ABSTRACT

The Internet is the biggest repository of online knowledge wherein the end-user employs tools for searching and sharing information. It has revolutionized the world of computer and communications like never before and within a span of few years it has changed the way of business and communication. The World Wide Web is a hyperlinked repository of trillions of hypertext documents lying in different websites, distributed over far-end distant geographical locations. With such a huge web repository, finding the right information at the right time is a challenging task. Hence, there is a need for a very effective way of retrieving information from the web.

Web search engines act as a bridge between web users and web pages. Without search engines, the unlimited source of information stored in web pages remain hidden for us. Web search engines employ web crawlers to continuously collect web pages from the web, index and store them in a database. To maintain the freshness of pages in the collection, a crawler has to revisit the websites again and again due to which resources employed become overloaded. In traditional crawling, the pages from all over the web are brought to the search engine side and then processed, that results in a lot of network traffic.

Mobile (migrating) agents are computational software processes that can be employed to roam around the wide area networks such as the web, to gather the information on behalf of its owner and may come back to its mother system. In distributed crawling based on migrating agents, the agents allow packaging a conversation and dispatching it to a destination host, where the interactions can take place locally. When very large volumes of data are stored at remote hosts, these data should be processed in the locality of the data rather that transferred over the network. Due to their inherent nature of roaming around the web, the migrating agent may cause security threats for the data it carries, platform it uses and with itself.

This work utilizes some of the best features of migrating crawling agents (migrants) for managing the process of selection; filtration and compression at server. It allows local access of web documents and minimizes network utilization significantly. It uses a more network efficient approach for extracting the volatile information from the web server.
using migrants with the help of table of volatile information, which further helps in reducing the network load significantly. This enables migrating agents to select only modified information and bring it to the search engine side, leaving all other static information which already is in the search engine repository. An alternate approach has been used for optimizing the revisit frequency of migrants for visiting websites based on user’s interest. The architecture manages the process of revisiting to a website with a view to maintain fairly fresh documents at the search engine side. It adjusts the frequency of revisit by dynamically assigning a priority of revisiting to a side by computing the priority based on previous experience that how many times the crawler founds changes in contents and the interest of the users shown in the websites. The pages visited by the users more, be given high priority as compared to those that are less or rarely visited. A user perception based inverted index mechanism is being used to organise large indices that provides high-speed access to sets of documents satisfying user queries. It helps the web search engines to provide better results to the query based on the user perception i.e. the word for which he is intended to search. To fix issues like maintaining security and integrity of the agent, data it carries and the remote platform on which it executes, a remote platform oriented reliability based security system approach has been used. It manages the security of a migrant by computing its reliability factor. The level of restriction is in proportionate to the reliability that means as the reliability increases, the restriction of the environment decreases. In case the reliability of a migrant is very low, it provides a restricted secured environment to the migrating agent and its restriction be decreased in incremental manner as reliability of the agent increases. The results of the proposed approach are found as per our expectation.