

Chapter 84

Ranking Documents Based on the Semantic Relations Using Analytical Hierarchy Process: Query Expansion and Ranking Process

Ali I. El-Dsouky

Kafrelsheikh University, Egypt

Hesham A. Ali

Kafrelsheikh University, Egypt

Rabab Samy Rashed

Kafrelsheikh University, Egypt

ABSTRACT

With the rapid growth of the World Wide Web comes the need for a fast and accurate way to reach the information required. Search engines play an important role in retrieving the required information for users. Ranking algorithms are an important step in search engines so that the user could retrieve the pages most relevant to his query. In this work, the authors present a method for utilizing genealogical information from ontology to find the suitable hierarchical concepts for query extension, and ranking web pages based on semantic relations of the hierarchical concepts related to query terms, taking into consideration the hierarchical relations of domain searched (sibling, synonyms and hyponyms) by different weighting based on AHP method. So, it provides an accurate solution for ranking documents when compared to the three common methods.

1. INTRODUCTION

Web based information retrieval systems; especially search engines are the basic tools to assist users to find information on the World Wide Web. Despite the vital role in reaching information, many of the returned results are irrelevant to the user's needs as they are ranked based on the string matching of the

DOI: 10.4018/978-1-5225-5191-1.ch084

user's query. This has created a semantic gap between the meanings of the keywords in the retrieved documents and the meanings of the terms used in users' queries.

Search is the most popular applications on the Web. The bulk of traditional retrieval systems usually make use of metadata keywords matching with the query. However, these systems do not take into account the semantic relationships between query terms and other concepts that might be significant to users. Thus, the addition of explicit semantics can improve the search process. Semantic search is an application of the Semantic Web to search. It tries to improve traditional search results (based on Information Retrieval technology) using data from the Semantic Web (Preethi & Devi, 2012). This approach offers an enhancement to traditional search as it allows retrieval to incorporate the underlying terms semantics (Lee, Hendler & Lassila, 2001). It improves the traditional search that focuses on word frequency by trying to understand hidden meanings in the retrieved documents and users' queries (Huan, Duan, Tang et al., 2011; Sharma, Kandpa, Bhakuni et al., 2013). The problem of poor retrieval information system exists when users cannot clearly express their information needs or poor ranking methods to evaluate pages if they are related to query or not.

In order to overcome the irrelevant documents that result from search process, there are many solutions such as: using query expansion (QE), taking into account the semantic meaning; or by improving the ranking of documents, taking into account not only the occurrence of query terms, but also the semantic relation between the user search and the document context.

QE is considered a viable solution, expanding process by expanding query keywords with related terms. With an expanded query, the retrieved documents are not only based on the query terms, but also on the related terms to that query which can improve the search process. This is suitably broadened and more accurate results may be obtained by retrieving more relevant documents. Web search ranking algorithms play an important role in ranking web pages so that the user could get good results more relevant to the user's query.

This paper presents two methods to solve these problems. The first is an expansion query method taking into consideration the relations between expanded query terms in the ranking process of documents, by organizing all terms of an expanded query as a tree model of multi-levels, regarding their hierarchical relationships defined in a specific ontology. The second method is a ranking process for documents based on the semantic relation between document contents and the query terms.

2. RELATED WORK

Search engines accuracy is improved based on how they will search for the meaning of query terms, and how they will present the results to users by evaluating the documents containing the query terms. There are many solutions for improving the search engine: by expanding query taking into account the semantic meaning related to user's query terms; or by improving the evaluation of documents not only by the occurrence of terms, but also by how it semantically relates to the topic search.

Query expansion (QE) is a technique used to aid users to express their requirements. There are many works in QE techniques, such as the mechanisms of relevance feedback (Lin, Lin & He, 2012), and statistical term co-occurrence (Chu, Liu & Mao, 2002). The drawback of relevance feedback and statistical term co-occurrence methods is the analysis of previous results documents which may provide a relationship between extracted terms and the original query. But this cannot be ensured if there are no sufficient documents used for analysis before a search process.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/ranking-documents-based-on-the-semantic-relations-using-analytical-hierarchy-process/198628

Related Content

Methods for Gene Selection and Classification of Microarray Dataset

Mekour Norreddine (2018). *Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management* (pp. 66-77).

www.irma-international.org/chapter/methods-for-gene-selection-and-classification-of-microarray-dataset/197695

Domain Specific Custom Search for Quicker Information Retrieval

Tushar Kanti Saha and A. B. M. Shawkat Ali (2013). *International Journal of Information Retrieval Research* (pp. 26-39).

www.irma-international.org/article/domain-specific-custom-search-for-quicker-information-retrieval/101358

Information Retrieval Model using Uncertain Confidence's Network

Fatiha Naouar, Lobna Hlaoua and Mohamed Nazih Omri (2017). *International Journal of Information Retrieval Research* (pp. 34-50).

www.irma-international.org/article/information-retrieval-model-using-uncertain-confidences-network/177281

Adoption of Institutional Repositories Towards Realization of Digital Libraries: The Southern African Perspective

Mass Masona Tapfuma and Ruth Geraldine Hoskins (2022). *Innovative Technologies for Enhancing Knowledge Access in Academic Libraries* (pp. 156-175).

www.irma-international.org/chapter/adoption-of-institutional-repositories-towards-realization-of-digital-libraries/306435

Latent Semantic Analysis for Text Mining and Beyond

Anne Kao, Steve Poteet, Jason Wu, William Ferng, Rod Tjoelker and Lesley Quach (2012). *Intelligent Multimedia Databases and Information Retrieval: Advancing Applications and Technologies* (pp. 253-280).

www.irma-international.org/chapter/latent-semantic-analysis-text-mining/59963