Abstract

Information Retrieval plays a vital role in our daily activities and its most prominent role marked in search engines. Retrieval of the relevant natural language text document is of more challenge. Typically, search engines are low precision in response to a query, retrieving lots of useless web pages, and missing some other important ones. In this paper, we present linguistic
phenomena of NLP using shallow parsing and Chunking to extract the Noun Phrases. These noun phrases are used as key phrases to rank the documents (typically a list of titles and snippets returned by a certain Web search engine). Organizing Web search results in to clusters facilitates user's quick browsing through search results. Traditional clustering techniques are inadequate since they don't generate clusters with highly readable names. Here, we also proposed an approach for web search results clustering based on a phrase based clustering algorithm Known as Optimized Snippet Flat Clustering (OSFC). It is an alternative to a single ordered result of search engines. This approach presents a list of clusters to the user. Experimental results verify our method's feasibility and effectiveness.

Reference

- Yahoo! Search BOSS (Build your Own Search Service) http://developer.yahoo.com/search/boss/
- A Vector Space Model For Automatic Indexing, G. Salton, A. Wong and C. S. Yang, Cornell University.
- SharpNLP - open source natural language processing tools,
http://www.codeplex.com/sharpnlp

Index Terms

Computer Science
Information Retrieval

Key words

Noun Phrases
Document Clustering

Information Retrieval

Natural Language Processing

Web Mining