

# *MFCRank*: A Web Ranking Algorithm Based on Correlation of Multiple Features

Yunming Ye<sup>1</sup>, Yan Li<sup>1</sup>, Xiaofei Xu<sup>1</sup>, Joshua Huang<sup>2</sup>, and Xiaojun Chen<sup>1</sup>

<sup>1</sup> Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China

yym\_sjtu@yahoo.com.cn

<sup>2</sup> E-Business Technology Institute, The University of Hong Kong, Hong Kong  
jhuang@eti.hku.hk

**Abstract.** This paper presents a new ranking algorithm *MFCRank* for topic-specific Web search systems. The basic idea is to correlate two types of similarity information into a unified link analysis model so that the rich content and link features in Web collections can be exploited efficiently to improve the ranking performance. First, a new surfer model *JBC* is proposed, under which the topic similarity information among neighborhood pages is used to weigh the jumping probability of the surfer and to direct the surfing activities. Secondly, as *JBC* surfer model is still query-independent, a correlation between the query and *JBC* is essential. This is implemented by the definition of *MFCRank* score, which is the linear combination of *JBC* score and the similarity value between the query and the matched pages. Through the two correlation steps, the features contained in the plain text, link structure, anchor text and user query can be smoothly correlated in one single ranking model. Ranking experiments have been carried out on a set of topic-specific Web page collections. Experimental results showed that our algorithm gained great improvement with regard to the ranking precision.

**Keywords:** Ranking, Search Engine, Link Analysis, *PageRank*, Web

## 1 Introduction

The enormous volume of the Web presents a big challenge to Web search, as there are always too many results returned for specific queries, and going through the entire results to find the desired information is very time-consuming for the user. To improve the information retrieval efficiency, Web search engines need to employ a suitable page ranking strategy to correctly rank the search results so that the most relevant (or important) pages will be included in the top list of the search results.

In traditional information retrieval, ranking measures, such as *TF\*IDF* [1], usually rely on the text features alone to rate plain text documents. This strategy can give poor results on the Web, due to the fact that the indexed Web