Abstract

With the vast growth of the Internet, many web pages are available online. Search engines use a component called as web crawlers for collecting these web pages from the web for storage and indexing. Many web pages are autonomous and are updated independent of the users. As the web pages are updated autonomously; users do not come to know of how often the sources change. An incremental crawler visits the web repeatedly after a specific interval of time for the updation of its collection. Users are benefited by knowing the page importance based upon the page accessing frequency. This paper finds out the page importance based on page accessing frequency and also architecture for the same is also proposed.

Index Terms
Keywords

Search Engine  Web Crawler  Page access Frequency.