Abstract

Traditional search engines use a thin client, distributed model for crawling. This crawler based approach has certain drawbacks which could be removed with a proposed rich client based model. The rich client based search engine offers faster crawling and better updation time using lesser resources than thin client model, and it covers more of the World Wide Web than normal crawler based search engines. Although modern day search engine giants have
improvised on various features such as ergonomics and utilities, along with several added
goodies, little work is done to improve energy efficiency of such Large Scale Search Engines.
As the Internet is increasing exponentially the search engines will involve more and more
servers thus costing more and more energy. This ever increasing demand of search engines
needs to be curbed down. Rather than multiplying server resources it is better to use existing
servers which work in a congenial environment, using communication methods to reduce
redundant downloading of data from different servers by the crawlers. This paper proposes a
rich client based architecture for search engines along with analysis and comparison with
present search engines. This could help into reducing the challenges of global warming,
keeping up the speed and efficiency requirements.

References

- Dunham, M., 2003. "Data Mining: Introductory And Advanced Topics";
  1/e, Pearson Education.
  "Crawler-friendly web servers";
  Department of Electrical & Computer Engineering, University of California, San Diego, Citeseer.
  Association of Small Computer Users in Education &quot;Our Second Quarter Century of Resource Sharing&quo;
  pp. 84.
  "Searching the web";
  ACM Transactions on Internet Technology (TOIT), ACM, Vol. 1(1), pp. 43.

Index Terms

Computer Science

Information Sciences
Keywords
Search Engines  Thick Client  Rich Client  Updation Delay  And Crawler.