**Big Data Project Proposal**

**Abstract**

In this project, our team will implement a data portal which is for movie and actor searches. The input is a query which is a combination of terms related to movies or actors and the output is the detailed information of the movie or actor or a sorted list of movies or actors. The original data is crawled down from web site, data mining technologies such as feature extraction, entity identification, entity disambiguation, etc will be used to extract information from the web site and to build the data base.

**Data portal usage**

1. Basic query:
2. Given an actor/actress, the portal will show all films in which he/she has a role.
3. Given an actor/actress combing with the role position (i.e. protagonist, male supporting role), the portal will show all films reach the conditions.
4. Given the language or country, the portal will show the films in that language originally.
5. Given a song name, the portal will show the films that use the song as main music or background music.
6. Given a role name, the portal will show the films in which the role name is included.
7. Given a genre tag such as "action" or "comedy", the portal will show the related films
8. Given an award, the portal will show the movie/actor.
9. Given a director, the portal will show all the films that he/she has directed
10. Advanced query:
11. Given a blurred description such as “most popular movie in 2015”, a list of sorted movies will be returned.
12. Given a theme keyword such as “movie of second world war”, a list of movies will be returned.

However, the final implementation of the functions might be tuned accordingly.

**Attributes for movie**:

name, Genre tag(action, comedy, fiction), major actors, theme song, language, country, awards, release year, box office, role, director, popularity, keywords

**Attributes for actor**:

name, country, Female/male, role, movies, awards, keywords

**Data source**:

1) https://www.wikipedia.org/

api: http://www.programmableweb.com/api/wikipedia

2) http://www.imdb.com/

api: http://www.omdbapi.com/

3) https://www.themoviedb.org/

api: <https://www.themoviedb.org/documentation/api>

**Data base**:

MySQL

**Programming language**:

Ruby, Python, Java

**Data structure**:

XML, JSON

**Team Member**:

QI, Xiaoxu 20298859 xqiab@connect.ust.hk

JIN, Yue 20295728 yjinah@connect.ust.hk

CHEN, Guanhao 20292221 gchenah@connect.ust.hk