TITLE	METASEARCH ENGINE USING FUZZY LOGIC AND CONTENT-BASED CLUSTERING APPROACHES
AUTHOR	WERASAK YUOKOOLABDEE
DEGREE	MASTER OF SCIENCE PROGRAMME IN COMPUTER SCIENCE
FACULTY	FACULTY OF SCIENCE
ADVISOR	CHOMTIP PORNPANOMCHAI
CO-ADVISOR	DAMRAS WONGSAWANG

ABSTRACT

Due to the enormous data on the internet it is very difficult for the web users to find the information they want. Search engines have been created to solve the problem by matching the query with the entry data in the database and retrieving the information for web users. However, the amount of data on the internet is still increasing rapidly. Gathering all available data on the internet and keeping it up to date is very hard for a single search engine. Therefore, web developers have developed a system that can send a query to several search engines at one time and integrate the results into one list. The system is called a metasearch engine. Its features enhance time saving, increase searching coverage and help users to find relevant information more quickly than using several single search engines. Normally, in searching data, not all terms are equally important in a query. Thus, it is important to allow users to indicate the relative importance of various terms by weighting them. It is noteworthy that weighting may have an impact on the entire set of retrieved document. This thesis proposes the use of a fuzzy logic ranking. The fuzzy logic ranking will retrieve the information followed by terms that web users can weight by themselves. Furthermore, the ranking mechanism works well when web users can formulate a well-defined query for searches. However, a lot of users formulate very short queries and are unfamiliar with the topics they are looking for, therefore, most results satisfy their needs. In solving such a problem, a content-based clustering (CBC) technique is considered very helpful as it will divide the results into several groups, which may be matched to the areas of the users interest. CBC aims to yield the clusters that can fulfill a precision of results. In addition, aprototype of a metasearch engine using fuzzy logic and content-based clustering was developed as proof of the proposed concept. Although the system takes several minutes to retrieve all of the results pages, the experiments show that this system outperforms other systems.

KEYWORD

METASEARCH ENGINE / FUZZY LOGIC /CLUSTERING/ INFORMATION RETRIEVAL