The Web crawler is a computer program that downloads data or information from World Wide Web for search engine. Web information is changed or updated rapidly without any information or notice. Web crawler searches the web for updated or new information. Approximate 40% of web traffic is by web crawler. In this paper a web or network traffic solution has been proposed. The method of web crawling with filter is used. This approach is query based approach. The proposed approach solves the problem of revisiting web pages by crawler.

References

- Alpert, Jesse; Hajaj, Nissan (July 25, 2008). "We knew the web was big..." The Official Google Blog.
- Bal. S and Nath. R, "Filtering the web pages that are not modified at remote site without downloading using mobile crawler." Information Technology journal 9(2)2010 ISSN 1812-5638, Asian Network for Scientific information. (pp: 376-380)
- Shekhar mishra, anurag jain and A K Sachhan, "Smart approach to Reduce the Web Crawler Traffic of existing system using HTML based update file at web server".

Index Terms

Computer Science  Information Retrieval
Keywords
Web  Search Engine  Web Crawler  Web Crawling Traffic  Http Get Request