**Team Name: Funtastic Four (Hao Han, Jie Ren, Kai Wang and Gabriel Yang)**

**Background**

Imagine yourself are a screenwriter trying to craft the next blockbuster script. To make your work stand out among thousands of movie scripts and get producers’ attentions, using the right words becomes crucial. Our system aims to provide valuable information on word choices from hundreds of well-known movies. By ranking top/ bottom words used in their respective movies, we provide recommendations on the word to choose and to avoid given your specific genre and other inputs.

**Objectives**

This system consists of two parts: a search engine and word frequency charts based on movie genres. Specially, our goals will be the following:

1. Build a searchable database of movies with different genres from IMDb. Specifically we will have two groups of movies for each genre: great ones and mediocre ones.

2. Identify some words that differentiate great action movies from mediocre ones. In this way, screenwriters may choose to build their plots on these words. For example, if I want to write a great action movie about FBI, I can search FBI in the database finding out what great vs. so-so movies are produced about FBI and what words differentiate these two types of movies.

3. Build an attractive and interactive interface for the system

**Procedure**

The following two links contain all 2672 action movies with at least 1000 votes ranked by average user rating.

<http://www.imdb.com/search/title?at=0&genres=action&num_votes=1000,&sort=user_rating,desc&title_type=feature>

Here is how the movie project proceeds

1. Follow the link above to get to the list of top movies and mediocre movies (maybe 200 each)

2. Use BeautifulSoup and urlib2 to get the synopsis of the movies

3. Get word frequencies within these two groups of synopsis. Find which words differentiate great movies from mediocre movies

4. Build a database where people can use keywords to search and find great vs. so-so movies (as well as their word frequencies)

5. Build an interactive search interface with word cloud for showing differentiating words