A Job Search System using Breadth First Crawling and Inverted Index

Thane Sat Oo, Su Thawda Win

University of Computer Studies, Yangon

tsatoo@googlegmail.com

Abstract

The World Wide Web has grown from a few thousand pages to billion of pages at present. Due to this explosion in size, web search systems are becoming increasingly important as the primary means of locating relevant information. Such search systems collect web pages with the help of web crawlers, which transverse the link by following hyperlinks and storing downloaded pages in a database that is later indexed for efficient execution of user queries. In this paper, we develop a keyword based search system that can search jobs and help users who want to retrieve the information about the job from Internet. The major components of Job Search System are Web Crawlings, Indexing and Searching. The crawler in this system uses breadth first method to crawl the local sites. The resulted keywords are indexed using inverted index structure, so that it can answer user queries in fast respond time. This paper is implemented with Java SE (Crawler and Indexer) and Java Servlet (for searching interface).