Web search engine system is one of the most popular web applications. Typically, users can type any kind of queries and get ranked web pages. In recent years, both technique-oriented and user-oriented research has been done to improve users’ search experience. A particular interesting topic is that a query sometimes implies several different facets of user needs. For example, the query “Harry Potter” may indicate users’ information needs of getting movie ticket, watching the video, or reading the books. If only movie ticket pages are provided (considering there are hundreds of websites provide this information), then other kind of user needs may not be fulfilled.

One possible way of overcoming this problem is by clustering the ranked documents in different facets and provides an interface for representing different facets of user needs. The method of clustering and re-ranking documents based on certain criterions is not new, clustering algorithm such as K-means are adopted. One problem of those algorithms is that they are lacking of interpretation ability, which means they only cluster documents, but take no further steps of considering interpreting the semantic meaning of this cluster. A topic modeling method proposed recently can be used here (Blei & Jordan, 2003), because this method is intuitively in the latent topic level.

However, to achieve the goal of investigating multi-facets information needs representation, we need to at first build a prototype search system. A basic prototype of search system consists of three main components: the spider, the indexer and the searcher. Spider collects web pages from varieties of external resources; then, Indexer parses the collected documents and do further analysis. In the final step, a searcher ranks all documents based on certain models given users’ queries. In this project, we are going to develop a search engine system that can fulfill users’ multi-facet information needs.

Basically, we are planning to do three things:

• Construct a prototype web search system, which consist of Spider, Indexer and Searcher;

• Re-ranking documents based on the facets of user information needs, which we assumed that they are identical to topic analysis in documents;

• Developing a web user-interface that can represent multi-facet of results