Assignment: Crawl and parse information from IMDB and store in MySQL database

Pre-requisites: (Pip+lxml+requests) installed in the system, Python Programming Language, MySql database.
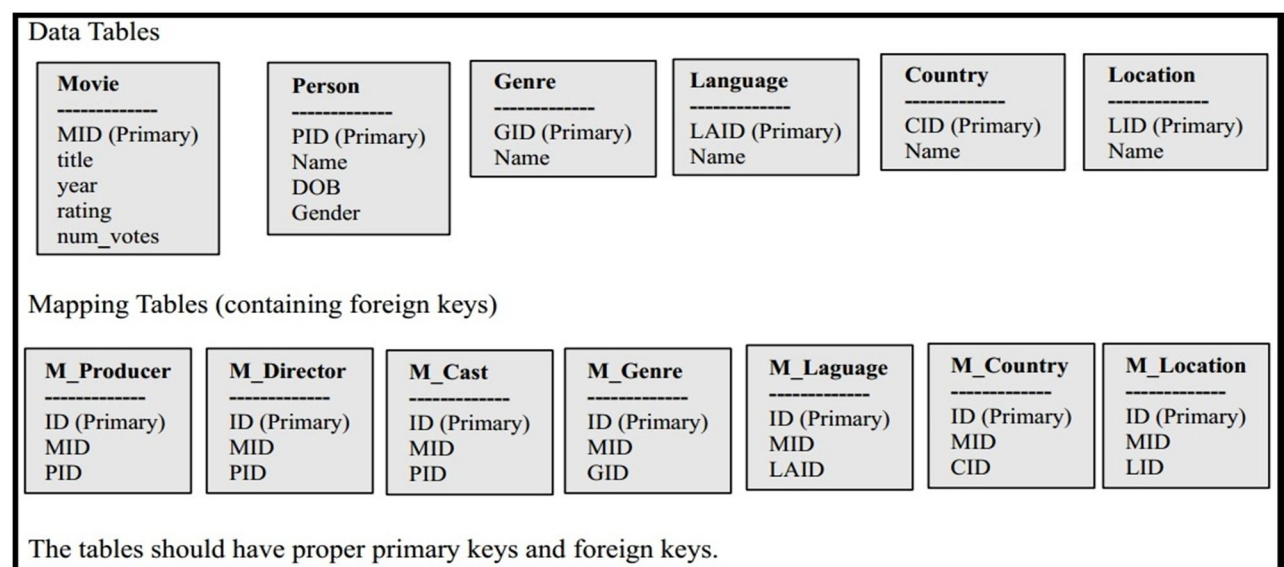
Description:

Given a link to a page of IMDB website containing sorted movie list you have to crawl the movie details, cast details and also the next pages containing the rest of the ranked movies. You have to crawl the first 5 pages of the given link.

The crawled information will be inserted in the MySQL Database tables created beforehand by you. The link to crawl and prefix (preferably your roll number) for table names should be taken from a text input file.

Before starting the assignment go through the "setup.txt" file, make your system ready for crawling and run the "sampleCode.py".

IMDB database schema:

```
Data Tables

 Movie            Person          Genre           Language        Country         Location
 -----------      -----------     -----------     -----------     -----------     -----------
 MID (Primary)    PID (Primary)   GID (Primary)   LAID (Primary)  CID (Primary)   LID (Primary)
 title            Name            Name            Name            Name            Name
 year             DOB
 rating           Gender
 num_votes

Mapping Tables (containing foreign keys)

 M_Producer    M_Director    M_Cast        M_Genre       M_Laguage     M_Country     M_Location
 -----------   -----------   -----------   -----------   -----------   -----------   -----------
 ID (Primary)  ID (Primary)  ID (Primary)  ID (Primary)  ID (Primary)  ID (Primary)  ID (Primary)
 MID           MID           MID           MID           MID           MID           MID
 PID           PID           PID           GID           LAID          CID           LID

The tables should have proper primary keys and foreign keys.
```

Link:

Highest Rated Hindi-Language Feature Films With At Least 50 Votes

http://www.imdb.com/search/title?languages=hi|1&title_type=feature&num_votes=50,&sort=user_rating,desc