IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-3, Issue-10 E-ISSN: 2347-2693

Survey on Mobile Optimized Search Crawler

Kukreja Kajal[1*], Gavali Nishigandha[2] and Khedlekar Gandhali[3]

[1*, 2, 3] *R. H. Sapat College of Engineering, Management Studies & Research, Nashik, Maharashtra, India*

www.ijcseonline.org

Received: Oct/03/2015 Revised: Oct/11/2015 Accepted: Oct/24/2015 Published: Oct/31/ 2015

***Abstract* —** The web crawler is the central component of a search engine which works like an indexer, finds out hyperlinks and computes keyword density of each web page. It assigns a page rank to each crawled web page by using some ranking algorithm and stores the visited links for the future use. Search results retrieve very fast from desktop browser, but it takes more time when user is on mobile. When a keyword is searched from a mobile browser, the traditional web servers take a long time to interpret this request. Also, the web server has to format search results into the form which mobile device can interpret. Thus, to eliminate this overhead on the web server, a Mobile Application Server has to be introduced instead of the web server to interpret requests from mobile devices.

***Keywords* —** Search engine; Crawler; Web server; Mobile application server; Wireless markup language; Wireless application protocol.

I. INTRODUCTION

A search engine is a system that searches for information on the World Wide Web and finds out the web pages that contain information related to the search keyword. This information can be in the form of images, audio, video or text. "Search engines automatically create web site listings by using spiders that crawl the web pages, index their information, and optimally follows that site's links to other pages.[12]" A web crawler is a program that crawls web pages on the World Wide Web to read visible text, hyperlinks and content of the various tags used in the site, such as keyword rich meta tags. Using this information that is gathered by the crawler, a search engine determines what the site is about, computes its keyword density and assigns it a page rank by using some ranking algorithm. A ranking algorithm ranks the web pages based on several factors like relevance, authority, novelty, etc. It is the core technique of search engine that is very useful to improve the search results' quality. Various page ranking algorithms are PageRank, ELO, HITS, Weighted PageRank, etc.

Mobile Application Server (MAS) is a server that hosts, installs and operates mobile applications and other services. It is similar to a web server which stores, processes and delivers web pages to the clients. It bridges the gap between existing infrastructures of servers to the mobile devices. Popular mobile application servers include IBM's WebSphere, IBM's MobileFirst Platform, Contec's Hornet, @Hand Mobile Application Server, etc.
[1]

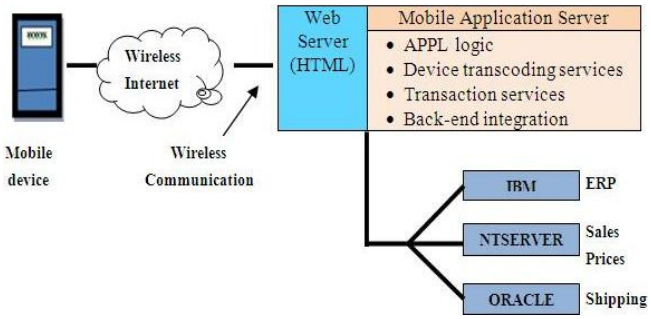


Fig. 1 "Architecture of a mobile application server [11]"

This paper is organized in six sections as follows - Section 2 discusses the motivation for this survey, Section 3 shows the related work, Section 4 demonstrates the comparative study and Section 5 concludes with objectives for the future work.

II. MOTIVATION

Whenever user sends any search request from mobile, the mobile devices sends this request to the service provider's gateway server using Wireless Application Protocol (WAP). The gateway server then retrieves the information via HTTP from the web server. It encodes this HTTP data in a language understood by the mobile device called as Wireless Markup Language (WML). This WML-encoded data is then sent to the device which made the search request. Thus, the user sees the wireless Internet version of the web pages. This conversion of data from WML to HTML and vice-versa consumes lots of resources and time.

Thus, there is a need to eliminate this conversion process and introduce a server, which uses same language used by the mobile devices. This can be done by introducing a mobile application server to interpret the requests made

[1]Contact Author - Kukreja Kajal (www.kajalkukreja.com)

© 2015, IJCSE All Rights Reserved

from the mobile devices. It will also provide fast, accurate and optimized search results to the mobile users within the optimum time span.

III. LITERARURE SURVEY

Feng Zhao, Jingyu Zhou, Chang Nie, Heqing Huang and Hai Jin proposed "SmartCrawler : A Two-stage Crawler for Efficiently Harvesting Deep-Web Interfaces" which performed site-based searching for the center pages with the help of search engines, avoiding visiting a large number of pages. "To achieve more accurate results for a focused crawl, *SmartCrawler* ranks websites to prioritize highly relevant ones for a given topic. In the second stage, *SmartCrawler* achieves fast in-site searching by excavating most relevant links with an adaptive link-ranking.[1]" The only drawback of this system was that crawler was implemented on web server that takes more time when a search request is made from a mobile device.

Mehdi Bahrami, Mukesh Singhal and Zixuan Zhuang proposed "A Cloud-based Web Crawler Architecture" which used cloud computing features and the MapReduce programming technique to crawl the web. Crawling was done by distributed agents with each agent storing its own finding on a Cloud Azure Table (NoSQL database). "A cloud-based web crawler allows people to collect and mine web content without buying, installing and maintaining any infrastructure.[2]" This web crawler stored unstructured and massive amount of data on Azure Blob storage. They analyzed the performance and scalability of web crawler and described its advantages over traditional distributed web crawlers.

Kausar M. A., Nasar M. and Singh S. K. implemented the concept of "Maintaining the repository of search engine freshness using mobile crawler". They proposed a system based on web crawler that used mobile agent. They used Java Aglets to crawl the web pages. The major advantage of web crawler based on Mobile Agents was that the analysis part of the crawling process was done locally at the residence of the data rather than in the remote server. This considerably reduced network load and traffic that improved the performance and efficiency of the crawling process.

Jian-Hong Liu, Jing Chen, Yi-Li Wu and Pei-Li Wang were motivated by the widespread use of cloud computing and popularity of mobile platforms, such as smart phones and tablet computers running Android system. Thus, they developed "AASMP - Android Application Server for Mobile Platforms". "The main purposes of developing AASMP include - to allow deploying Android applications on a server to be accessed by client-side users, not necessarily operating an Android platform, through a browser software without installing any plug-in modules; to support offloading the execution of existing Android applications to powerful server-side environment with neither modifying nor porting needed; to lay a foundation for developing server-side Android applications or services which are provided with the resources normally limited on mobile platforms.[4]"

Vinay Kancherla proposed "Smart Crawler for a Concept based Semantic Search Engine" which crawled all over the internet for collecting web pages and storing them in the form of text files. It performed an initial data analysis of unnecessary data before storing it. "We aim to improve the efficiency of the Concept Based Semantic Search Engine by using the Smart crawler.[5]" It drastically improved the efficiency of the Concept Based Semantic Search Engine. The main drawback of this system was that it took inputs only in the form of text files.

Pavalam S. M., S. V. Kasmir Raja, Jawahar M. and Felix K. Akorli researched on "Web Crawler in Mobile Systems" and explained various concepts of web crawlers. "Web crawler is the central part of the search engine which browses through the hyperlinks and stores the visited links for the future use.[6]" They also explained the ways in which crawlers can be used in mobile systems and explored the different kinds of software used in mobile phones for crawling purposes. Thus, they identified the advantages of crawlers in mobile communications.

Beena Mahar and C K Jha presented "A Comparative Study on Web Crawling for searching Hidden Web" for describing the concepts of web crawler, types of web crawler and architecture for searching the hidden web documents. They explained different crawling technologies and different ways to crawl in the search of hidden web documents. "Total quality content of the deep web is atleast 1000-2000 times greater than that of the surface web.[7]" Thus, they concluded that HiddenWeb is important because it retrieves high quality information. Therefore, there was a need to implement an indexing technique to efficiently index the high quality data.

Vladislav Shkapenyuk and Torsten Suel proposed "Design and implementation of a high-performance distributed web crawler" which ran on network of workstations and crawled hundreds of web pages simultaneously for finding out relevant information. "The crawler scales to (at least) several hundred pages per second, is resilient against system crashes and other events, and can be adapted to various crawling applications.[8]"

Jason Rennie and Andrew McCallum researched on "Using reinforcement learning to spider the web efficiently" and concluded that the creation of efficient web crawler is best framed and solved by reinforcement learning, a branch of

© 2015, IJCSE All Rights Reserved

machine learning that concerns itself with optimal sequential decision-making. "One strength of reinforcement learning is that it provides a formalism for measuring the utility of actions that give benefit only in the future.[9]"

Hardik P. Trivedi, Gaurav N. Daxini, Jignesh A. Oswal, Vinay D. Gor and Swati Mali presented "An Approach to Design Personalized Focused Crawler" in which they defined web page change detection policy for the design of a focused crawler. "The motivation behind developing a Personalized Focused Crawler is to provide targeted information to user i.e. providing information based on user's interest solely.[10]"

IV. COMPARATIVE STUDY

The following table shows a comparative study between the time (in seconds) taken by desktop browser, Google mobile widget and mobile browser to search a string/keyword.

String/keyword (input)	Time taken (in seconds)		
	Desktop browser	*Google widget*	*Mobile browser*
create a website	0.39	1.841	5.223
learn maths in few minutes	0.41	3.637	6.622
convert word to pdf	0.50	3.270	6.647
Structure ^	0.41	2.26	1.84
What is meant by definition ^	0.52	3.468	1.53
constructor	0.49	4.535	6.508
function^	0.49	2.786	4.371
instance	0.60	2.716	3.917
how to use android studio	0.51	3.005	3.895
create a structure	0.58	4.428	4.6
make something public	0.40	2.426	3.944

a. ^ denotes irrelevant results were displayed by existing search engine.

From this table, we can conclude that desktop browser requires minimum amount of time to fetch the search results whereas mobile browser requires maximum amount of time to retrieve the same results. We can also conclude that Google mobile widget takes considerable time to retrieve search results but it's less as compared to mobile browser.

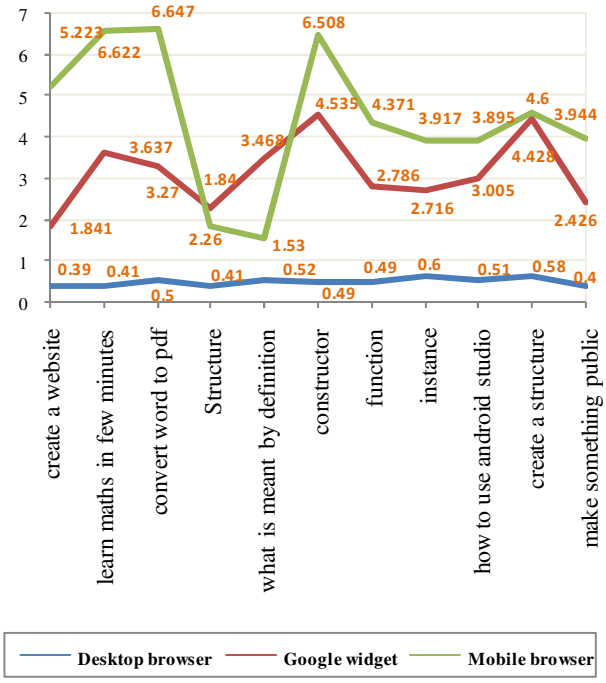


Fig. 2 Graph showing time (in seconds) taken by desktop browser, Google mobile widget and mobile browser.

V. CONCLUSION

Through this survey and comparative study of time (in seconds) taken by desktop browser, mobile browser and a Google mobile widget for searching a keyword, we can conclude that mobile browser requires more amount of time than the desktop browser to retrieve the search results. So, there is a need to develop a web crawler which will be specially optimized for mobile devices. As compared to traditional web crawlers, it must drastically reduce the amount of time required for retrieving search results on mobile devices. It must fulfill three main objectives - reduce the time required for retrieving the search results, eliminate the conversion process required for fetching the results and obtain accurate search results for the mobile users in optimized time. In this way, we have concluded that web crawler has a big scope for mobile systems.

ACKNOWLEDGMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without mentioning the people who made it possible. We are grateful to a number of individuals whose professional guidance along with encouragement have made it very pleasant endeavor to complete this survey.

We have a great pleasure in presenting "Survey of Mobile Optimized Search Crawler" under the guidance of Prof. Dr.

© 2015, IJCSE All Rights Reserved

S. V. Gumaste and our project coordinator Prof. C. R. Barde. We are truly indebted and grateful to Head of Computer Engineering Department Prof. N. V. Alone for his valuable guidance and encouragement.

We would also like to thank Gokhale Education Society's R. H. Sapat College of Engineering, Management Studies and Research, Nashik-5 for providing the required facilities, internet access and important books. At last, we must express our heartfelt gratitude to all the teaching and non-teaching staff members of Computer Engineering Department who helped us by giving their valuable time, support, comments and suggestions.

REFERENCES

[1] Feng Zhao, Jingyu Zhou, Chang Nie, Heqing Huang and Hai Jin, "SmartCrawler : A two-stage crawler for efficiently harvesting deep-web interfaces", Services Computing, IEEE Transactions, Volume-**PP**, Issue-**99**, DOI-**10.11.09**, **2015**.

[2] Mehdi Bahrami, Mukesh Singhal and Zixuan Zhuang, "A Cloud-based Web Crawler Architecture", Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference, Page No **(216-233)**, 17-19 Feb **2015**.

[3] Kausar M. A., Nasar M. and Singh S. K., "Maintaining the repository of search engine freshness using mobile crawler", Emerging Research Areas and 2013 International Conference on Microelectronics, Communications and Renewable Energy (AICERA/ICMiCR), 2013 Annual International Conference, Page No (**1-6**), **2013**.

[4] Jian-Hong Liu, Jing Chen, Yi-Li Wu and Pei-Li Wang, "AASMP - Android Application Server for Mobile Platforms", IEEE 2013 16th International Conference on Computational Science and Engineering, Page No (**643-650**), **2013**.

[5] Vinay Kancherla, "Smart Crawler for a Concept based Semantic Search Engine", Master's Theses and Graduate Research at SJSU ScholarWorks, Paper **380**, 12 Jan **2014**.

[6] Pavalam S.M, S.V. KumarRaja, M. Jawhar and Felix K. Akorli, "Web crawler in mobile systems", International Journal of Machine Learning and Computing, Volume-**02**, Issue-**04**, Page No (**531-534**), August **2012**.

[7] Beena Mahar and C K Jha, "A comparative study on Web crawling for searching Hidden web", International Journal of Computer Science and Information Technologies, Volume-**02**, Issue-**03**, Page No (**2159-2163**), **2015**.

[8] Vladislav Shkapenyuk and Torsten Suel, "Design and implementation of a high-performance distributed Web crawler", Data Engineering, 2002. IEEE Proceedings. 18th International Conference, Page No (**357-368**), **2002**.

[9] Jason Rennie and Andrew McCallum, "Using reinforcement learning to spider the web efficiently", Proceedings of the 16th International Conference on Machine Learning (ICML), Page No (**335-343**), **1999**.

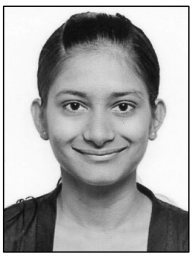
[10] Hardik P. Trivedi, Gaurav N. Daxini, Jignesh A. Oswal, Vinay D. Gor and Swati Mali, "An Approach to Design Personalized Focused Crawler", International Journal of Computer Sciences and Engineering, Volume-**02**, Issue-**03**, Page No (**144-147**), March **2014**.

[11] "Mobile Application Server", http://www.mobileinfo.com/application_servers.htm , 4 July, 2015.

[12] "Search Engine", http://websearch.about.com/od/enginesanddirectories/a/searchengine.htm , 15 July, 2015.

AUTHORS' PROFILE

Miss Kajal Kukreja has completed Diploma in Information Technology with distinction from Government Polytechnic, Nashik. She is pursuing Bachelors degree in Computer Engineering from R. H. Sapat College of Engineering, Management Studies & Research, Nashik. After working as a web developer for three months in a company, she is currently working as a freelancer and has deployed more than 65 projects till date. She is a passionate web developer and software programmer with an ability to convert client requirements into innovative solutions. She has designed her professional website with URL http://www.kajalkukreja.com



Miss Nishigandha Gavali has completed Diploma in Computer Technology from Government Polytechnic, Nashik. She is pursuing Bachelors degree in Computer Engineering from R. H. Sapat College of Engineering, Management Studies & Research, Nashik. She looks forward to have a successful career in Graphic Designing, Virtualization, Cloud Computing and Distributed Computing.

Miss Gandhali Khedlekar is pursuing Bachelors degree in Computer Engineering from R. H. Sapat College of Engineering, Management Studies & Research, Nashik. She has completed HSC (12^{th}) from R. Y. K. College of Science, Nashik. Her areas of interest include Mobile Computing and Cloud Computing.

© 2015, IJCSE All Rights Reserved