Project Status Report - Actor Recommendation for Casting Directors

Team: Vijayakumar Gedigeri, Yiming Liu **11/02/17, INF 558**

Abstract:
While making a movie, casting directors have to spend lot of time and effort to find the best match for character description of an actor to be casted for the movie. In this project, we aim to build a knowledge graph of actors combining the popularity of actors on social media, their previous movie records and physical characteristics. The resulting system will rank the actors that match the character description and will serve as a recommendation engine for casting directors for the actor consideration.

Project Scope:
Data Sources: IMDB, Twitter, Facebook, Rotten Tomatoes, Instagram
Movie Language: English
Knowledge Graph Size: 35000 actors that have appeared in movies from 1970 to 2017

Project Status: Currently data from IMDB, Twitter and Facebook is collected for actors and stored in cdr format after data cleaning. Web crawling is done using Scrapy and data extraction using XPath Selector from Scrapy. Below is the detailed steps

- From IMDB, crawled around 270,000 actor URLs from around 20,000 English movies and TV shows having rating of more than 5 and minimum 500 votes released between 1970 and 2017.
- Removed duplicate actor URLs from a list of 270,000 to get unique 79,000 actors.
- Removed actors that do not have birth information, or are already dead to get around 35000 distinct actors.
- Crawled actors' pages from IMDB and extracted useful information like URL's of movies, Twitter and Facebook usernames, and personal information.
- Crawled Twitter to get user profile information for actors having twitter accounts via Twitter Search API.
- Crawled Twitter and extracted tweets for each actor with actor names as the keywords.
- Crawled Facebook to get user profile information for actors having Facebook accounts.

Next Steps:

- Crawl movie pages from IMDB and Rotten Tomatoes, and actors information from Instagram.
- Implement natural language processing methods like sentiment analysis and topic modeling to analyse the tweets and movie reviews
- Build a scoring system for actors' information from each source.
- Construct the database and address entity resolution while combining the sources.
- Build the user interface which offers searching the knowledge graph for actors.