Movie search

Users initiate informational queries regarding movies where they can obtain few results about a movie. We need to assure not to make the results too similar to each other to provide diversity among the available results. In movies query there is no problem of one phrase with many meanings unlike in the web where terms like “apple” could refer to fruit or company but here we know the context and it is about movies.
Users expect similar results for phrases like "movies 2007", "2007 movies" and "movies released in 2007".Query expansion using global analysis using some form of thesaurus would help. For each term in a query, the query can be automatically expanded with synonyms and related words from the dictionary. Use of a dictionary can be combined with ideas of term weighting.
Dictionary could use controlled vocabulary or an automatically derived thesaurus generated by analyzing documents and building a co occurrence matrix and then applying dimensionality reduction. At present we are not looking into personalized results which make use of query log mining as that would require huge user query data to be collected in advance.

Once we have applied query reformulation, we query the SQL database containing movie information accordingly. Users can check for top 3 movies similar to a movie which are determined based on data like genres (a movie can be upto in maximum 3 genres), IMDB rating of the movie, release date etc..

In our github repository, the codes similar.py and recom.py do the queries. We are working now on query reformulation.

References
https://nlp.stanford.edu/IR-book/html/htmledition/global-methods-for-query-reformulation-1.html:
https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36337.pdf