

Requisiti

Python 2.7.x

Selenium 3.8.1

Phantomjs 2.1.3

Contenuto cartella

Script Python:

- *CCMixerGetSongList.py*

Serve per creare un file con una lista di url dei remix presenti sul sito, a partire dalla pagina "<http://ccmixter.org/view/media/remix/>"

esempio di utilizzo: `python CCMixerGetSongList.py listRemixUrl`

- *CCMixerSongMining.py*

A partire dalla lista creata in precedenza, url per url, vengono estratte e scritte su file le informazioni relative ai remix. Lo script produce tre file in output:

- un file in formato csv con le feature estratte;
- un file contenente la lista degli url dei profili degli autori dei remix esplorati (senza duplicati);
- un file contenente la lista degli url dei sample (canzoni a partire dalle quali i remix sono creati) (senza duplicati).

esempio di utilizzo: `python CCMixerSongMining.py listRemixUrl Songs.csv listAuthorsUrl listSamplesUrl`

- *CCMixerAuthorMining.py*

Serve per estrarre le informazioni relative agli autori, partendo dalla lista generata come output da *CCMixerSongMining.py*. Lo script produce in output un file in formato csv con le feature estratte relative agli autori.

esempio di utilizzo: `python CCMixerAuthorMining.py listAuthorsUrl Authors.csv`

Risultati:

- *CCMixerSongs.csv* (aggiornato al 5/2/2018)

File contenente le feature estratte relative alle canzoni. Le feature estratte comprendono:

- *Title*, titolo della canzone
- *Author*, autore della canzone
- *dateUpload*, data di upload della canzone sul sito
- *dateUpdate*, data dell'ultima modifica della canzone, se presente
- *isRemixed*, valore booleano che risulta vero se la canzone è un remix (creata utilizzando pezzi di altre canzoni), falsa nel caso contrario (sarà una canzone originale)
- *featuring*, lista di autori con cui si è collaborato, se presenti
- *samplesFrom*, lista di canzoni a partire dalle quali si è creato il remix, se presenti
- *recommends*, numero di "like" che ha ricevuto la canzone

- *CCMixerAuthors.csv* (aggiornato al 5/2/2018)

File contenente le feature estratte relative agli autori. Le feature estratte comprendono:

- *Username*, username dell'autore sul CCMixer
- *Uploads*, numero degli upload effettuati sul sito dall'autore
- *HasAvatar*, valore booleano, falso quando l'avatar dell'autore è quello di default, vero in caso contrario

- *SignUpDate*, data di iscrizione sul sito
- *remixDone*, numero di remix effettuati
- *remixReceived*, numero di volte in cui l'autore è stato remixato, ovvero sue canzoni sono state base per dei remix
- *PlaylistsWithAuthor*, numero di playlist contenenti canzoni realizzate dall'autore
- *forumMessage*, numero di messaggi scritti dall'autore sul forum
- *reviewLeft*, numero di review a canzoni scritte dall'autore
- *reviewReceived*, numero di review ricevute
- *CCMixerSongsNEW.csv* (aggiornato al 18/2/2018)
File contenente informazioni riguardo alle canzoni sul sito. Dopo aver modificato lo script *CCMixerSongMining.py*, è stato rilanciato per le necessità di includere 2 ulteriori nuove feature:
 - *samplesIn*, lista dei remix che contengono sample delle canzoni, se presenti
 - *reviews*, numero di review ricevute dalla canzone

Liste url:

- *ListaLinkCanzoni* (aggiornato al 5/2/2018)
Lista degli url relativi alle canzoni presenti su CCMixer. File necessario se si volesse rilanciare lo script *CCMixerSongMining.py*
- *ListaLinkAutori* (aggiornato al 5/2/2018)

Cartella API_files:

contiene file generati indirettamente utilizzando le API fornite da CCMixer

- *ListaLinkAutoriAPI* (aggiornato al 20/2/2018)
Lista degli url relativi ai profili degli autori su CCMixer, generata usando le API. File necessario se si volesse rilanciare lo script *CCMixerAuthorMining.py*
- *ListaLinkCanzoniAPI* (aggiornato al 20/2/2018)
Lista degli url relativi alle canzoni presenti su CCMixer, generata usando le API. File necessario se si volesse rilanciare lo script *CCMixerSongMining.py*
- *CCMixerSongsAPIlist.csv* (aggiornato al 21/2/2018)
43108 item
- *CCMixerAuthorsAPIlist.csv* (aggiornato al 20/2/2018)
4633 item

Guida

1. Modifica gli script in modo da inserire il percorso per l'eseguibile di phantoms
2. Avvia *CCMixerGetSongList.py* per stampare la lista degli url dei remix
3. Lancia *CCMixerSongMining.py* fornendogli in input la lista degli url dei remix
4. Rilancia *CCMixerSongMining.py* dandogli in input la lista degli url dei samples. Poi unisci il csv ottenuto con il csv ottenuto precedentemente, facendo attenzione ai possibili duplicati. Fai lo stesso con le liste degli url dei profili degli autori.
5. Lancia *CCMixerAuthorMining.py* per estrarre le feature relative agli autori