Application of ARIMA(1,1,0) Model for Predicting Time Delay of Search Engine Crawlers

Jeeva JOSE[1], P. Sojan LAL[2]
[1] Department of Computer Applications, BPC College, Piravom, Kerala, India
[2] School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India
vijojeeva@yahoo.co.in, padikkakudy@gmail.com

World Wide Web is growing at a tremendous rate in terms of the number of visitors and number of web pages. Search engine crawlers are highly automated programs that periodically visit the web and index web pages. The behavior of search engines could be used in analyzing server load, quality of search engines, dynamics of search engine crawlers, ethics of search engines etc. The more the number of visits of a crawler to a web site, the more it contributes to the workload. The time delay between two consecutive visits of a crawler determines the dynamicity of the crawlers. The ARIMA(1,1,0) Model in time series analysis works well with the forecasting of the time delay between the visits of search crawlers at web sites. We considered 5 search engine crawlers, all of which could be modeled using ARIMA(1,1,0).The results of this study is useful in analyzing the server load.

Keywords: *ARIMA, Search Engine Crawler, Web logs, Time delay, Prediction*

1 Introduction

Crawlers also known as ‘bots’, ‘robots’ or ‘spiders’ are highly automated programs which are seldom regulated manually[1][2]. Crawlers form the basic building blocks of search engines which periodically visit the web sites, identify new web sites, update the new information and index the web pages in search engine archives. The log files generated at web sites play a vital role in analyzing user as well as the behavior of the crawlers. Most of the works in web usage mining or web log mining is related to user behavior as they have application in target advertising, online sales and marketing, market basket analysis, personalization etc. There is open source software available like Google Analytics which measures the number of visitors, duration of the visits, the demographic from which the visitor comes etc. But it cannot identify search engine visits because Google Analytics track users with the help of JavaScript and search engine crawlers do not enable the JavaScript embedded in web pages when the crawlers visit the web sites [3].

The search engine crawlers initially access the robots.txt file which specifies the Robot Exclusion Protocol. Robots.txt is a text file kept at the root of the web site directory. The crawlers are supposed to access this file first before it crawls the web pages. The crawlers which access this file first and proceeds to crawling are known as ethical crawlers and other crawlers who do not access this file are called unethical crawlers. The robots.txt file contains the information about which pages are allowed for crawling and which all folders and pages are denied access. Certain pages and folders are denied access because they contain sensitive information which is not intended to be publically available. There may be situations where two or more versions of a page will be available one as html and other one as pdf. The crawlers can be made do avoid crawling the pdf version to avoid redundant crawling. Also certain files like JavaScript, images, style sheets etc. can be avoided for saving the time and bandwidth. There are two ways to do this. First one is with the help of robots meta tag and the other one is with the help of robots.txt file. The robots.txt file contains the list of all user agents and the folders or pages which are disallowed [30]. The structure of a robots.txt file is follows.

```
User-agent:
Disallow:
```

“User-agent:” is the search engine crawler and “Disallow:” lists the files and directories to be excluded from indexing. In addition to

DOI: 10.12948/issn14531305/17.4.2013.03

"User-agent:" and "Disallow:" entries, comment lines are included by putting the # sign at the beginning of the line. For example all user agents are disallowed from accessing the folder /a.# All user agents are disallowed to see the /a directory.

```
User-agent: *
Disallow: /a/
```

The crawlers which initially access the robots.txt and then the subsequent files or folders are known as ethical crawlers whereas others are known as unethical crawlers. Some crawlers like "Googlebot", "Yahoo! Slurp" and "MSNbot" cache the robots.txt file for a web site and hence during the modification of robots.txt file, these robots may disobey the rules. Roughly, a crawler starts off with the URL for an initial page p_0. It retrieves p_0, extracts any URLs in it, and adds them to a queue of URLs to be scanned. Then the crawler gets URLs from the queue (in some order), and repeats the process. Every page that is scanned is given to a client that saves the pages, creates an index for the pages, or summarizes or analyzes the content of the pages [26]. Certain crawlers avoid too much load on a server by crawling the server at a low speed during peak hours of the day and at a high speed during late night and early morning [2]. A crawler for a large search engine has to address two issues. First, it has to have a good crawling strategy, i.e., a strategy for deciding which pages to download next. Second, it needs to have a highly optimized system architecture that can download a large number of pages per second while being robust against crashes, manageable, and considerate of resources and web servers [24]. There are two important aspects in designing efficient web spiders, i.e. crawling strategy and crawling performance. Crawling strategy deals with the way the spider decides to what pages should be downloaded next. Generally, the web spider cannot download all pages on the web due to the limitation of its resources compared to the size of the web [28].

The mobile crawlers that always stay in the memory of the remote system occupy a considerable portion of it. This problem will further increase, when there are a number of mobile crawlers from different search engines.

- all these mobile crawlers will stay in the memory of the remote system and will consume lot of memory that could have otherwise been used for some other useful purposes;
- it can also happen that the remote system may not allow the mobile crawlers to reside permanently in its memory due to security reasons;
- in case a page changes very quickly then the mobile crawler immediately accesses the changed page and sends it to the search engine to maintain up-to-date index. This will result in wastage of network bandwidth and CPU cycles etc [30].

Recently web crawlers are used for focused crawling, shopbot implementation and value added services on the web. As a result more active robots are crawling on the web and many more are expected to follow which will increase the search engine traffic and web server activity [4]. The Auto Regressive Integrated Moving Average (ARIMA) Model was used to predict the time delay between two consecutive visits of a search engine crawler. We used the differenced first-order autoregressive model, ARIMA(1,1,0) for forecasting the time delay between two consecutive visits of search engine crawlers.

2 Background Literature

There are several works that mentions about the search engine crawler behavior. A forecasting model is proposed for the number of pages crawled by search engine crawlers at a web site [3]. Sun et al has conducted a large scale study of robots.txt [2]. A characterization study and metrics of search engine crawlers is done to analyze the qualitative features, periodicity of visits and the pervasiveness of visits to a web site [4]. The working of a search engine crawler is explained in [5]. Neilsen NetRatings is one of the leading internet and digital media audience information and analysis services. NetRatings have provided a study on the usage statistics of search engines in United States [6]. Com-

DOI: 10.12948/issn14531305/17.4.2013.03

mercial search engines play a lead role in World Wide Web information dissemination and access. The evidence and possible causes of search engine bias is also studied [7]. An empirical pilot study is done to see the relationship between JavaScript usage and web site usage. The intention was to establish whether JavaScript based hyperlinks attract or repel crawlers resulting in an increase or decrease in web site visibility [8]. The ethics of search engine crawlers is identified using quantitative models [9]. Analysis of the temporal behavior of search engine crawlers at web sites is also done [10]. There is a significant difference in the time delay between and among various search engine crawlers at web sites [11]. Search engines do not index sites equally, may not index new pages for months, and no engine indexes more than about 16% of the web [23]. A crawling technique to reduce the load of the network using mobile agents were developed by Bal and Nath [25]. The working of a comprehensive full text search engine called WebCrawler is also studied [27]. An optimal algorithm for distributed web crawling is done by compressing the crawled web data before sending it to the central database of the search engine and thereby reducing the load and processing bottleneck of the search engine database [29].

3 Methodology

3.1 Pre Processing

Web log files need considerable amount of preprocessing. The user traffic needs to be removed from this file as this work focuses on the search engine behavior. Improper preprocessing may bias the data mining tasks and lead to incorrect results. About 90% of the traffic generated at web sites is contributed by search engine crawlers [13]. The advantages of preprocessing are:

- the storage space is reduced as only the data relevant to web mining is stored;
- the user visits and image files are removed so that the precision of web mining is improved.

The web logs are unstructured and unformatted raw source of data. Unsuccessful status codes and entries pertaining to irrelevant data like JavaScript, images, stylesheets etc. including user information are removed. The most widely used log file formats are Common Log File Format and Extended Log File Format. The Common Log File format contains the following information: a) IP address b) authentication name c) the date-time stamp of the access d) the HTTP request e) the URL requested f) the response status g) the size of the requested file. The Extended Log File format contains additional fields like a) the referrer URL b) the browser and its version and c) the operating system or the user agent[14][15]. Usually there are three ways of HTTP requests namely GET, POST and HEAD. Most HTML files are served via GET method while most CGI functionality is served via POST or HEAD. The status code 200 is the successful status code [14].Search engines are identified from their IP addresses and user agents used for accessing the web. The log file of a business organization www.nestgroup.net of 30 days ranging from May 1, 2011 to May 31, 2011 comprising of 31 days. Table 1 shows the results of preprocessing.

Table 1. Results of Preprocessing

Total number of records	2,68,858
Number of successful search engine requests	21,230
Number of distinct search engine crawlers	17
Number of search engine crawlers after pre processing	13
Number of visits chosen	100

Those search engines whose number of visits less than 5 in a month is eliminated before further analysis. There were 13 distinct search engine crawlers. Certain search engine crawlers made several visits on one day itself where as some others made one or two visits

DOI: 10.12948/issn14531305/17.4.2013.03

a day. The prominent crawlers were Baiduspider, Bingbot, Discobot, Ezooms, Feedfetcher-Google, Googlebot, Gosospider, Ichiro, MJ12bot, MSNbot, Slurp, Sogou, Sosospider and Yandex. Some crawlers were not significant because they made less than 5 visits a month. It includes Alexa, Exabot, Magpie and Yrspider.

We have chosen 5 prominent crawlers from our data set for study. It includes Baiduspider, Bingbot, Googlebot, Feedfetcher-Google and Slurp. These crawlers were consistent in their visits and hence chosen for modeling. It is a Chinese search engine crawler which crawls the server depending on the server load. Baidu has several user agents like Baiduspider for web search, Baiduspider-mobile for mobile search, Baiduspider-image for image search, Baiduspider-video for video search, Baiduspider-news for news search, Baiduspider-favo for bookmark search and Baiduspider-ads for business search. Bingbot is the crawler for bing search engine. Both Googlebot and Feedfetcher-Google are crawlers from Google while Slurp is the crawler for Yahoo [12].

3.2 Auto Regressive Integrated Moving Average Model (ARIMA)

Forecasting is an important aspect of statistical analysis that provides guidance for decisions in all areas. It is important to be able to make sound forecasts for variables such as sales, production, inventory, interest rates, exchange rates, real and financial asset prices for both short and long term business planning. Autoregressive Integrated Moving Average (ARIMA) models provide a unifying framework for forecasting. These models are aided by the abundance of high quality data and easy estimation and evaluation by statistical packages [16]. We found the time delay between the visits of search engine crawlers could be predicted using the ARIMA Model.

ARIMA(p,d,q): ARIMA models are, in theory, the most general class of models for forecasting a time series which can be made stationary by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random-walk and random-trend models. The fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors. Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series [20]. Lag 1 is the time period between two observations y_t and y_{t-1}. time series can also be lagged forward, y_t and y_{t+1}.

A non-seasonal ARIMA model is classified as an ARIMA(p,d,q) model, where:

- p is the number of autoregressive terms,
- d is the number of non-seasonal differences, and
- q is the number of lagged forecast errors in the prediction equation.

The autoregressive element, p, represents the lingering effects of preceding scores, the integrated element, d, represents trends in the data and q represents the lingering effects of preceding random shocks. When the time series is long, there are also tendencies for measures to vary periodically called seasonality or periodicity in time series. Time series analysis is more appropriate for data with autocorrelation. If all patterns are accounted for in the model, the residuals are random. In many applications of the time series, identifying and modeling the patterns in the data are sufficient to produce an equation, which is then used to predict the future of the process.

Model Identification

Let $y_1, y_2, y_3 \ldots y_T$ represent a sample of T observations of a variable of interest y and $\{y_t\}$ represents the time series. Since the stationary property is essential for the identification of an ARIMA model, the first step is always to test for stationary property of the underlying series. Many data in real time including the web data chosen for our study is not stationary. The series can be made stationary by

DOI: 10.12948/issn14531305/17.4.2013.03

differencing with or without pre-transformations. Formally, {y_t} is said to be stationary if the mean, E$(y_t)=\mu$, the variance $Var(y_t)$=E$(y_t - \mu)^2$ and the covariance Cov$(y_t, y_{t-s})=$ E$(y_t - \mu)(y_{t-s} - \mu)= \gamma_s$ are all stable over time. For the series to be stationary, it must not exhibit any stochastic trend (changing mean) or varying volatility (changing variance) [16] [21].

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The principle way to determine which Auto-Regressive (AR) or Moving Average(MA) model is appropriate is to look at the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the time series. The plot of the autocorrelation function and partial autocorrelation function also serves as a visual test for stationary property [18] [19]. At lag k, the ACF is computed by

$$\text{ACF}(k) = \frac{E[(yt-E[yt])(yt-k-E[yt-k])]}{\sqrt{Var[yt]\ Var[yt-k]}} \qquad (1)$$

In time series, we may want to measure the relationship between y_t and y_{t-k} when the effects of other time lags 1, 2,...,k − 1 have been removed. The autocorrelation does not measure this. However, Partial autocorrelation is a way to measure this effect. The partial autocorrelation of a time series at lag k is denoted α_k and is found as follows (1) Fit a linear regression of y_t to the first k lags (i.e. fit an AR(k) model to the time series):

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_k y_{t-k} + ȩ_t. \quad (2)$$

Then $\alpha_k = \varphi\hat{}_k$, the fitted value of φ_k from the regression (Least Squares).The set of partial autocorrelations at different lags is called the partial autocorrelation function (PACF) and is plotted like the ACF. The Box-Jenkins procedure is concerned with fitting an ARIMA model to data [17]. It has three parts: identification, estimation, and verification. Figure 1 shows the Box-Jenkin's model building process.

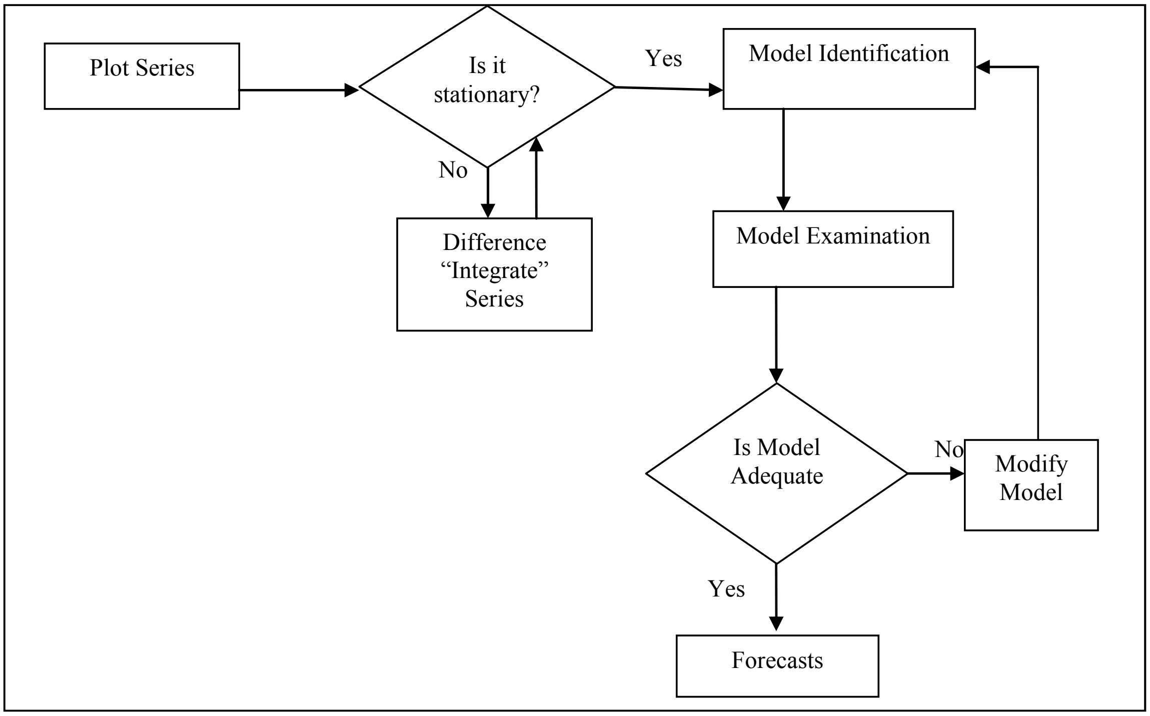

Fig. 1. Box-Jenkins Model Building Process

The Box-Jenkins approach suggests short and seasonal (long) differencing to achieve stationary in the mean, and logarithmic or power transformation to achieve stationary property in the variance. In case the series are seasonal, the Box-Jenkins methodology pro-

DOI: 10.12948/issn14531305/17.4.2013.03

poses multiplicative seasonal models coupled with long-term differencing, if necessary, to achieve stationary property in the mean. The difficulty with such an approach is that there is practically never enough data available to determine the appropriate level of the seasonal ARMA model with any reasonable degree of confidence. Users therefore proceed through trial and error in both identifying an appropriate seasonal model and also in selecting the right long-term (seasonal) differencing. In addition, seasonality complicates the utilization of ARMA models as it requires using many more data while increasing the modelling options available and making the selection of an appropriate model more difficult [22].

We have chosen 100 time delay between consecutive visits for the crawlers Baiduspider, Bingbot, Googlebot, Feedtetcher-Google and Slurp. The time delay in seconds were plotted and Autocorrelation Function(ACF) and Partial Autocorrelation Function(PACF) were plotted. The obtained plots for Baiduspider are given in Figure 2 and Figure 3 respectively.

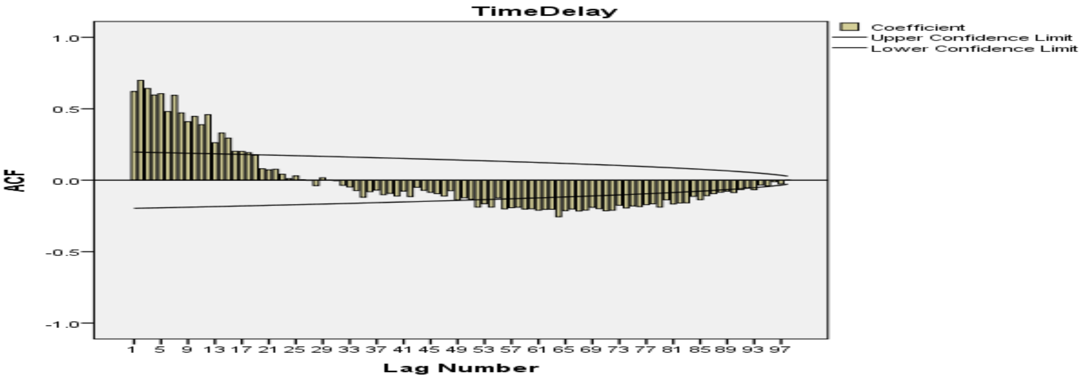

Fig. 2. ACF for Baiduspider

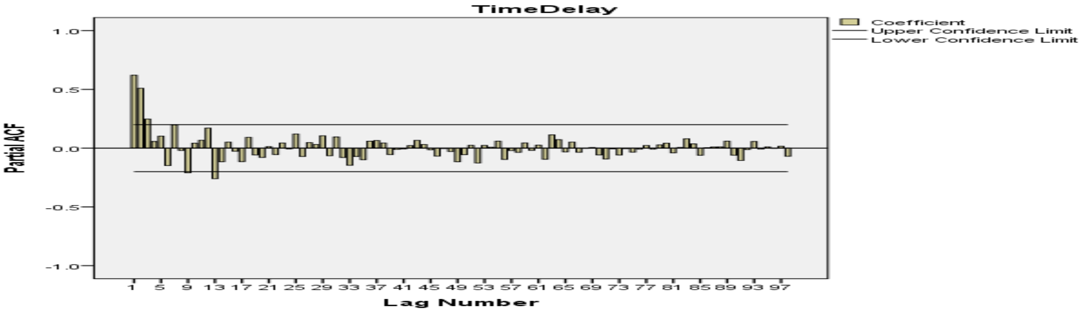

Fig. 3. PACF for Baiduspider

Similarly the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of other crawlers Bingbot are shown in Figure 4 and Figure 5 respectively.

DOI: 10.12948/issn14531305/17.4.2013.03

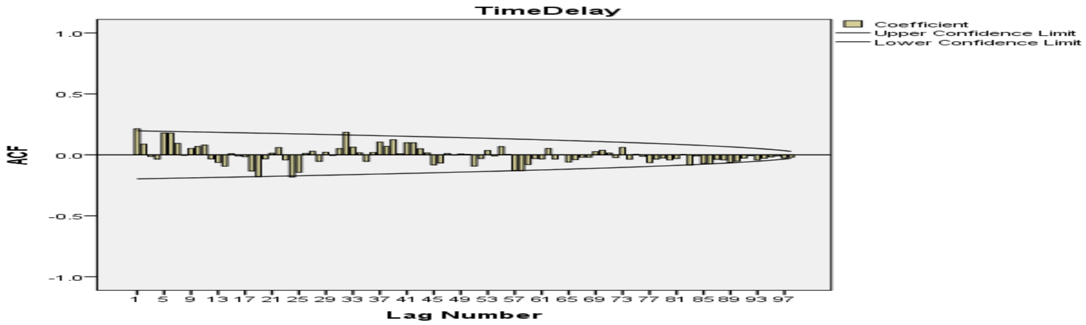

Fig. 4. ACF for Bingbot

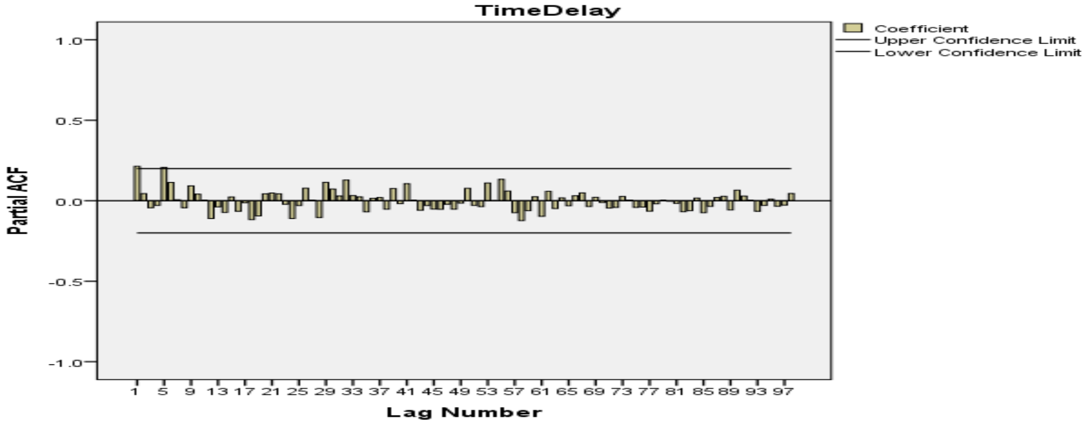

Fig. 5. PACF for Bingbot

The ACF and PACF plots of Feedfetcher-Google, Googlebot and Slurp are shown in Figure 6, Figure 7, Figure 8, Figure 9, Figure 10 and Figure 11 respectively.

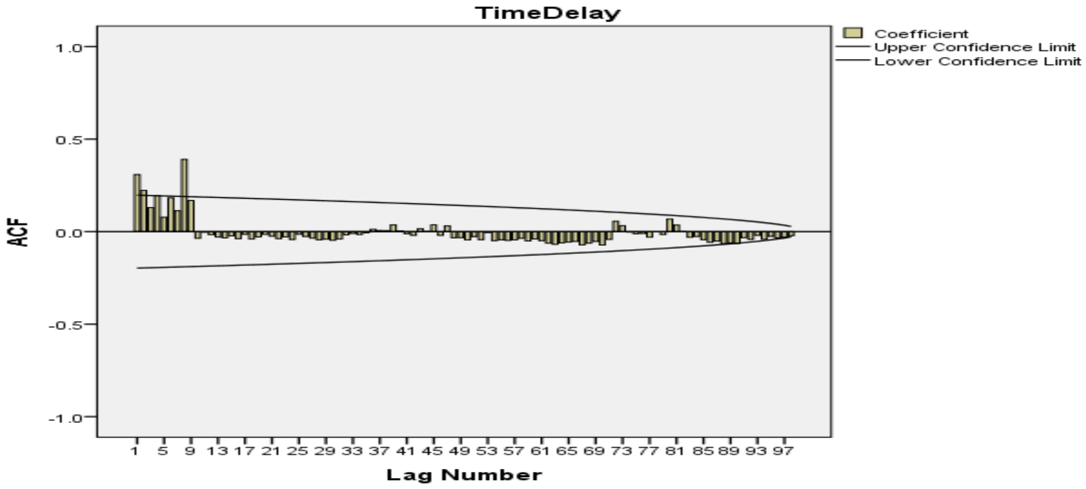

Fig. 6. ACF for Feedfetcher-Google

DOI: 10.12948/issn14531305/17.4.2013.03

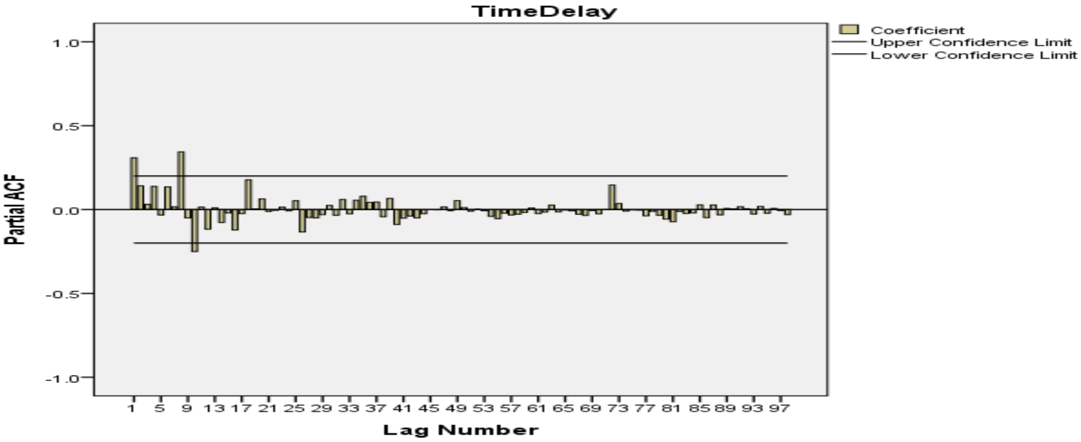

Fig. 7. PACF for Feedfetcher-Google

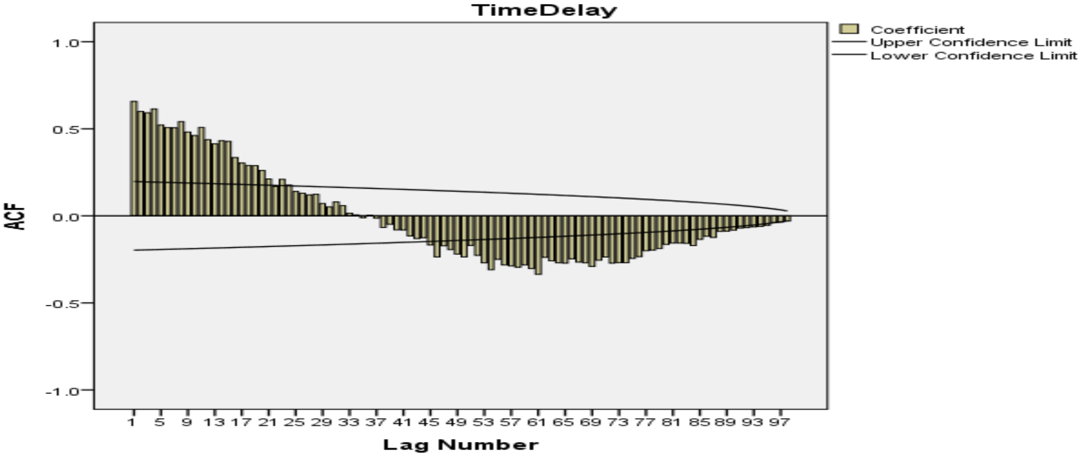

Fig. 8. ACF for Googlebot

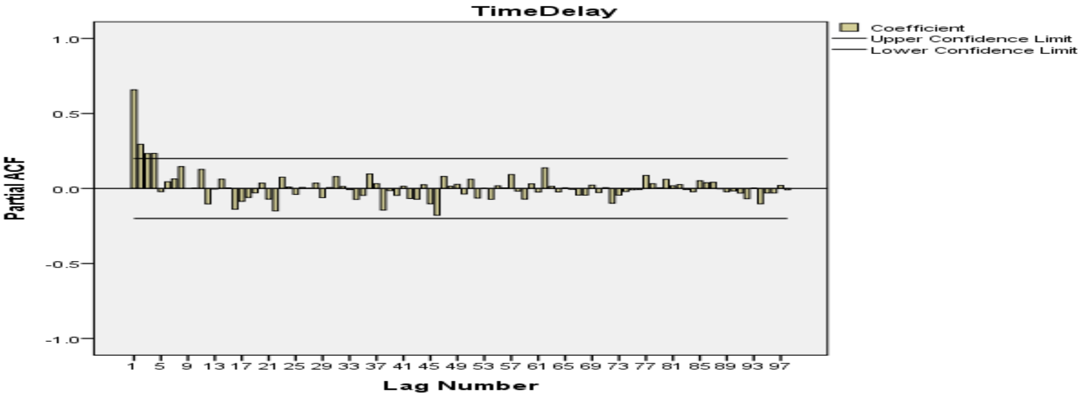

Fig. 9. PACF for Googlebot

DOI: 10.12948/issn14531305/17.4.2013.03

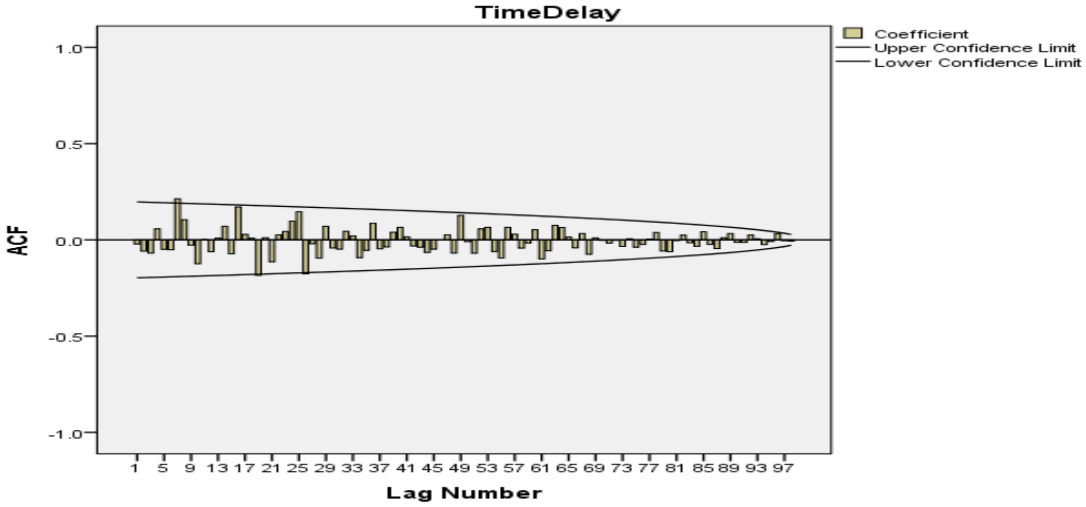

Fig. 10. ACF for Slurp

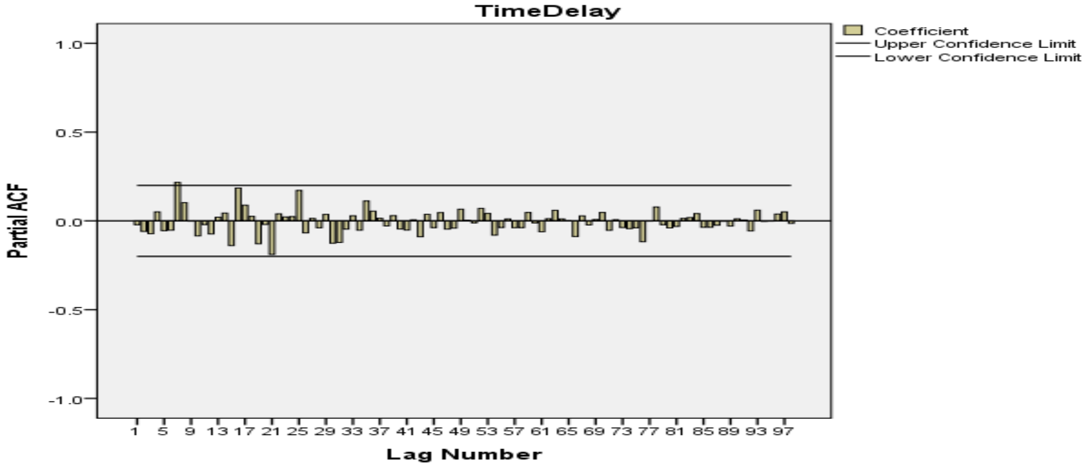

Fig. 11. PACF for Slurp

The ACF and PACF plots revealed that the data series could be modelled using Autoregressive Integrated Moving Average Model ARIMA(1,1,0) the number of autoregressive terms and number of non-seasonal differences as 1 and number of lagged forecast errors to 0.

ARIMA(1,1,0)

ARIMA(1,1,0) is known as the differenced first order autoregressive model. It is represented by the equation

$$Y^{\wedge}(t)=\mu+Y(t-1)+\varphi(Y(t-1)-Y(t-2)) \qquad (3)$$

where μ represents the constant and φ is the autoregressive coefficient.

The observed and forecasted values of time delay between visits of crawlers namely Baiduspider, Bingbot, Feedfetcher-Google, Googlebot and Slurp are shown in Figure 12, Figure 13, Figure 14, Figure 15 and Figure 16 respectively.

DOI: 10.12948/issn14531305/17.4.2013.03

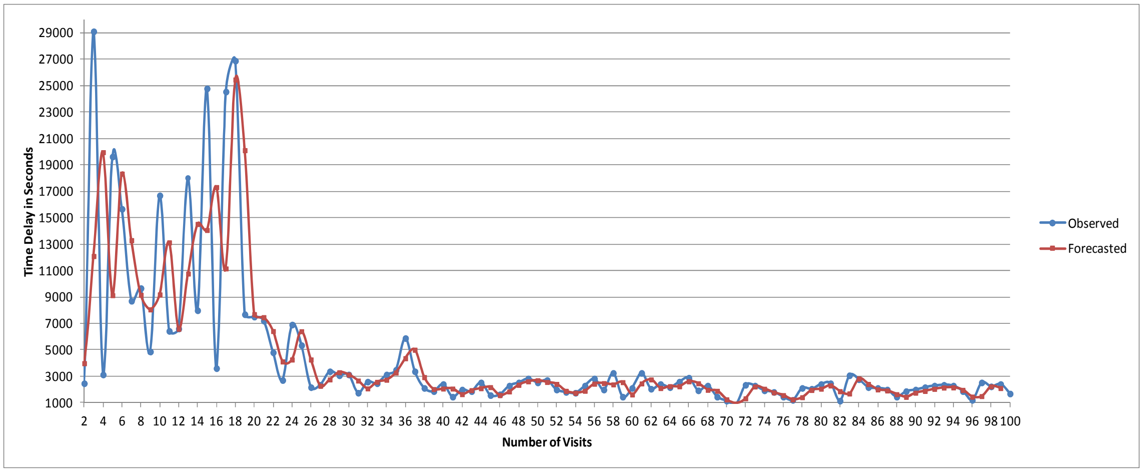

Fig. 12. Observed and forecasted values for Baiduspider

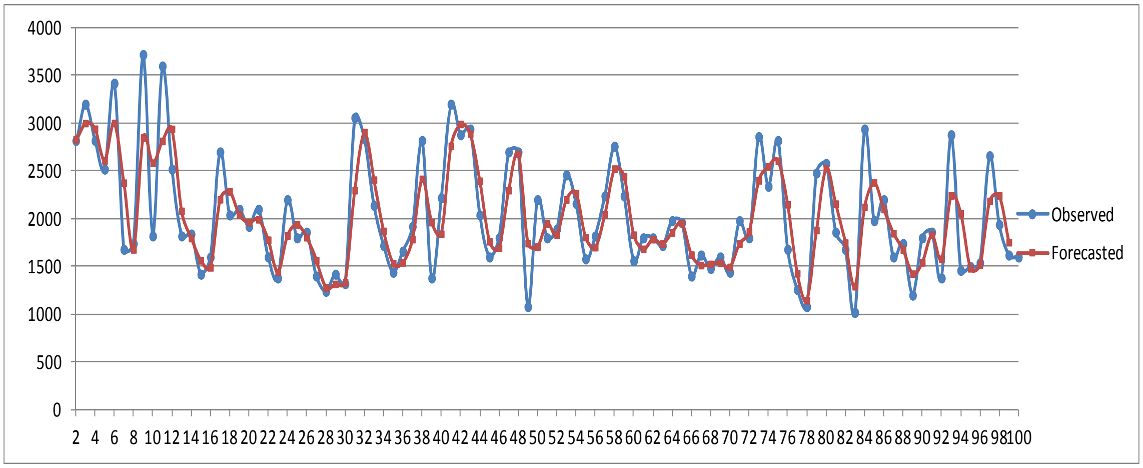

Fig. 13. Observed and forecasted values for Bingbot

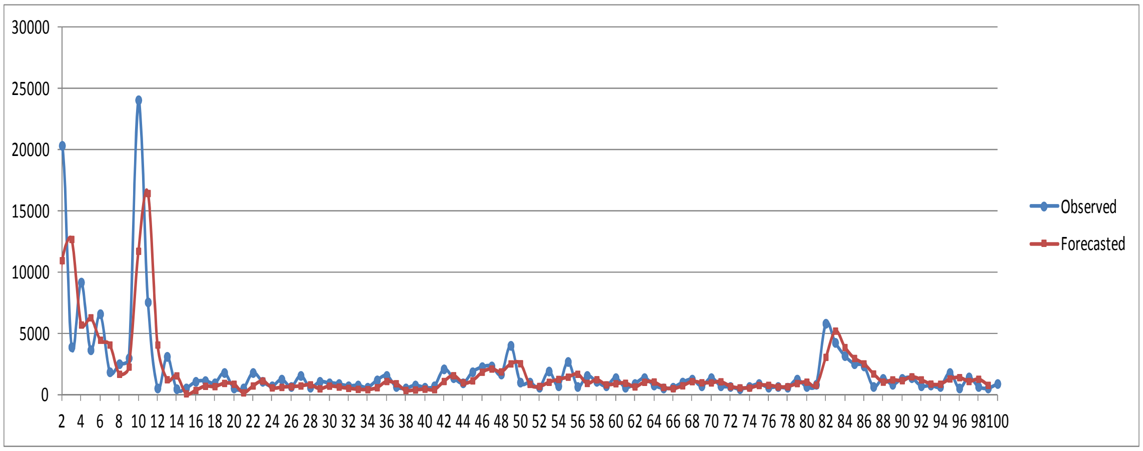

Fig. 14. Observed and forecasted values for Feedfetcher-Google

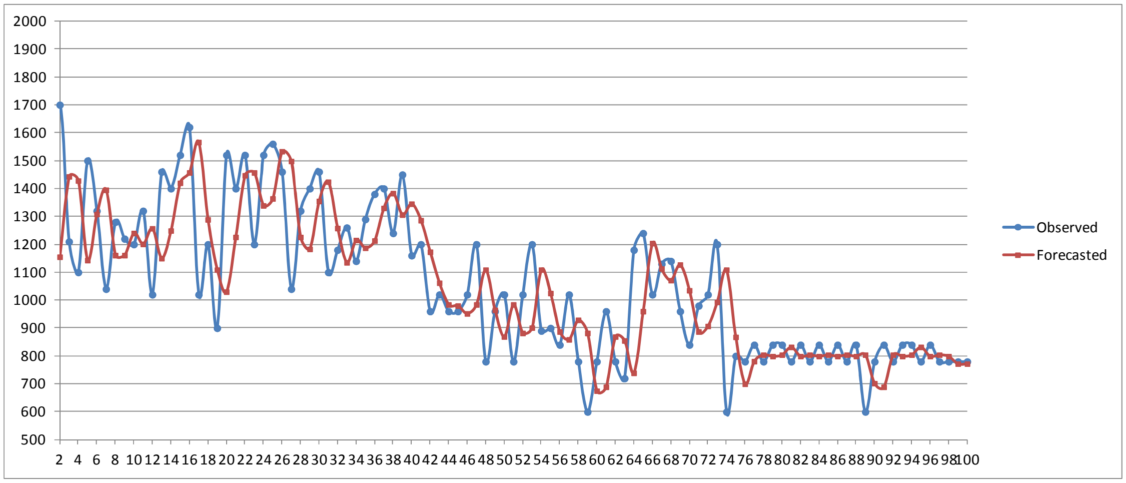

Fig. 15. Observed and forecasted values for Googlebot

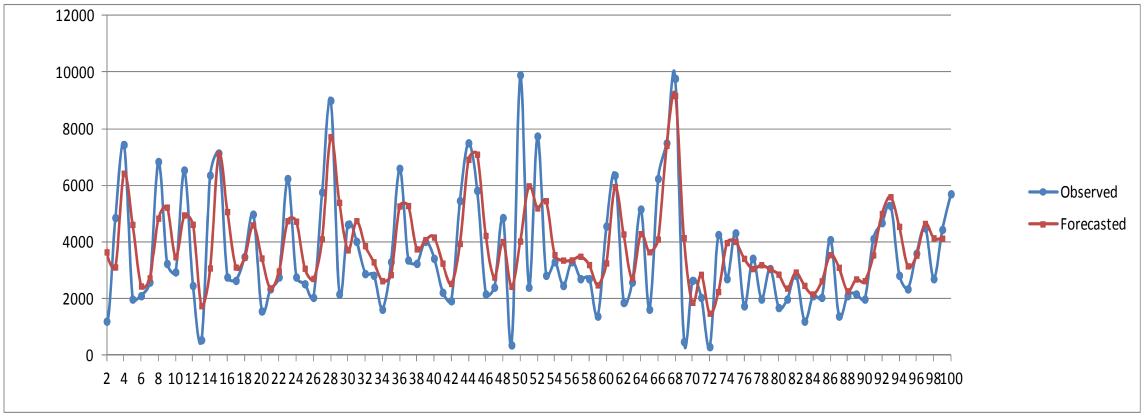

Fig. 16. Observed and forecasted values for Slurp

4 Conclusion

The results revealed that Autoregressive Integrated Moving Average, ARIMA(1,1,0) model suits well for predicting the time delay between visits of search engine crawlers like Baiduspider, Bingbot, Feedfetcher-Google, Googlebot and Slurp. The Autocorrelation Function (ACF) and Partial Autocorrelation Function suggested to opt for ARIMA(1,1,0) model. The crawlers like Baiduspider, Bingbot and Feedfetcher-Google showed more accuracy with this model than Googlebot and Slurp. This forecasting is helpful to calculate the server load and traffic. This work can be extended to find the time delay between visits of crawlers on hourly basis to identify the crawlers visiting the web site during peak hours. The visits of such crawlers can be regulated and assigned to off hours so that the server load could be minimized.

Acknowledgement

This research work is funded by Kerala State Council for Science Technology and Environment, Thiruvananthapuram, Kerala State, India as per order No. 009/ SRSPS/2011/CSTE.

References

[1] C. Lee Giles, Y. Sun and G. Issac, Council, “Measuring the Web Crawler Ethics,” WWW2010, ACM, 2010, pp. 1101-1102.

[2] Y. Sun, Z. Zhuang and C. L. Giles,” A Large-Scale Study of Robots.txt”, WWW2007, ACM, 2007, pp.1123–1124.

[3] J. Jose, P. S. Lal, “A Forecasting Model for the Pages Crawled by Search Engine Crawlers at a Web Site”, International Journal of Computer Applications(IJCA), Vol 68,Issue 13, 2013, pp.19-24.

[4] M.P. Dikaikos, S. Athena and P. Loizos, ”An Investigation of Web Crawler Be-

DOI: 10.12948/issn14531305/17.4.2013.03

havior: Characterization and Metrics", Computer Communications, Vol 28, 2005, pp.880-897.

[6] S. Brin and L. Page, The Anatomy of a Large Scale Hypertextual Web Search Engine, In Proceedings of the 7th International WWW Conference, Elsevier Science, New York, 1998.

[7] D. Sullivan, "Webspin: Newsletter", http://contentmarketingpedia.com/Marketing-Library/Search/industryNewsSeptA1.pdf

[8] L. Vaughan and M. Thelwal, "Search Engine Coverage Bias: Evidence and Possible causes", Information Processing and Management, Vol 40, pp. 693-707.

[9] F. Schwenke and M. Weideman, "The Influence that JavaScript has on the visibility of a web site to search engines – a pilot study", Informatics & Design Papers and Reports, Vol 11, pp. 1-10.

[10]C. L. Giles, Y. Sun and I.G. Council, "Measuring the Web Crawler Ethics," WWW2010, ACM, 2010, pp. 1101-1102.

[11]J. Jose, P. S. Lal, Analysis of the Temporal Behavior of Search Engine Crawlers at Web sites, COMPUSOFT,Vol. 2, Issue 6, 2013, pp.136-142.

[12] J. Jose, P. S. Lal, Differences in Time Delay between Search Engine Crawlers at Web Sites, International Journal of Software and Web Sciences, Vol 2, Issue 5, 2013, pp.112-117.

[13]D. Mican and D. Sitar-Taut," Preprocessing and Content/ Navigational Pages Identification as Premises for an Extended Web Usage Mining Model Development", Informatica Economica, 2009,Vol. 13, Issue 4, pp.168-179.

[14]A.H.M. Wahab, H.N.M. Mohd, F.H. and M.F.M. Mohsin," Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm", World Academy of Science, Engineering and Technology, 2008, pp.190-197.

[15]M. Spiliopoulou, "Web Usage Mining for Web Site Evaluation", Communications of the ACM, 2000.Vol.43, Issue 8, pp.127-134.

[16] S. Jaggia, "Forecasting with ARMA Models", CS-BIGS, 2010, Vol. 4, Issue 1, pp. 59-65.

[17] G. Box and G. Jenkins, "Time-Series Analysis: Forecasting and Control", second edition. San Francisco, CA: Holden Day, 1984.

[18] G. Weisang and Y. Awazu, "Vagaries of the Euro: an Introduction to ARIMA Modeling", CS-BIGS, Vol. 2, Issue, pp.45.

[19]W.S. Wei, "Time Series Analysis: Univariate and Multivariate Methods", Pearson Education, 2006.

[20] P.J. Brockwell and R.A. Davis, "Time Series: Theory and Methods", Springer, 1991.

[21] G.P. Zhang, "Time Series Forecasting using a Hybrid ARIMA and Neural Network Model", Neurocomputing, Issue 50, pp. 159-175, 2003.

[22] S. Makridakis and M. Hibon," Arma Models And The Box Jenkins Methodology", INSEAD, 1984.

[23]S. Lawrence and C. L. Giles, "Accessibility of Information on the Web", Nature, 400:107-109, 1999.

[24] V. Shkapenyuk and T. Suel, "Design and Implementation of a High-Performance Distributed Web Crawler", Proceedings of the 18th International Conference on Data Engineering, IEEE CS Press, pp. 357-368, 2002.

[25] S. Bal and R. Nath, "Filtering the Web Pages that are not modified at Remote Site without Downloading using Mobile Crawlers", Information Technology Journal,Vol.9, Issue 2, pp. 376-380, 2010.

[26] J. Cho, H.G. Molina, and L. Page, "Efficient crawling through URL ordering", In 7th International World Wide Web Conference, May 1998.

[27] B. Pinkerton, "WebCrawler: Finding What People Want", University of Washington, PhD Thesis, 2000.

[28] K. Koht-Arsa, "High Performance Cluster Based Web Spiders", Kasetsart University, Master of Engineering Thesis, 2003.

DOI: 10.12948/issn14531305/17.4.2013.03

[29] O. Papapetrou and G. Samaras, “Minimizing the Network Distance in Distributed Web Crawling”, Springer-Veralag, LNCS, pp. 581-596, 2004.

[30] R. Nath and S. Bal, “A Novel Mobile Crawler System Based on Filtering off Non-Modified Pages for Reducing Load on the Network”, The International Arab Journal of Information Technology”, Vol. 8, Issue 3, pp. 272-279, 2011.

[31] M. Koster, “Robots in the Web: threat or treat?”, www.robotstxt.org/threat-or-treat.html, 1995.

Ms. Jeeva JOSE, Assistant Professor, BPC College, Piravom, India**,** received her Masters Degree in Computer Science from Bharatiar University, Coimbatore in 2000; Master of Philosophy from Bharathidasan University, Trichy in 2006. Currently she is a Ph.D research scholar at School of Computer Sciences, Mahatma Gandhi University, Kottayam, India. She has suceesfully completed a Minor Research Project funded by University Grants Commission, New Delhi, India and has published 15 technical papers. Her research work is funded by Kerala State Council for Science, Technology and Environment, Thiruvananthapuram, Kerala, India.

Dr. P. Sojan LAL, Professor, In charge of Academic Research, MBITS, Kerala received his BE in Mechanical Engineering from Bangalore University, Karnataka, in 1985; M.Tech in Computer Science from NIT, Warangal, AP and Ph.D from School of Computer Science, Cochin University of Science and Technology, Kerala, India in 2002. He is also a Research Guide with School of Computer Science, Mahatma Gandhi University, Kerala. Dr. Sojan, who has 26 years of experience, has published 56 technical papers and a textbook in *Computer Programming with Numerical Methods*. He has also obtained MBA from Strathclyde Business School, Scotland, UK and is a Fellow of The Institution of Engineers (India) since 2004. He is a member of ISTE, ASME, CSI and Engineering Council (UK).

DOI: 10.12948/issn14531305/17.4.2013.03