Fall 2016
BBM103: Introduction to Programming Laboratory 1
T.A. : Res. Assist. (Necva BOLUCU, Selma DILEK, Burcu YALCINER, Selim YILMAZ)

PROGRAMMING ASSIGNMENT 5- Movie Reviews

Due Date: 04.01.2017

Introduction

Movie reviews are a fairly commonly used tool used by consumers to understand if a movie is worth the price and time. There are different methods to create reviews about movies. One of them is rating the movies by different users. GroupLens Research has collected and made available rating datasets from the MovieLens web site (http://movielens.org). We used extra information about movies, Dennis Schwartz's reviews.

In this assignment, you will implement a python program that analyzes GroupLens' data and compares them with Dennis Schwartz's reviews. This program will create html files for movies which are both in Dennis Schwartz's reviews and in GroupLens' data and try to guess genres of movies based on the data which obtained from movies.

Fall 2016
BBM103: Introduction to Programming Laboratory 1
T.A. : Res. Assist. (Necva BOLUCU, Selma DILEK, Burcu YALCINER, Selim YILMAZ)

Stage 1: Create HTML Files for Movies

Step 1: Understand the GroupLens' data

In this assignment, we will give you different files to analyze. The most important stage is understanding the data.

- **u.item**

Information about the items (movies);

```
movie id | movie title | video release date | IMDb URL | unknown | Action | Adventure | Animation | Children's | Comedy | Crime | Documentary | Drama | Fantasy | Movie-Noir | Horror | Musical | Mystery | Romance | Sci-Fi | Thriller | War | Western |
```

The last 19 fields are the genres, a 1 indicates the movie is of that genre, a 0 indicates it is not; movies can be in several genres at once.

The movie ids are the ones used in the u.data.

Example: The content of the data

```
1|Toy Story (1995)|01-Jan-1995|http://us.imdb.com/M/title-exact?Toy%20Story%20(1995)|0|0|0|1|1|1|0|0|0|0|0|0|0|0|0|0|0|0|0

2|GoldenEye (1995)|01-Jan-1995|http://us.imdb.com/M/title-exact?GoldenEye%20(1995)|0|1|1|0|0|0|0|0|0|0|0|0|0|0|0|0|1|0|0

3|Four Rooms (1995)|01-Jan-1995|http://us.imdb.com/M/title-exact?Four%20Rooms%20(1995)|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|1|0|0

4|Get Shorty (1995)|01-Jan-1995|http://us.imdb.com/M/title-exact?Get%20Shorty%20(1995)|0|1|0|0|0|1|0|0|1|0|0|0|0|0|0|0|0|0|0

1176|Welcome To Sarajevo (1997)|01-Jan-1997 |http://us.imdb.com/M/title-exact?Welcome+To+Sarajevo+(1997)|0|0|0|0|0|0|0|0|1|0|0|0|0|0|0|0|0|1|0
```

Analyzing a line:

```
Movie id        : 1176

Movie title     : Welcome To Sarajevo (1997)

Release date    : 01-Jan-1997

IMDB Link       :http://us.imdb.com/M/title-exact?Welcome+To+Sarajevo+(1997)

Genre           : 0|0|0|0|0|0|0|0|1|0|0|0|0|0|0|0|0|1|0
```

- **u.genre**

This file contains a list of the genres.

genre | genre id

You will use this file to format genre field which are taken from u.item.

Example: Convert genre by taking genre names from u.genre file

```
Movie id        :1176

Movie title     : Welcome To Sarajevo (1997)

Genre           : Drama War
```

- **u.user**

This file contains demographic information about the users; (The user ids are the ones used in the u.data data set.)

user id | age | gender | occupation id | zip code

- **u.occupation**

This file consists of list of the occupations. (The occupation ids are the ones used in the u.user data

occupation id | occupation

Analyzing a line of u.user file by using u.occupation file:

```
User id   : 1
User Age  : 24
Gender    : M
Occupation: technician
Zip Code  : 85711
```

- **u.data**

The full data set, 100000 ratings by 943 users on 1682 items comprised of this file.

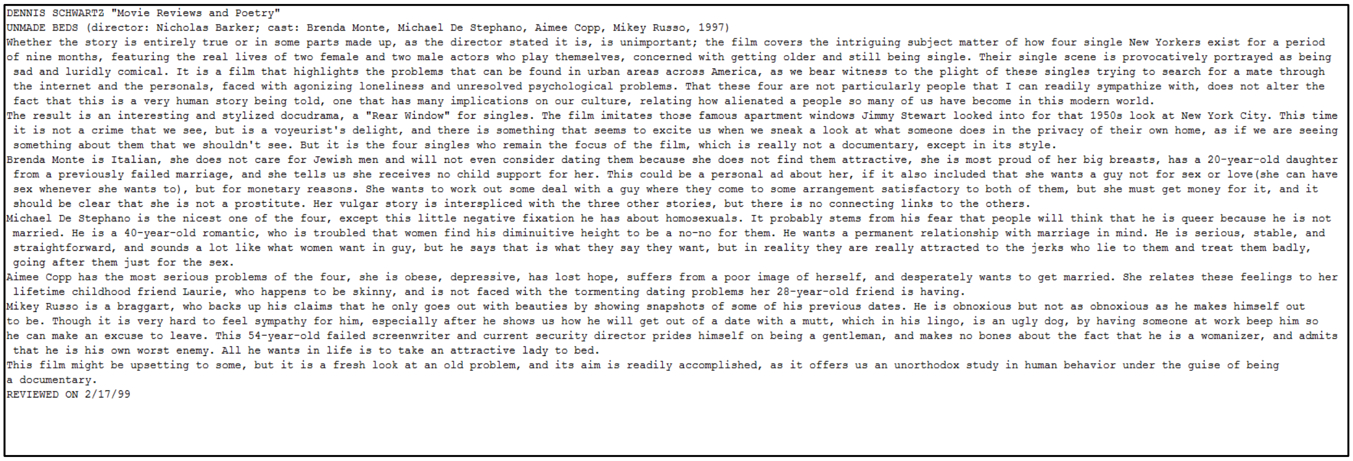
This is a tab separated list of

user id   movie id   rating timestamp.

Step 2: Understand the Dennis Schwartz's data

Dennis Schwartz' review data is taken from (*https://www.cs.cornell.edu/people/pabo/movie-review-data/ You can look here to get information about Dennis Schwartz)*. This data consists of different txt files.

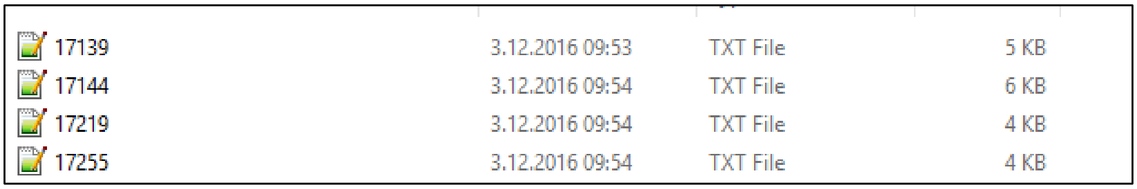
Example: Content of a file in this folder (16748.txt)



```
DENNIS SCHWARTZ "Movie Reviews and Poetry"
UNMADE BEDS (director: Nicholas Barker; cast: Brenda Monte, Michael De Stephano, Aimee Copp, Mikey Russo, 1997)
Whether the story is entirely true or in some parts made up, as the director stated it is, is unimportant; the film covers the intriguing subject matter of how four single New Yorkers exist for a period
of nine months, featuring the real lives of two female and two male actors who play themselves, concerned with getting older and still being single. Their single scene is provocatively portrayed as being
 sad and luridly comical. It is a film that highlights the problems that can be found in urban areas across America, as we bear witness to the plight of these singles trying to search for a mate through
 the internet and the personals, faced with agonizing loneliness and unresolved psychological problems. That these four are not particularly people that I can readily sympathize with, does not alter the
 fact that this is a very human story being told, one that has many implications on our culture, relating how alienated a people so many of us have become in this modern world.
The result is an interesting and stylized docudrama, a "Rear Window" for singles. The film imitates those famous apartment windows Jimmy Stewart looked into for that 1950s look at New York City. This time
 it is not a crime that we see, but is a voyeurist's delight, and there is something that seems to excite us when we sneak a look at what someone does in the privacy of their own home, as if we are seeing
 something about them that we shouldn't see. But it is the four singles who remain the focus of the film, which is really not a documentary, except in its style.
Brenda Monte is Italian, she does not care for Jewish men and will not even consider dating them because she does not find them attractive, she is most proud of her big breasts, has a 20-year-old daughter
 from a previously failed marriage, and she tells us she receives no child support for her. This could be a personal ad about her, if it also included that she wants a guy not for sex or love(she can have
 sex whenever she wants to), but for monetary reasons. She wants to work out some deal with a guy where they come to some arrangement satisfactory to both of them, but she must get money for it, and it
 should be clear that she is not a prostitute. Her vulgar story is interspliced with the three other stories, but there is no connecting links to the others.
Michael De Stephano is the nicest one of the four, except this little negative fixation he has about homosexuals. It probably stems from his fear that people will think that he is queer because he is not
 married. He is a 40-year-old romantic, who is troubled that women find his diminuitive height to be a no-no for them. He wants a permanent relationship with marriage in mind. He is serious, stable, and
 straightforward, and sounds a lot like what women want in guy, but he says that is what they say they want, but in reality they are really attracted to the jerks who lie to them and treat them badly,
 going after them just for the sex.
Aimee Copp has the most serious problems of the four, she is obese, depressive, has lost hope, suffers from a poor image of herself, and desperately wants to get married. She relates these feelings to her
 lifetime childhood friend Laurie, who happens to be skinny, and is not faced with the tormenting dating problems her 28-year-old friend is having.
Mikey Russo is a braggart, who backs up his claims that he only goes out with beauties by showing snapshots of some of his previous dates. He is obnoxious but not as obnoxious as he makes himself out
to be. Though it is very hard to feel sympathy for him, especially after he shows us how he will get out of a date with a mutt, which in his lingo, is an ugly dog, by having someone at work beep him so
he can make an excuse to leave. This 54-year-old failed screenwriter and current security director prides himself on being a gentleman, and makes no bones about the fact that he is a womanizer, and admits
 that he is his own worst enemy. All he wants in life is to take an attractive lady to bed.
This film might be upsetting to some, but it is a fresh look at an old problem, and its aim is readily accomplished, as it offers us an unorthodox study in human behavior under the guise of being
a documentary.
REVIEWED ON 2/17/99
```

Each of files is about only a movie review. These files are supposed to be in a folder which is named **film.** You are expected to read these files one by one from the folder. These files count can be changed, so you must read them in a loop.

Example: film folder



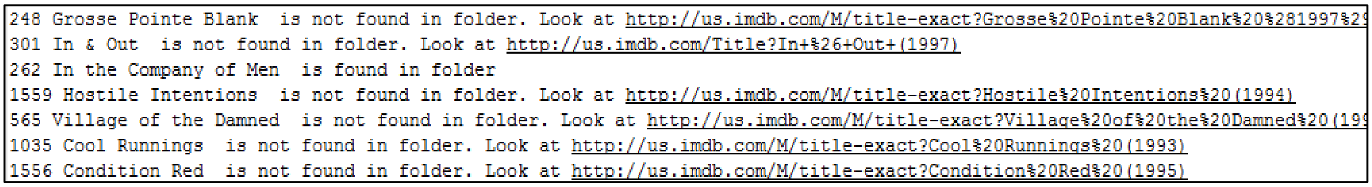
```
17139    3.12.2016 09:53    TXT File    5 KB
17144    3.12.2016 09:54    TXT File    6 KB
17219    3.12.2016 09:54    TXT File    4 KB
17255    3.12.2016 09:54    TXT File    4 KB
```

Step 3: Combine the GroupLens' data and Dennis Schwartz's data

In order to create html files for movies, you must combine the datasets. You are expected to create html files for the movies which are in **film** folder. **In this step, we expected to use list comprehensions.**

Firstly, you compare the both dataset (movies in film folder and u.item) and select the movies which are in both datasets. You will create **review.txt** file to write messages for movies which are in u.item but not in film folder and movies which are found in folder. Use user-defined exception to take messages.

Example: review.txt



```
248 Grosse Pointe Blank  is not found in folder. Look at http://us.imdb.com/M/title-exact?Grosse%20Pointe%20Blank%20%281997%2
301 In & Out  is not found in folder. Look at http://us.imdb.com/Title?In+%26+Out+(1997)
262 In the Company of Men  is found in folder
1559 Hostile Intentions  is not found in folder. Look at http://us.imdb.com/M/title-exact?Hostile%20Intentions%20(1994)
565 Village of the Damned  is not found in folder. Look at http://us.imdb.com/M/title-exact?Village%20of%20the%20Damned%20(19
1035 Cool Runnings  is not found in folder. Look at http://us.imdb.com/M/title-exact?Cool%20Runnings%20(1993)
1556 Condition Red  is not found in folder. Look at http://us.imdb.com/M/title-exact?Condition%20Red%20(1995)
```

After selecting movies, you will find user ids who rate them from u.data and get detail information about these users from u.user.

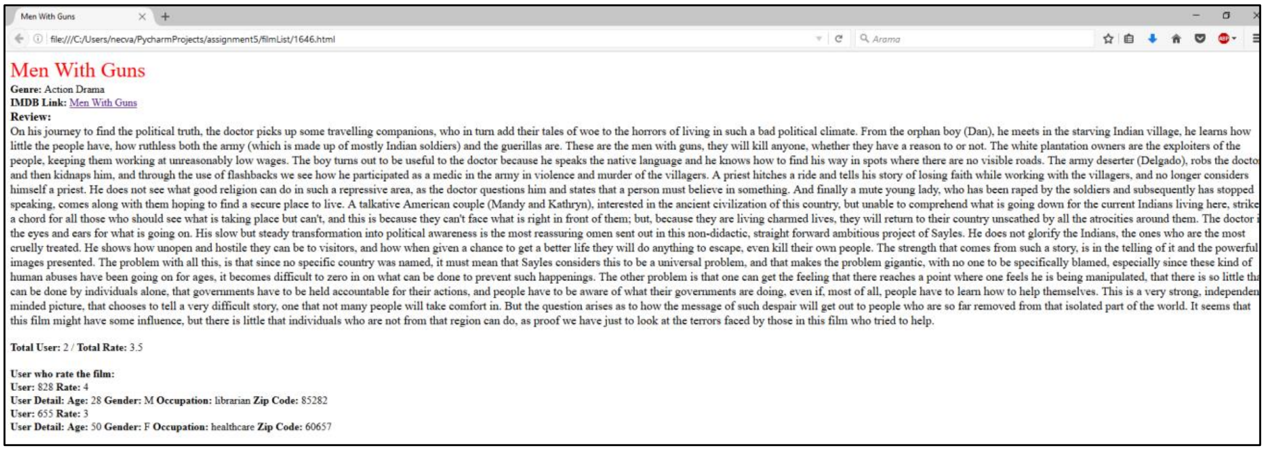
Step 4: Write review to html file

When you extract information from given data for movies, you are going to use this data to create html files which are located in **filmLis**t folder. In html file, the necessary fields are shown;

```
<html>

<title> NAME OF THE FILM</title>

<body>

 Times New Roman size="6" color="red" bold NAME OF THE FILM

Genre

IMDB Link

Times New Roman size="4" color="black” boldReview (taken from Dennis
Schwartz ‘s data)

Total User/Total Rate

User who rate the film:

User id     User rate

User Detail: Age - Gender - Occupation - Zip Code

</body>

</html>
```

! The file name is must be the film id which are given u.item.

Example: One of the output file


Men With Guns
file:///C:/Users/necva/PycharmProjects/assignment5/filmList/1646.html
Arama

Men With Guns

Genre: Action Drama
IMDB Link: Men With Guns
Review:
On his journey to find the political truth, the doctor picks up some travelling companions, who in turn add their tales of woe to the horrors of living in such a bad political climate. From the orphan boy (Dan), he meets in the starving Indian village, he learns how
little the people have, how ruthless both the army (which is made up of mostly Indian soldiers) and the guerillas are. These are the men with guns, they will kill anyone, whether they have a reason to or not. The white plantation owners are the exploiters of the
people, keeping them working at unreasonably low wages. The boy turns out to be useful to the doctor because he speaks the native language and he knows how to find his way in spots where there are no visible roads. The army deserter (Delgado), robs the docto
and then kidnaps him, and through the use of flashbacks we see how he participated as a medic in the army in violence and murder of the villagers. A priest hitches a ride and tells his story of losing faith while working with the villagers, and no longer considers
himself a priest. He does not see what good religion can do in such a repressive area, as the doctor questions him and states that a person must believe in something. And finally a mute young lady, who has been raped by the soldiers and subsequently has stopped
speaking, comes along with them hoping to find a secure place to live. A talkative American couple (Mandy and Kathryn), interested in the ancient civilization of this country, but unable to comprehend what is going down for the current Indians living here, strike
a chord for all those who should see what is taking place but can't, and this is because they can't face what is right in front of them; but, because they are living charmed lives, they will return to their country unscathed by all the atrocities around them. The doctor i
the eyes and ears for what is going on. His slow but steady transformation into political awareness is the most reassuring omen sent out in this non-didactic, straight forward ambitious project of Sayles. He does not glorify the Indians, the ones who are the most
cruelly treated. He shows how unopen and hostile they can be to visitors, and how when given a chance to get a better life they will do anything to escape, even kill their own people. The strength that comes from such a story, is in the telling of it and the powerful
images presented. The problem with all this, is that since no specific country was named, it must mean that Sayles considers this to be a universal problem, and that makes the problem gigantic, with no one to be specifically blamed, especially since these kind of
human abuses have been going on for ages, it becomes difficult to zero in on what can be done to prevent such happenings. The other problem is that one can get the feeling that there reaches a point where one feels he is being manipulated, that there is so little tha
can be done by individuals alone, that governments have to be held accountable for their actions, and people have to be aware of what their governments are doing, even if, most of all, people have to learn how to help themselves. This is a very strong, independen
minded picture, that chooses to tell a very difficult story, one that not many people will take comfort in. But the question arises as to how the message of such despair will get out to people who are so far removed from that isolated part of the world. It seems that
this film might have some influence, but there is little that individuals who are not from that region can do, as proof we have just to look at the terrors faced by those in this film who tried to help.

Total User: 2 / **Total Rate:** 3.5

User who rate the film:
User: 828 **Rate:** 4
User Detail: Age: 28 **Gender:** M **Occupation:** librarian **Zip Code:** 85282
User: 655 **Rate:** 3
User Detail: Age: 50 **Gender:** F **Occupation:** healthcare **Zip Code:** 60657

Stage 2: Guess Genre of Movie Based on Film Given in Film Folder

In this stage, it is expected to guess movies genre or genres based on the movies given in film folder.

Step 1: Getting Unique Words Without Stop Words

Firstly, you should extract unique words for all genres by taking movies genres and their review data, then make stop word elimination by stop word list which are taken **stopwords.txt**.

Step 2: Guess Genres of Movies

In this step, you will read the movie reviews from the **filmGuess** folder and implement step1 for getting unique words.

After getting words, if the intersection of words of movie and genres words is higher or equal **20,** we remark the genre for the movie.

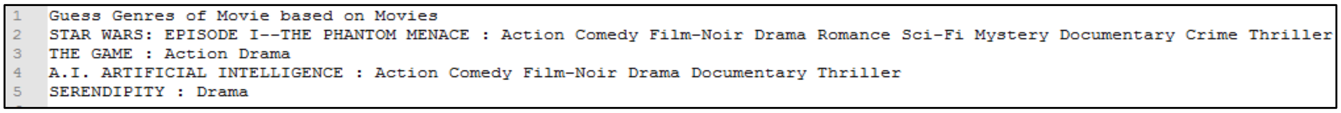
Step 3: Write to File

After getting genre or genres for the all movie in the **filmGuess** folder, you will write the film names and genres to **filmGenre.txt**

Example: Movies in The filmGuess Folder

```
18485.txt      STAR WARS: EPISODE I--THE PHANTOM MENACE

21168.txt      THE GAME

18687.txt      A.I. ARTIFICIAL INTELLIGENCE

29852.txt      SERENDIPITY
```

Output: filmGenre.txt for the given movies in filmGuess folder



```
Guess Genres of Movie based on Movies
STAR WARS: EPISODE I--THE PHANTOM MENACE : Action Comedy Film-Noir Drama Romance Sci-Fi Mystery Documentary Crime Thriller
THE GAME : Action Drama
A.I. ARTIFICIAL INTELLIGENCE : Action Comedy Film-Noir Drama Documentary Thriller
SERENDIPITY : Drama
```

Notes specified to this assignment

- Feel free to employ any built-in function.
- Use list comprehension and user-defined exception to take message in your project.
- Be careful to open and create files in try except block to avoid forum IOError.
- Be sure your submitted work exactly matches the hierarchy detailed below, as the submission with 0 score will not be considered for evaluation.
- Should you have a question, do not hesitate to as but consider office hours for BBM103 of TA in charge (Necva BÖLÜCÜ).
- You will use **static file and folder names**, but be careful to name files correctly.
- **Due date for this assignment is 04.01.2017**

Fall 2016
BBM103: Introduction to Programming Laboratory 1
T.A. : Res. Assist. (Necva BOLUCU, Selma DILEK, Burcu YALCINER, Selim YILMAZ)

Notes

- Do not miss the deadline.
- Compile your code on dev.cs.hacettepe.edu.tr before submitting your work to make sure it compiles without any problems on our server.
- Save all your work until the assignment is graded.
- The assignment must be original, individual work. Duplicate or very similar assignments are both going to be considered as cheating.
- You can ask your questions via Piazza
- (https://piazza.com/hacettepe.edu.tr/fall2016/bbm101) and you are supposed to be aware of everything discussed in Piazza. You cannot share algorithms or source code. All work must be individual! Assignments will be checked for similarity, and there will be serious consequences if plagiarism is detected.
- The submissions whose upload score is 0 will not be considered for evaluation.
- Do not include any other text files in your submission. Input files will have the same names but different content than those you worked on. Output files should be created when your program is executed.
- You will submit your work from *https://submit.cs.hacettepe.edu.tr/index.php* with the file hierarchy as below:

This file hierarchy must be zipped before submitted (Not .rar, only .zip files are supported by the system)

→ <student id>
  → assignment5.py

Who is Dennis Schwartz?

Dennis Schwartz is editor of the Vermont based movie magazine "Ozus's World Movie Reviews." He has been a prolific online movie reviewer since 1998, also contributing to various publications all over the globe and maintaining an active website--where it's not uncommon for him to review as many as 365 movies a year.