

# Time Based Web Usage Mining Using Ant Colony Algorithm

<sup>1</sup>Er. Reena, <sup>2</sup>Er. Jyoti Arora

<sup>1</sup>M.Tech Scholar, <sup>2</sup>Assistant Professor, Desh Bhagat University, Gobindgarh Mandi, Punjab, India.

---

**Abstract:** As the web is growing rapidly, the various search engines play an inefficient role in retrieving the data from the internet. When the user requests for data, the search engines provide a huge amount of links related to data directly or indirectly as a search result. Thus it becomes impossible for the user to go through every link provided as a result to the user to retrieve the data required at the moment. To minimize this problem, ranking web pages is one of the most basic and preferable algorithms to sort the data in the manner that the maximum required data or efficient data will be shown in minimum links or WebPages. In this algorithm, the different web pages are assigned a rank as per their usability and user interest and time taken by the server to provide data to the client on the client side. In the presented work, I am working on the same issue i.e. webpage ranking on the basis of the page time taken to execute on the client side and posted date. The complete work is inspired by the behavior of the real ant colonies. As per the artificial ant visits the webpage, the rank is assigned to the different WebPages as per their interest calculated through their pheromone, time consumed to cover the whole path.

**Keywords:** Web mining, Ant colony, Page rank.

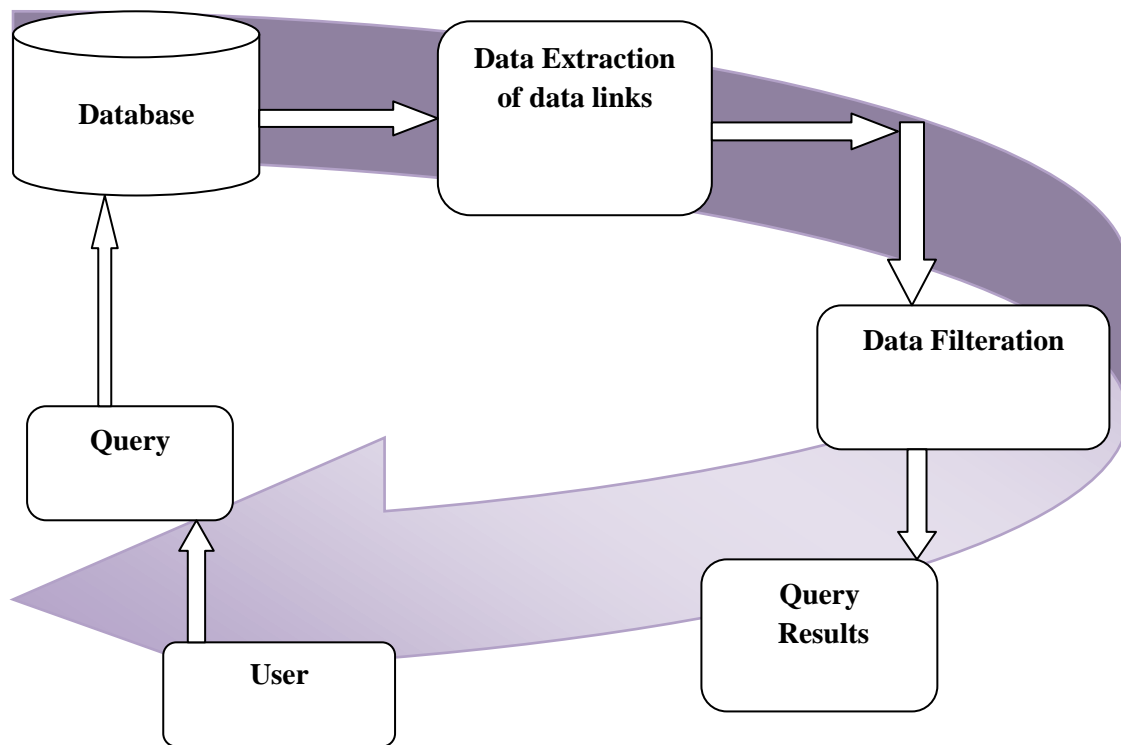
---

## I. INTRODUCTION

Web Mining is that area of Data Mining which deals with the extraction of interesting knowledge from the World Wide Web. Web mining is the process of extraction of data as per the user interest in the manner to provide the maximum useful and efficient data to the users next time. In web mining, the topic web usage mining (WUM) is one of the most interesting topics for the researchers to research in the manner to provide maximum useful and efficient data. In the presented work, we are working on the same topic in the manner to retrieve the maximum data in minimum links to visit so that the user will satisfy as per their requirement and can get maximum data. In the time based system used to assign the rank to different pages is based on the real ant colonies (ACO). In this algorithm, when the virtual ant (user) visits the web page, a hit is logged into the database in reference to the webpage and time is saved that is used by the server to load the webpage at the client side. Since in this algorithm, the key time constraint that is measured by the time consumed to load at the client side is a major key to assign the rank of the different webpages.

From the past surveys, it is clear that the satisfaction level of the users for search results from the different search engines reduces day by day. One of the most famous search engines from which every person of the world is familiar is Google search engine. In the algorithm of search engine Google, there are more than 40 factors based on which the different WebPages are ranked or sorted as per search result. One of the basic factors is page rank in which Google used to allot the ranks to different pages based on inbound and outbound link due to which the WebPages ranking system is not perfectly work and the unnecessary or irrelevant data is displayed as a search result many times. To overcome this, the time based algorithm is introduced in this research paper.

In Fig.1 shows the architecture of the Time based Web Usage Mining using ACO. The three main phases of the system can be discussed here: In this the system preprocesses the raw data, as per the requested query and initially selects the different links to display the various results as per the query. Now, the different results are examined and the different relevant links are discovered as per the request.



**Fig.1 Architecture of the Time based WUM using ACO**

The whole process is completed under the second phase of the process. In third phase, the search result is arranged in the matter to provide the exact, maximum and latest information can be available in minimum data links.

## II. RELATED WORK

The idea of formulating a link analysis problem as an Eigen value problem was probably first suggested in 1976 by Gabriel Pinski and Francis Narin, who worked on scientometrics ranking scientific journals. PageRank was developed at Stanford University by Larry Page and Sergey Brin in 1996 as part of a research project about a new kind of search engine. Sergey Brin had the idea that information on the web could be ordered in a hierarchy by "link popularity": a page is ranked higher as there are more links to it. The first paper about the project, describing PageRank and the initial prototype of the Google search engine, was published in 1998. Shortly after, Page and Brin founded Google Inc., the company behind the Google search engine.

The PageRank transferred from a given page to the targets of its outbound links upon the next iteration is divided equally among all outbound links. If the only links in the system were from pages **B**, **C**, and **D** to **A**, each link would transfer 0.25 PageRank to **A** upon the next iteration, for a total of 0.75.

$$PR(A) = PR(B) + PR(C) + PR(D).$$

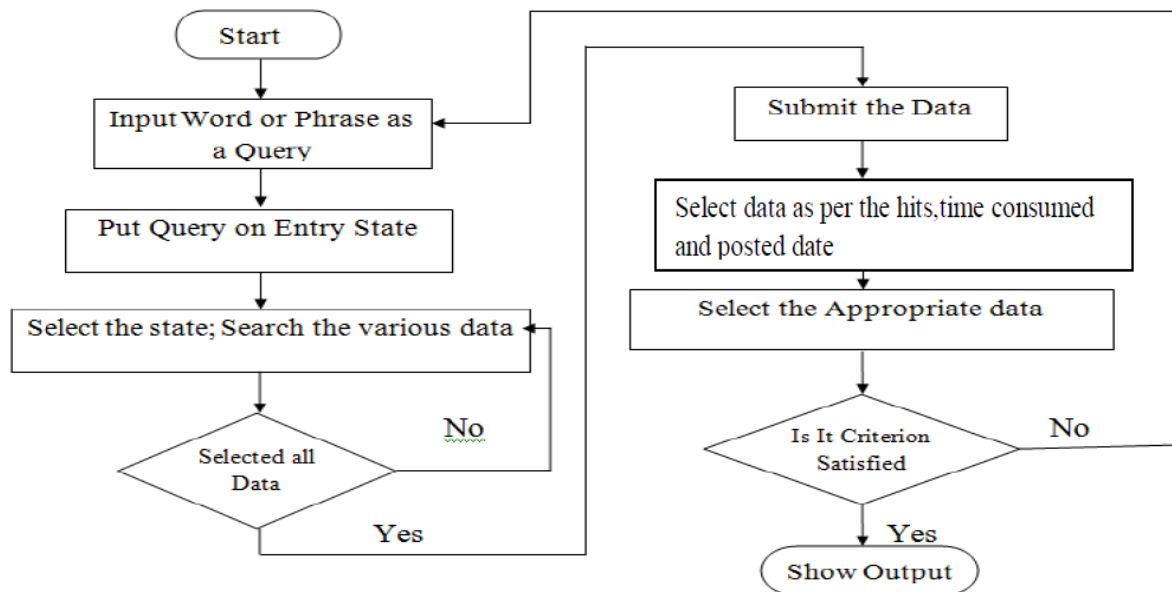
If the only links in the system were from pages **B**, **C**, and **D** to **A**, each link would transfer 0.25 PageRank to **A** upon the next iteration, for a total of 0.75.

Another work done by the G.Anuradha[8] and her co.team members who suggested the algorithm ANTRANK for the ranking of web pages. They suggested interesting suggestion of using the no. of hits on the link as a probability value of the link. Maximum no. of user visited the web page, the page probability will be maximum for the link and in the manner the links are shown as a result of maximum probability of the page and the this manner the whole search engine will work.

Also another idea is suggested by the richa gupta[7] who suggested that the pages can be classified on the basis of the user session which also one of the good idea to maintain the probability of the different pages.

### III. METHODOLOGY

In the proposed system, it works on the three phases defined by the web mining as pre-processing, pattern discovery and pattern analysis. Initially the query requested by the user is forwarded to the database in which the different data are selected as per the result of the requested query. After the selection of different links, the different data discovered as the result of search is analysed. In this, the different data are compared on the basis of their hits, time consumed by server to load page at client side and the posted date of the data on server. The data having maximum hits or is latest update on the database and the minimum time consumed to load over the client side is selected and shown as a result of the search.



### IV. ALGORITHM

The algorithm comprise of following steps:

1. Initialization of values.
2. Set  $S :=$  set of selected web pages;  $N :=$  no. of web pages selected,  $INPUT[] :=$  Input data which is to be search;  $S_f[] :=$  selected web pages after filtering;  $W_n :=$  all web pages within the database;  $W_a :=$  Web page name;  $L_A =$  All logs;  $L_f :=$  Filtered log;  $Q :=$  requested query
3. Set  $S=0$ ; Select  $W_n, Q=INPUT[]$ .
4. While  $INPUT []$  matches  $W_n$  || index, select as the search result until  $W_a=W_n$ .
5. Select  $S_f[] \in \{0,1,2,\dots,n\}$  where  $S_f[] =$  selected Web pages from the  $W_a$  after the analysis of the selected data.
6. For all  $S_f[]$ , arrange them as per the user hits ||  $Creation\_Date$  ||  $avg(time\_taken)$  in ascending order
7. Do steps 5, 6 and 7 until the value of  $S_f[]=n$ . When the no. of values of  $S_f[]$  i.e. the different links are opened in new page as a search result as per their hits or date, the user need to open the link as per there requirement. At the moment user can open the multiple links. When the user open the link, then as per there feedback, an increment will be added to either the like or dislike attribute of the table.
8. When the users open any link, add increment to the webpage as per feedback, Increment +1 to  $W_n$  {likes/dislikes}. And hence based on the algorithm, the performance and accuracy of the website will be improved.

## V. RESULTS

In the first screenshot, the database is shown which contains all the collected data with their proper attribute in the manner to select the data of good quality whenever the user requests.

Content ID	Content Heading	Content Image	Status	Posted At	Posted By	Likes/Dislike	Hits	Details	View Images	Action
35	Ant Rank		true	9/2/2014 3:15:36 AM	reena	0/0	0	<a href="#">View</a>	<a href="#">View</a>	<a href="#">Edit</a> <a href="#">Delete</a>
34	Architecture Web Usage mining		true	9/2/2014 1:30:36 AM	reena	/		<a href="#">View</a>	<a href="#">View</a>	<a href="#">Edit</a> <a href="#">Delete</a>
33	area of Web mining		true	9/2/2014 1:29:31 AM	reena	0/0	0	<a href="#">View</a>	<a href="#">View</a>	<a href="#">Edit</a> <a href="#">Delete</a>
32	Definition Web mining		true	9/2/2014 1:28:47 AM	reena	0/0	0	<a href="#">View</a>	<a href="#">View</a>	<a href="#">Edit</a> <a href="#">Delete</a>

**Fig.2 Screenshot Database Collection**

In screenshot 3, the different data can be shown in the manner the maximum rank of the page among all the research. In this, the data before any condition upon which the searching can be done is not concerned and the query is run simply on the system.

In this, the result shown for the search is before the implementation of the algorithm. The different links are shown as per the search result for the query which displays number of results with inefficient and poor quality of data due to which the search becomes useless and unsuitable due to which an improvement is required in the manner to improve the results of the search engine.

Searched For : Web usage mining  
 About 22 results found in (25.60 Seconds)

**Wikipedia on Web mining**  
 It is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining. Web usage mining: Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out ...[Read More](#)  
 Views: 8

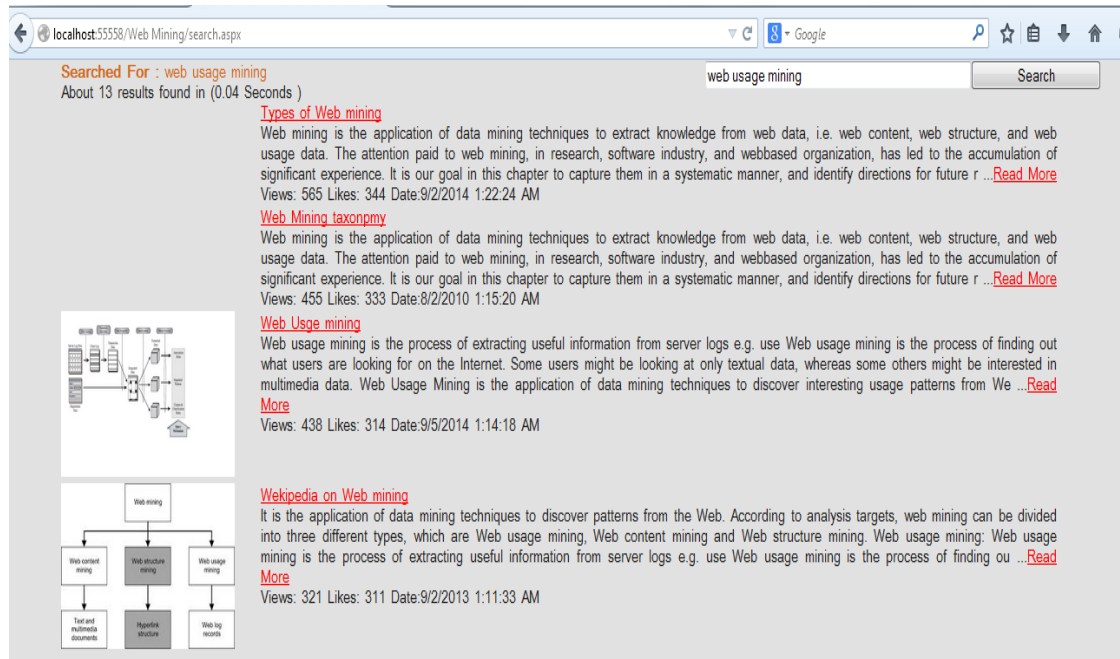
**Web Mining using Artificial Ant Colonies**  
 Web Mining using Artificial Ant Colonies Ant Colony Optimization (ACO) is a technique for optimization that was introduced in nearly 1990s. The inspiration of ACO algorithms are the behavior of real ant colonies. The complex social behaviors of ants have been much studied by Science, and computer scientists are now finding that these behavior patterns can provide models for solving difficult co ...[Read More](#)  
 Views: 0

**Ant Colony Algorithms**  
 The pioneer in understand how ant foraging works was Jean-Louis Deneubourg et al. [Jean-Louis Deneubourg et al., 1989]. They proposed that the reason ants are seen creating "highways" to and from their food is because of a chemical pheromone. The way in which this type of intelligent behavior emerges from a group of otherwise unintelligent entities has been labeled as "swarm intelligence". Each an ...[Read More](#)  
 Views: 0

**Wikipedia on Ant colony optimization**  
 With substantial decline in data storage cost, rapid advancement in computer networks, technology advances in data acquisition, improvement in computer performance and explosive growth in the generation of electronic information, huge amount of data are being collected and stored in databases. The amount of data stored in databases continues to grow fast. This large amount of stored data

**Fig. 3 Screenshot of the search result before implementation**

In screenshot 4, the requested query data is searched over the system and the no. of links are displayed after implementation of the algorithm within the system.



**Fig. 4. Screenshot of the search result after implementation of algorithm**

## VI. CONCLUSION

In this work, the goal of this algorithm is to assign rank for web pages based on Users Interest (Ant) and the time consumed to complete the whole cycle. The time consumed by artificial ant to return back to home with food. The Proposed algorithm is based on the behavior of real ant colonies. The concept has been utilized for Web mining. The main focus is to show the performance of the system in the manner that the WebPages are assigned by the ranking. In this, individual page in initially ranked with 0. Based on the user interest, time taken to load webpage on the client side and posted date of the data on the server, the ranking of the web page in increased or decreased within the system and hence it is as per the ranking, shown to the user within the search results.

Further in this, I am introducing an improved system to improve the search result. In this, Concept has shown improved results after elimination of the user misbehaving parameters in the search algorithm.

## VII. FUTURE SCOPE

It is an important issue for the further study to implement the proposed scheme on the search engine in the manner to improve the results. The proposed work focused on the elimination of poor quality of data in the manner that the unrequired data would be neglected or removed from the results shown to the user. In future it is very interesting research to improve the search result with proposed algorithm. Moreover this research is useful in saving the resources but still there is more research is required to be done in the manner to use the logic in case of large and huge domains containing web data in millions or in which, on the daily basis, the webpages are posted or updated.

## VIII. ACKNOWLEDGEMENT

This project has been a dedicated effort towards development of a highly autonomous artificial intelligence, which primarily would not have been possible at the first place without the apt guidance of the respected Asst. Prof. Er. Jyoti Arora . Her motivation and encouragement during the project made me go further than expected in the development process and the redaction of this current report. I would also like to thank Asst. Prof. Er. Amandeep Kaur who the first explained to me the concept of Ants Algorithm and provided me with a relevant example of implementation.

**REFERENCES**

- [1] Xing, W., Ghorbani, A. 2004. Weighted PageRank Algorithm. Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04), IEEE.
- [2] Identifying User Behavior by Analyzing Web Server Access Log File, K. R. Suneetha, and Dr. R. Krishnamoorthi, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009.
- [3] .Keong, B.V., Anthony, P. 2011. PageRank: A Modified Random Surfer Model. International Conference on IT in Asia (CITA), IEEE, pp. 1-6.
- [4] Sara Setyash, Ali Harounabadi, Amir Masoud Rahmani, "Presentation of an Extended Version of the PageRank Algorithm to Rank Web Pages Inspired by Ant Colony Algorithm", International Journal of Computer Applications (0975 – 8887) , January 2014.
- [5] Sara Setyash, Ali Harounabadi, Amir Masoud Rahmani, "Presentation of an Extended Version of the PageRank Algorithm to Rank Web Pages Inspired by Ant Colony Algorithm", International Journal of Computer Applications (0975 – 8887) , January 2014.
- [6] Bhaiyalal Birla, Sachin Patel, "An Implementation on Web Log Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014, ISSN: 2277 128X.
- [7] Richa Gupta, "Web Mining using Artificial Ant Colonies: A Survey", International Journal of Computer Trends and Technology (IJCTT) – volume 10 number 1 – Apr 2014.
- [8] G. Anuradha, G. Lavanya Devi and M.S Prasad Babu3, "ANTRANK: An ant colony algorithm for ranking web pages", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3 Issue 2, March – April 2014
- [9] Esra Saraç and Selma AyGe Özel, "An Ant Colony Optimization Based Feature Selection for Web Page Classification", Scientific World Journal Volume 2014, Article ID 64926
- [10] Ajith Abraham, "Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming".
- [11] Nicholas Holden, "Web Page Classification with an Ant Colony Algorithm".
- [12] M.Krishna Kumar, Dr. S. Poonkuzhali, "Enhancing the Search Results through Web Structure Mining Using Frequent Pattern Analysis and Linear Correlation Method", International Journal of Electronics Communication and Computer Engineering, Volume 4, Issue 3, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209.
- [13] Yezheng Liu, Haifeng Ling, Shanlin Yang, "An artificial ant colony methodology for users pattern navigation mining".
- [14] JaideepSrivastava, Robert Cooley, Mukund Deshpande, "Web Usage mining: Discovery and Applications of pattern usage from web data".