

DESIGN OF ONLINE QUERY CACHING PREDICTION MODEL FOR WEB DATA RETRIEVAL

BISWAJIT SAHA AND P. MAJUMDER

Abstract

The World Wide Web (WWW) is an ever growing sea of information. The volume of data available online increases exponentially, as more and more people and entities publish data on the web. The web has no authorship standards, no editorial board and no imposed topical hierarchy. In this chaotic arena of text, images, voice and video, hundreds of millions of people from all over the world are surfing in search of information on every possible work of life. Web search is made possible by complex information retrieval system called search engines. These systems, which combine such disciplines as database technology, distributed computing and storage, statistical linguistics and graph algorithms are presented with queries of users. A user expects their queries to be instantaneously answered with ranked list of the most relevant URLs available online for each query. In order to meet these demands search engines must collect and index billions of resources, and develop highly efficient retrieval and ranking algorithms that are capable of effectively answering queries in almost blink of the eye. This research will focus on the caching and prefetching of search results in search engines. Successful caching of results can significantly lower the computational resources required for executing the stream of queries with which the engines are faced. In this paper, we will propose a strategy for effective paging of search results that minimizes the expected computational cost also examine the link structure analysis on the retrieval and ranking components of search engines.

Keywords : WWW, Search engine, Caching, Paging, Probability.