Abstract

A focused crawler downloads web pages that are relevant to a user specified topic. The downloaded documents are indexed with a view to optimize speed and performance in finding relevant documents for a search query at the search engine side. However, the information will be more relevant if the context of the topic is also made available to the retrieval system. This paper proposes a technique for indexing the keyword extracted from the web documents along with their contexts wherein it uses a height balanced binary search (AVL) tree, for indexing purpose to enhance the performance of the retrieval system.

References

- Yang Yongsheng and Wang Hui, “Implementation of Focused Crawler”,
COMP630D Course Project Report.

Index Terms

Computer Science Information Retrieval

Keywords

AVL tree contextual repository balance factor