Homework 2

Python Programming

November 12, 2014

Problem 1. IMDB: Exercise 20 in course slides

Please obtain the IMDB movies data from: *http://had.co.nz/data/movies/*
It is the file named *movies.tab*. The website also has a pdf document which describes the data in detail.
Write a program which follows the guidelines for Exercise 20 (see the course slides).

```
from imdb_ratings import condition, GroupMovieByElements
```

condition() checks if a record in the input file satisfies the following:

- The movie title present in a record has edit distance ≤ 3 with one of the user provided valid movie titles. The length of the provided movie titles will always be ≥ 5 characters.

- **Note**:

    - The provided movie titles reside in a text file *needed_movies.txt* with one in each line.
    - You can either write the program to compute the edit distance between two strings or you can use a suitable library which does that for you. Edit distance is also known as Levenshtein distance.

GroupMovieByElements uses **regular expressions** to extract the following from a record which satisfies the above condition:

- Votes

- Rating

- Year

- Title

There should be classes with methods for printing the extracted attributes of qualifying records in the following formats:

- CSV

- HTML table

Also, implement a test suite and try to achieve ˜100% coverage.