Abstract

The World Wide Web is a huge source of hyperlinked information contained in hypertext documents. Search engines use web crawlers to collect these documents from web for the purpose of storage and indexing. However, many of these documents contain dynamic information which gets changed on daily, weekly, monthly or yearly basis and hence we need to refresh the search engine side storage so that latest information is made available to the user. An incremental crawler visits the web repeatedly after a specific interval for updating its collection. In this paper to regulate the revisiting frequency a novel mechanism and a novel architecture for incremental crawler is being proposed.

Reference

3. Alexandros Ntoulas, Junghoo Cho, Christopher Olston, “What’s new on the Web? The

"Searching the Web." ACM Transactions on Internet Technology, 1(1): August 2001

Architecture For Incremental Web Crawler”, International Journal of Computer Science and

www.isoc.org/internet/history.


Services & Use”, 25 (2005) 137-147, IOS Press, 2005

vol. 2, no. 4, pp. 219-229, 1999.

the 8th International World Wide Web Conference (WWW8), pages 1467-1479, 1999.

implications for an incremental crawler”, In Proceedings of the 26th International Conference on
Very Large Databases.

Submitted to VLDB 2000, Research track.

(DSIM)”, International Journal of Computer Science and Network Security (IJCSNS), Vol 8,
No12, Dec 2008.


Web Techniques Magazine, 2(5), May 1997

107—117, April 1998.

through examples”, Proceedings of the 25th International Conference on Very Large Databases

Index Terms

Computer Science

Internet Applications
**Key words**

web search engine  
incremental crawler  
dynamic  
information  
crawl workers