

MOVIE RECOMMENDATION BASED ON EMOTION

PYTHON PROJECT ASSIGNMENT

CS-1435(EL-II)

ABHIGYAN BARUAH | ROLL NO: 13 | REG.NO.201700172

*CONTENTS*

1. *INTRODUCTION…................................................................2*
2. *MODULES………………………………………………………………………….4*
3. *EXECUTABLE PYTHON CODE……………………………..………….6*
4. *FUNTIONS USED……………………………..……………………………….8*

*INTRODUCTION*

One of the underlying targets of movies is to evoke emotions in their viewers. IMDb offers all the movies for all genre. Therefore the movie titles can be scraped from the IMDb list to recommend to the user. IMDb does not have an API, for accessing information on movies and TV Series. Therefore we have to perform scraping. Scraping is used for accessing information from a website which is usually done with APIs.

# Installation

Install **BeautifulSoup** and **lxml**,  
Open terminal and write



The scraper is written in Python and uses lxml for parsing the webpages. BeautifulSoup is used for pulling data out of HTML and XML files.

# Emotion associated with Genre of Movies

There are 8 classes of emotion that would be effective to classify a text. These are: *‘Anger’, ‘Anticipation’, ‘Disgust’, ‘Fear’, ‘Joy’, ‘Sad’, ‘Surprise’, ‘Trust’*. Here these are taken as input and the corresponding movies would be displayed for the emotion.  
The correspondence of every emotion with genre of movies is listed below:

Sad – Drama  
Disgust – Musical  
Anger – Family  
Anticipation – Thriller  
Fear – Sport  
Enjoyment – Thriller  
Trust – Western  
Surprise – Film-Noir

Based on the input emotion, the corresponding genre would be selected and all the top 5 movies of that genre would be recommended to the user.

*MODULES*

# BeautifulSoup

Beautiful Soup is a Python library designed for quick turnaround projects like screen-scraping. Three features make it powerful:

1. Beautiful Soup provides a few simple methods and Pythonic idioms for navigating, searching, and modifying a parse tree: a toolkit for dissecting a document and extracting what you need. It doesn't take much code to write an application
2. Beautiful Soup automatically converts incoming documents to Unicode and outgoing documents to UTF-8. You don't have to think about encodings, unless the document doesn't specify an encoding and Beautiful Soup can't detect one. Then you just have to specify the original encoding.
3. Beautiful Soup sits on top of popular Python parsers like [lxml](http://lxml.de/) and [html5lib](http://code.google.com/p/html5lib/), allowing you to try out different parsing strategies or trade speed for flexibility.

Beautiful Soup parses anything you give it, and does the tree traversal stuff for us. You can tell it "Find all the links", or "Find all the links of class externalLink", or "Find all the links whose urls match "foo.com", or "Find the table heading that's got bold text, then give me that text."

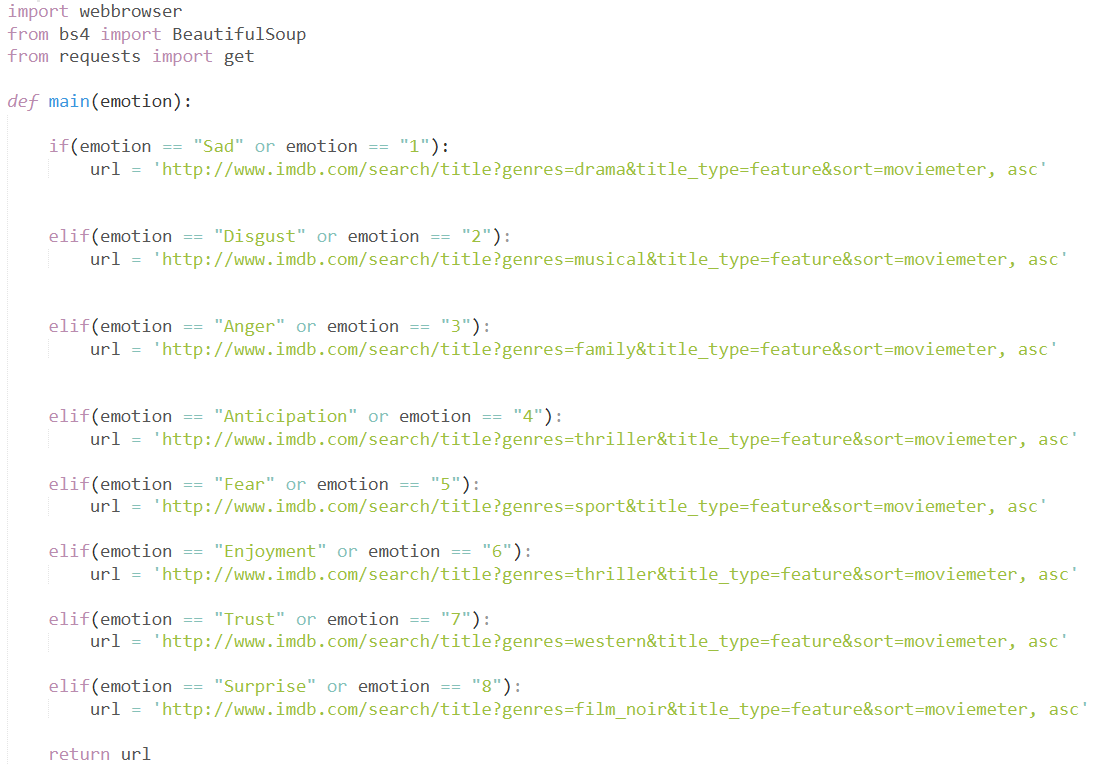
Valuable data that was once locked up in poorly-designed websites is now within our reach. Projects that would have taken hours take only minutes with Beautiful Soup.

# Lxml

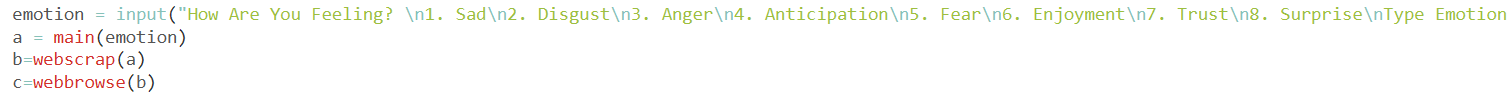
lxml is the most feature-rich and easy-to-use library for processing XML and HTML in the Python language.

The lxml XML toolkit is a Pythonic binding for the C libraries libxml2 and libxslt. It is unique in that it combines the speed and XML feature completeness of these libraries with the simplicity of a native Python API, mostly compatible but superior to the well-known ElementTree API. The latest release works with all CPython versions from 2.7 to 3.7.

*EXECUTABLE CODE*







*FUNCTIONS USED*

# Webscarp

This function uses BeautifulSoup and Lxml html parser to parse the selected item from the given input url from the main function into a database. In this case, the Name and the IMDb rating of the movie is parsed from the website. It also contains the display function, which uses the concept of Dictionary.

# Main

This function returns the url of the genre of movies of the selected emotion by the user.