float	confidence measure	url
release	string	album name	
release 7digitalid	int	ID from 7digital.com or −1	url
sections confidence	array float	confidence measure	url
sections start	array float	largest grouping in a song, e.g. verse	url
segments confidence	array float	confidence measure	url
segments loudness max	array float	max dB value	url
segments loudness max time	array float	time of max dB value, i.e. end of attack	url
segments loudness max start	array float	dB value at onset	url

News

April 25, 2012
The MSD Challenge has launched!

October 20, 2011
We release the Last.fm dataset of tags and similarity!

April 12, 2011
We release the musiXmatch dataset of lyrics!

March 15, 2011
We release the SecondHandSongs dataset of cover songs!

February 8, 2011
We release the dataset! (and get Dan to blog)

Quick links

LabROSA
The Echo Nest
Musicbrainz
Infochimps
7digital
Last.fm
musiXmatch
SecondHandSongs

Main contact

MSD mailing list

Random track

[new track]

Random ISMIR paper

[new paper]

Subscribe to MSD Mailing List

Email: [   ]

[Subscribe]

Search the site

(NOT THE DATA):

[   ] [Search]

AddThis

segments pitches	2D array float	chroma feature, one value per note	url
segments start	array float	musical events, ~ note onsets	url
segments timbre	2D array float	texture features (MFCC+PCA–like)	url
similar artists	array string	Echo Nest artist IDs (sim. algo. unpublished)	url
song hotttnesss	float	algorithmic estimation	
song id	string	Echo Nest song ID	
start of fade out	float	time in sec	url
tatums confidence	array float	confidence measure	url
tatums start	array float	smallest rythmic element	url
tempo	float	estimated tempo in BPM	url
time signature	int	estimate of number of beats per bar, e.g. 4	url
time signature confidence	float	confidence measure	url
title	string	song title	
track id	string	Echo Nest track ID	
track 7digitalid	int	ID from 7digital.com or –1	url
year	int	song release year from MusicBrainz or 0	url

Login to post comments