

README File

Mingfei Yang (Group 7)

CSE 515 - Group Project Phase 1

Repository Structure

Code directory - All python source code

Outputs directory - All output csv files, contains results for sample inputs

Report directory - Project phase report, with docx format and pdf format

phase1_dataset directory - All the given data sets

Code Directory - Source code

** All Written in Python 3.6.2 Environment

FunctionCaller.py - Main entry of the program/application, please run this file

HelperFunctions.py - Wrapped in a number of functions to reduce code duplication

ActorToTag.py - The functioning python file for Task 1

GenreToTag.py - The functioning python file for Task 2

UserToTag.py - The functioning python file for Task 3

GenreDifferentiation.py - The functioning python file for Task 4

Outputs Directory - Resulting csv files

Actor2Tag-TF - Tag-weight vector for given actor under TF model

Initial input actor ID for generating this file is 1793283, as shown in header

Actor2Tag-TF-IDF - Tag-weight vector for given actor under TF-IDF model

Initial input actor ID for generating this file is 1793283, as shown in header

Genre2Tag-TF - Tag-weight vector for given genre under TF model

Initial input genre for generating this file is Thriller, as shown in header

Genre2Tag-TF-IDF - Tag-weight vector for given genre under TF-IDF model

Initial input genre for generating this file is Thriller, as shown in header

User2Tag-TF - Tag-weight vector for given user under TF model

Initial input user ID for generating this file is 146, as shown in header

User2Tag-TF-IDF - Tag-weight vector for given user under TF-IDF model

Initial input user ID for generating this file is 146, as shown in header

GenreDifferentiation-P-DIFF1 - Tag-weight vector for two given genres under P-DIFF-1 model

Initial input genres for generating this file are Thriller and Horror, as shown in header

GenreDifferentiation-P-DIFF2 - Tag-weight vector for two given genres under P-DIFF-2 model

Initial input genres for generating this file are Thriller and Horror, as shown in header

GenreDifferentiation-TF-IDF-DIFF-g1 - Tag-weight vector for first given genre under TF-IDF-DIFF model

Initial input genres for generating this file are Thriller and Horror, this file is for genre Thriller, as shown in header

GenreDifferentiation-TF-IDF-DIFF-g2 - Tag-weight vector for second given genre under TF-IDF-DIFF model

Initial input genres for generating this file are Thriller and Horror, this file is for genre Horror, as shown in header