Semantic Clustering Approach Based Multi-agent System for Information Retrieval on Web

Bassma S. Alsulami, Maysoon F. Abulkhair, Fathy A. Essa

Faculty of Computing and Information Technology
King AbdulAziz University
Jeddah, KSA

Summary
Document clustering is an important technology which helps users to organize the large amount of online information, especially after the rapid growth of the Web. This paper focuses on semantic document clustering method and its application in search engine. We proposed a multi-agent based information retrieval system to enhance the search process. The agents retrieve the results of Web search engine and organize the results by clustering them into different categories for a given query. We utilized WordNet ontology and several approaches to cluster results in appropriate category according to WordNet synsets. The experiment shows that semantic clustering work better than original clustering.

Key words:
Document Clustering, WordNet, Semantic Clustering, information retrieval, Multi-Agent System

1. Introduction

Information Retrieval plays an important role in our daily life and its largest role is observed in search engines. Most users rely on Web search engines to look for specific information from the Web. These search engines often return a long list of search results that would be ranked by their relevance to the given query. Web users have to go through the long list and inspect the titles, and snippets sequentially to recognize the required results. Filtering the search engines' results consumes the users' effort and time especially when multiple sub-topics of the given query are mixed together [1]. This problem raise because the current search engines make little effort to understand users' queries and they use traditional techniques relay on matching terms and phrases; the search engine return a page just when the page includes the target keywords. Also, Web users sometime sent very short queries (only one word) and often it has multiple meanings that make a task of finding relevant information from the Web using just few words is a tremendous challenge.

Manuscript received January 5, 2012
Manuscript revised January 20, 2012

Search results' clustering attempts to solve this problem by automatic organizing a list of search result returned by a search engine into a set of meaningful thematic categories. The available clustering search engines tend to use string matching, rather than true linguistic analysis, to identify keywords and phrases which documents share in common. Then they generate clusters based on these shared keyword and phrases [2]. In this work, the words are semantically analyzed using WordNet to produce precise analysis of how the documents might relate to each other.
This paper describes how to overcome some of the major limitations in the current search engines. We proposed a multi-agent based information retrieval system to enhance the search process. We used different types of agents each of them has its own responsibility. We organize the results of Web search engine by clustering them into different categories for a given query. We utilized WordNet ontology and several approaches to cluster results in appropriate category according to WordNet synsets.
The paper is organized as follows. Some related works are introduced in Section 2. The basic concepts are defined in Section 3. In Section 4, we describe the proposed system Architecture. The implantation of the proposed system is presented in Section 5. Section 6, we describe the results from the applied clustering technique and its evaluations, which associated with examples. Finally we conclude the paper and give some future works in Section 7.

2. Related work

One of the very well-known approaches for query ambiguity is search result clustering. The basic idea of this approach is to cluster a given query with the list of snippets returned from search engines based on some measures of similarity. Algorithms for clustering Web search result have been reported in many papers such as [3, 4, 5]. One of the pioneer works in this respect is the Scatter/Gather project [3], but this project has some limitations because it used a traditional heuristic clustering algorithms. Y. Wang et al. [4] have proposed a clustering Web search results