FULL SIMILARITY-BASED PAGE RANKING

Jingqian Xu

Prof. Xinhua Zhuang, Thesis Supervisor

ABSTRACT

The World Wide Web creates many new challenges to information retrieval. The sheer mass and almost anarchic structure of the Web makes effective search difficult. Some good search engines 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. Current page ranking algorithms 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 and uses them in page ranking. The experimental results also show that FSBR provides much higher accuracy than the HITS page ranking algorithm.