Project Proposal

[Free Topic] Search Engine for Movie Adaptations of Books

1. Team

Daiyun Xu (net id: daiyunx2) - working individually

2. Description

My topic is a search engine for book-to-movie or movie-to-book adaptations. This is interesting in that it is highly possible that people who like a movie will also want to read the book the movie is based on, and vice versa.

So, the basic idea is building a recommendation system on top of a classic search engine to recommend books to a movie query or movies to a book query. For datasets, I will crawl data mainly from IMDB and Good Reads. Then I can borrow some ideas from ExpertSearch System to build a classic search engine, which allows users to choose either movie or book as the query category. Lastly, BM25 or other content-based similarity algorithms can be used to find the top relevant documents from the other category for recommendation. My expected outcome is when querying a novel-adapted movie, at the very least the original novel should show up among the top few recommendations. Other valid recommendations would be books from the same series or books written by the same writer, etc. To evaluate it's working, I can test it against a dataset with known book-to-movie relationships.

3. Languages

Python, Javascript, HTML/CSS

4. Workload estimation

1. crawl data from IMDB and Good Reads (5~10h)
2. build the movie/book search engine based on the idea of ExpertSearch System (5~10h)
3. recommendation system implementation / technology review on the same topic (5~10h)
4. performance review and documentation (5~10h)

Total workload should be at least 20 hours.