
musiclyrics_{*nlp*}*Documentation*

Release latest

Jan 23, 2022

CONTENTS

musiclyrics_nlp automates the creation of a corpus of music lyrics, instantiates classes for aggregating lyrics by genre or identified artists, and supports passing of these lyrics objects to nlp tasks including model training and testing.

individual lyrics files have the extension .lyr. Each file contains all the lyrics for one artist.

The genre_aggregator class organizes data for a particular genre, such as included artists and their song lists.

The musical_meg class is a virtual container for lyrics and statistics on those lyrics, and is instantiated using an instance of genre_aggregator. Musical_Meg (MM) objects have a generator-iterator to pass raw lyrics or lyrics and tags as TaggedDocuments, and can iterate all words in lyrics or filter based on tf*idf values.

musiclyrics_nlp has scripts and methods to pass MM objects to several nlp tasks. the MM class has instance methods to get word count, word frequency, gensim Dictionary, bag of words and tf*idf values. MM can be passed to a method to display tf*idf plot for each artist.

With the following model training, there are methods to split lyrics into train and test. MM objects can be passed to generate a gensim word2vec model MM objects can be passed to generate a gensim doc2vec model, there are built in scripts to test the model for familiarity with unseen lyrics, and record and plot the results. MM objects can be passed to create a phrases model and analyze top phrases MM objects can be passed to create an LDA topic model, there are scripts and methods to analyze topics and terms and plot results.