

Web Search Model for Dynamic and Fuzzy Directory Search

Bumghi Choi¹, Ju-Hong Lee¹, Sun Park², Tae-Su Park²

School of Computer Science and Engineering, Inha University, Incheon, Korea

¹{neural, juhong}@inha.ac.kr

²{sunpark, taesu}@datamining.inha.ac.kr

Abstract. In web search engines, index search used to be evaluated at a high recall rate. However, the pitfall is that users have to work hard to select relevant documents from too many search results. Skillful surfers tend to prefer the index searching method, while on the other hand, those who are not accustomed to web searching generally use the directory search method. Therefore, the directory searching method is needed as a complementary way of web searching. However, in the case that target documents for searching are obscurely categorized or users have no exact knowledge about the appropriate categories of target documents, occasionally directory search will fail to come up with satisfactory results. That is, the directory search method has a high precision and low recall rate. With this motive, we propose a novel model in which a category hierarchy is dynamically constructed. To do this, a category is regarded as a fuzzy set which includes keywords. Similarly extensible subcategories of a category can be found using fuzzy relational products. The merit of this method is to enhance the recall rate of directory search by reconstructing subcategories on the basis of similarity.

1 Introduction

The index searching method has an advantage in that it quickly searches the documents indexed by an input keyword. However, it may exhibit a critical defect by generating too many results or failing to search even a single one of the targeted documents. It is because that given keywords can't be satisfactorily matched with the subjects of the target documents, or they happen to be heteronyms or homonyms, or the target documents may not be properly indexed by the keywords inside them.

In spite of many advantages of the index searching method and many efforts to improve its efficiency, we absolutely need the directory search as a complementary method of the index search. Especially for beginners, the directory searching method is preferred because it can zoom in more detailed subjects by reconstructing the subcategories of a category in a fast manner if they are familiar with the exact information of the categorization of the search subjects. However, if users don't know the categories regarding the subjects of the target documents, or if documents are not exactly categorized, it can't provide users with satisfactory results, and occasionally it causes inconvenience by navigating too many categories before reaching the targets[4, 5, 6].