

Public Abstract

First Name:Jingqian

Middle Name:

Last Name:Xu

Adviser's First Name:Xinhua

Adviser's Last Name:Zhuang

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:SS 2008

Department:Computer Science

Degree:MS

Title:Full Similarity-based Page Ranking

The World Wide Web creates many new challenges to information retrieval [1]. As the information on Web grows so rapidly, the need of a user efficiently searching some specific piece of information becomes increasingly imperative. The sheer mass and almost anarchic structure of the Web makes effective search difficult. Some good search engine alleviate the problem to some extent by ranking the search results based on the relevancy of the Web pages to user's query. They aim to place the most prominent pages at high ranks.

Most of current search engines work by first retrieving a set of Web pages based on traditional text-based search engine and then applying link-based page ranking algorithms to rank this set of Web pages. Though those link-based ranking algorithm do extract the potential relationship among the Web pages and improve the Internet search to some extent, they have several problems. One of the most important problems is computation complexity since the convergence of those eigenvector-based ranking algorithms requires iteration which is computationally expensive.

Full Similarity-based Ranking (FSBR) using densely connected clustering, a novel approach for Web page ranking, is proposed by Prof. Xinhua Zhuang. Under his advising, I did thorough literature overview, proposed a novel Subgraph Chaining Expansion algorithm, built testbed, implemented FSBR algorithm, and conducted simulation and extensive experiments.

FSBR is a generic full similarity-based ranking scheme. It allows similarity measures built on link structure and other ranking contributable features. It finds similarity-based densely connected clusters, which are believed to play a significant role in page ranking and topic analysis, and use them in page ranking. Unlike popular eigenvector-based ranking algorithms, our scheme does not require complex iterations; hence it has a much lower computational complexity. The experimental results also show that FSBR provides much higher accuracy than the HITS page ranking algorithm.

Keywords: Information retrieval, search engine, page ranking, link-based page ranking, text-based page ranking, clustering.