ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

IMPLEMENTING THE WEB CRAWLER CONSIDERING VISITING FREQUENCY, WEIGHTAGE AND TTL VALUE TO INVESTIGATE THE PERFORMANCE

1Gaurav Chaudhary, 2Dr. Vijay Laxmi
1Research Scholar, Department of Computer Application, gaurav.gchaudhary@gmail.com
2A.P. Department of Computer Application, deanca@gku.ac.in
University college of Computer Applications Guru Kashi University, Talwandi Sabo (Punjab)

Abstract: Proposed research is considering the frequency of visiting the page. Another parameter for the crawling is weight age of the page i.e. how much time page get visited by the end users. The page which is most visited get high weight age. Moreover the research work would consider the TTL value of the page. TTL value shows what is the expiry time of the page . It is considering the pages that are having the less expiry time. The web crawling is capable to filter the web page list during searching considering visiting frequency, weight age and TTL value. In future research there is requirement to check the security against the malware attacks so that secure and optimized crawling could be performed.

Keywords: Web Crawling, Weight age, TTL value, visiting frequency

[1] INTRODUCTION

A web crawler is a program which systematically navigates the internet indexing WebPages. Tradition web crawlers are suffering from several limitations they are ignoring following key points.

Frequency of visiting page: The most frequently visited web pages must be ranked in order to prioritize during web crawling.

Life span of webpage: Existing researches are not considering lifespan of packets. Time-to-live (TTL) is a value in an Internet Protocol (IP) packet that limits the lifespan. It is defining a date and time at which the content is stale in HTML page headers. There is need to introduce a mechanism of TTL during web crawling so that the pages that are not live should be filtered out.

Interest of user: Time spent on web page shows the interest of user. The maximum time spent by user on web page should be considered during deciding the priority of web page for crawling.

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

Secure browsing: However some pages are highly insecure but they are considered in list during web crawling the pages that are unsecure must be filtered out. There is need to introduce a mechanism that could eliminate the unsecure pages from the list during web crawling. The most famous application of web crawling is Google's Search Engine. Below is a diagram of the internal workings of a typical web crawler:

Working of traditional web Crawler

In order to crawl a website or the web, you first need an entry point. Robots need to know that your website exists so they can come and have a look at it. Back in the days you would have submitted your website to search engines in order to tell them your website was online. Now you can easily build a few links to your website and Voilà you are in the loop!

Once a crawler lands on your website it analyses all your content line by line and follows each of the links you have whether they are internal or external. And so on until it lands on a page with no more links or if it encounters errors like 404, 403, 500, 503.

From a more technical point of view a crawler works with a seed (or list) of URLs. This is passed on to a Fetcher which will retrieve the content of a page. This content is then moved on to a Link extractor which will parse the HTML and extract all the links. These links are sent to both a Store processor which will, as its name says, store them. These URLs will also go through a Page filter which will send all interesting links to a URL-seen module. This module detects if the URL has already been seen or not. If not it gets sent to the Fetcher which will retrieve the content of the page and so on.

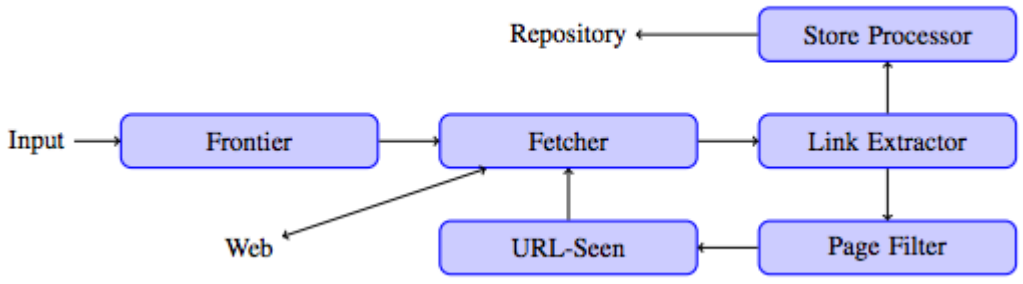


Figure 1 Architecture of a tradition web crawler

[2] LITERATURE REVIEW

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

In 2020, J. Xu, et al.[1] The Application of Web Crawler in City Image Research. Grasping the city image accurately is always a challenge for city manager. The main work of our study is to capture the web pages of a specific city using web crawler technology, process and analyze the captured data to get keywords representing the city. They propose a distributed crawler architecture and collect city-related data based on the architecture. Subsequently, several example specific cities are crawled successfully.

In 2019, V. V. Mahale, .et .al.[2]Advanced web crawler for deep web interface using binary vector page rank. Researcher are gaining more interest in deep web crawling. The issue of visiting the web pages is addressed by deep web, where pages are crawl from the deep website based on the query inputed by the user in the search form. Researcher are gaining more interest in crawling the hidden web. To crawls the pages the crawlers need to be empowered with special feature which will go beyond simply following links, like they should be capable to reveal search forms smartly that are entry points to the deep Web, fill in such forms, & follow certain paths to reach the deep Web pages with proper information. To enrich the crawling They present a unique way of crawling. To increase the performance of crawling the crawler They implemented calculates binary vector & page rank of pages & also return the count keywords which are mined from the URL. Implementing the proposed crawler will help in getting more precise result for a focused crawler with ranking. Experimental analysis is done in java where the performance and accuracy of the crawler is tested. Experimental results on a set of various domains depicts the agility & accuracy of our proposed crawler framework, which effectively retrieves deep-web interfaces from large-scale sites & attains higher collection rates as compare to the state of art crawlers

In 2018, Y. Wang, .et .al.[3] Research on LDA Model Algorithm of News-oriented Web Crawler. With fast development of big data, the data quantities and information types on the webpage are increasing tremendously. Consequently, it is becoming more difficult for users to obtain the valuable and interesting data and information from the webpage. The paper designs and implements a topic-focused crawler which utilizes Nodejs lightweight directional crawler to capture the data with great improvement on the efficiency of page retrievals. Firstly, the design idea and flow of the web crawler project is introduced. Furthermore, on the basis of crawled data

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

jieba package of python is used to achieve text participle. Finally, LDA model algorithm is used to classify keyword texts to reach the purpose of classifying different types of news

In 2018, S. M. Nakashe, .et .al.[4] Smart Approach to Crawl Web Interfaces Using a Two Stage Framework of Crawler. In present scenario, internet is very important part of our life.

In 2017, Z. Shi, .et .al.[5] The Implementation of Crawling News Page Based on Incremental Web Crawler. Web crawler technology is the technology which downloads web pages through the program. This paper implements incremental Python web crawler, uses Scrapy crawler framework, crawls news web pages from mainstream web sites incrementally in real time, and deposits data in the database.

In 2016, M. Kumar, .et .al.[6] Design of a mobile Web crawler for hidden Web. The World Wide Web (WWW) is a diverse source of information. A large part of the Web is hidden behind search forms and is reachable only when a user types in a set of keywords or queries. This part of Web is known as hidden Web or deep Web. The webpages in the hidden Web are not accessible by following hyperlinks and hence are not indexed by the search engine.

In 2015, N. Kumar, .et .al.[7] Framework for Distributed Semantic Web Crawler. Relevant information retrieval from the www mainly depends on the technique and efficiency of a crawler. So crawlers must be capable enough to understand the text and context of a link which they are going to crawl.

In 2015, S. Sharma, .et .al.[8] The anatomy of web crawlers. World Wide Web (www) is the gigantic and richest source of information. In this paper a survey of different architectures of web crawlers along with their comparisons has been carried out that takes into account various important features like scalability, manageability, page refresh policy, politeness policy etc

In 2015, A. Gupta, .et .al.[9] Focused web crawlers and its approaches. Rapid growth of WWW poses unpredictable challenges for the crawlers and search engines. In this work, They propose focused web crawler architecture to expose the underneath secrets of web crawling implementation.

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

In 2015, G. H. Agre, .et .al.[10] Keyword focused web crawler. Users and uses of internet is growing tremendously these days which causing an extreme trouble and efforts at user side to get web pages searched which are as per concern and relevant to user‘s requirement.

In 2013, A. Aghamohammadi, .et .al.[11] A novel defense mechanism against web crawlers intrusion. In this paper, a novel method to identify web crawlers is proposed to prevent unwanted crawler to access websites. This new method suggests Five-factor identification process to detect unwanted crawlers.

In 2012, K. S. Shetty, .et .al.[12] Symbolic verification of web crawler functionality and its properties. The aim of this paper is to model check the crawling process and crawler properties using a symbolic model checker tool called NuSMV. The basic operation of a hypertext crawler and the crawler properties has been modeled in terms of CTL specification and it is observed that the system takes care of all the constraints by satisfying all the specifications.

In 2012, W. Ma, .et .al.[13] Advanced deep web crawler based on Dom. Due to the fact that large amount of data today can only be stored in deep web. In view of the work done by others on deep web crawlers, it is extinct that no perfect, or even complete crawlers for deep web data has been made.

In 2012, W. Guo, .et .al.[14] A web crawler detection algorithm based on web page member list. The experiment shows that the new algorithm can detect the unknown crawlers and unfriendly crawlers who do not obey the Standard for Robot Exclusion.

In 2011, O. Jalilian, .et .al.[15] A new fuzzy-based method to weigh the related concepts in semantic focused web crawlers. In this paper a new method to weigh the concepts related to topic is suggested to be used as a main component in the architecture of semantic crawler to compute relevance of web page with the topic. The concepts related to the topic are retrieved by ontology graph and their weights are computed by a proposed fuzzy inference system. The results show that the proposed approach presents better precision rate compared with breadth-first and best-first search.

In 2011, F. Ahmadi-Abkenari, .et .al.[16] A clickstream-based web page significance ranking metric for web crawlers. This paper proposes an analysis on clickstream data in order to discover the popularity of Web pages in crawl frontier through proposing the metric itself and presenting the experimental results on ranking the UTM Web pages based on the proposed discussed metric

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

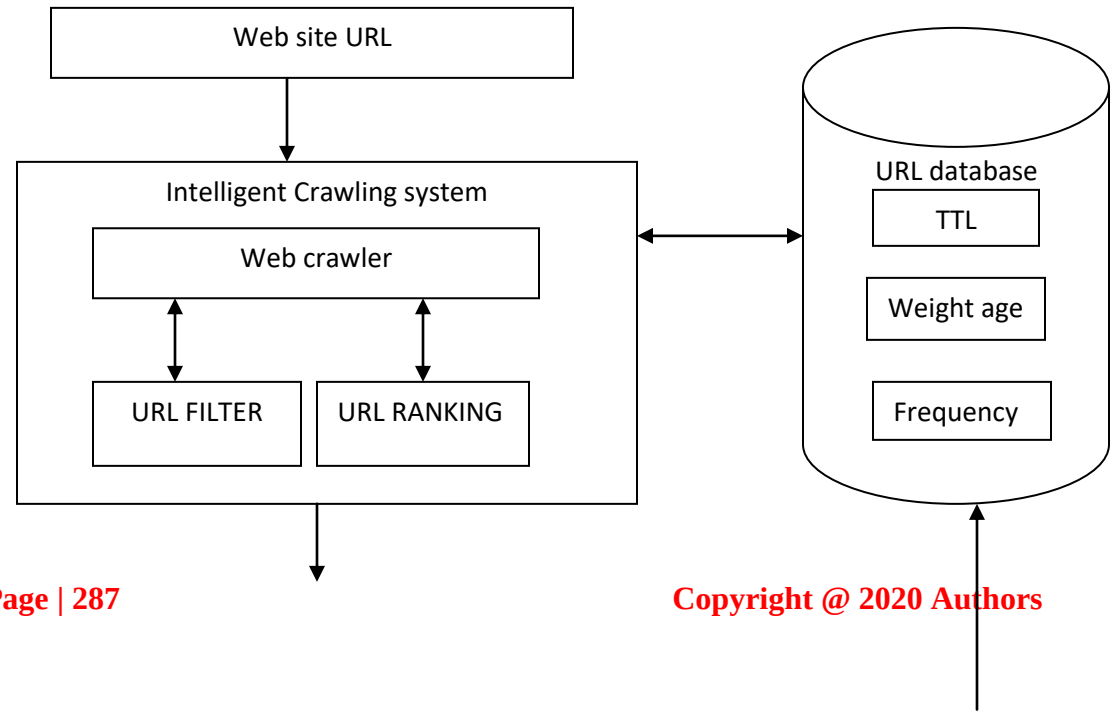
In 2010, M. S. Zhao, .et .al.[17] An intelligent topic web crawler based on DTB. This paper proposes an intelligent topic Web crawler based on DTB (dynamic topic base), which through studying on Web crawlers which filter URLs based on different methods. Experimental results show that the proposed Web crawler can fetch more topic relevant Web pages by crawling less Web space and in less time.

[3] PROBLEM STATEMENT

However there have been several researches in field of web crawling but still there is need of updates. Many researches focused on frequency of visiting web page but ignored the weight age and TTL. Some research considered weight age but they are not considering security. The researchers that consider security are not considering the performance during web crawling. Proposed system is suppose to resolve the issue of web page filtering considering TTL and issue of ranking considering weight age, frequency of web page visiting.

[4] PROPOSED WORK

In proposed system the website URL is passed to intelligent crawling system where web crawler perform the web crawling and URL filter is checking for the TTL , weight age and frequency of web page visiting. List of URL is ranked as per weight age and frequency of page visiting. The TTL represent the status of page whether is it is available or not at the time of browsing.



ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

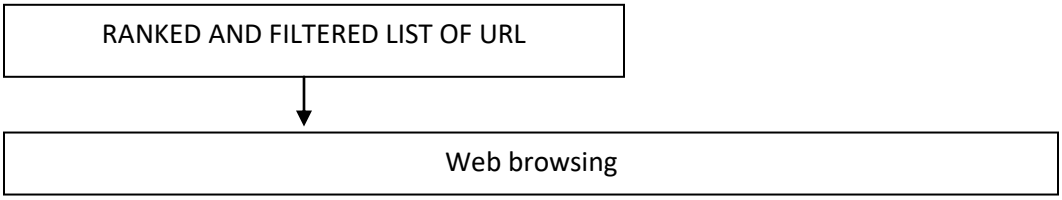


Fig 2 Process flow of proposed work

Proposed algorithm has considered the frequency of visiting the page. Another parameter for the crawling is weight age of the page represents the time page visited by users. Most visited page is getting high weight age so it is provided rank as per weight age . Proposed work has considered the TTL value of the page. TTL value shows the expiry time of the page to filter it. It is considering the pages that are having the less expiry time. The web crawling is capable to filter the web page list during searching considering visiting frequency, weight age and TTL value. The intelligent web crawling system is consisting of URL filter and URL ranking modules to eliminate the expired Url and set their rank. Rank are depending on the frequency of visiting and weight age. Filtering is performed as per TTL value. The database is getting updated as per user action.

[5] RESULTS

The proposed work need the database considering the URL, Frequency of web base visiting, weightage, and TTL value. This database would play significant role in filtering. Table has been shown below with schema that is used by proposed web crawling system.

Field	**Datatyple**
Sno	Integer
Url	Varchar
Frequency	Integer
Weightage	Float
TTL	Integer

ALGORITHM FOR DATABASE UPDATION AS USER VISIT URL

1. Get the url and check in table
2. If url not exist
    a. Make new entry for URL and
    b. Set frequency=1

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

c. Set weightage=time spent on url

d. TTL= 1 if available, 0 if not available

3. If url already exists

a. Update frequency = Frequency +1

b. Update weightage= weightage + time spent

c. TTL= TTL + (1 if available , 0 if not available)

Following dataset has been generated after implementing above algorithm

Table 1 URL dataset

SNO	URL	FREQUENCY	WEIGHTAGE	TTL
1	http://trdinfotech.com/page1	34	21	32
2	http://trdinfotech.com/page2	20	22	17
3	http://trdinfotech.com/page3	21	34	13
4	http://trdinfotech.com/page4	6	66	4
5	http://trdinfotech.com/page5	7	55	4
6	http://trdinfotech.com/page6	9	44	5
7	http://trdinfotech.com/page7	50	33	29
8	http://trdinfotech.com/page8	70	55	12
9	http://trdinfotech.com/page9	21	32	13
10	http://trdinfotech.com/page10	43	65	34
11	http://trdinfotech.com/page11	100	34	80
12	http://trdinfotech.com/page12	50	23	40
13	http://trdinfotech.com/page13	50	44	43
14	http://trdinfotech.com/page14	70	32	56
15	http://trdinfotech.com/page15	4	44	3
16	http://trdinfotech.com/page16	6	32	3
17	http://trdinfotech.com/page17	2	54	1
18	http://trdinfotech.com/page18	1	76	0
19	http://trdinfotech.com/page19	8	77	0
20	http://trdinfotech.com/page20	9	45	0
21	http://trdinfotech.com/page21	40	76	0
22	http://trdinfotech.com/page22	100	75	0
23	http://trdinfotech.com/page23	70	12	35

The line graph considering above listing is plotted below

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

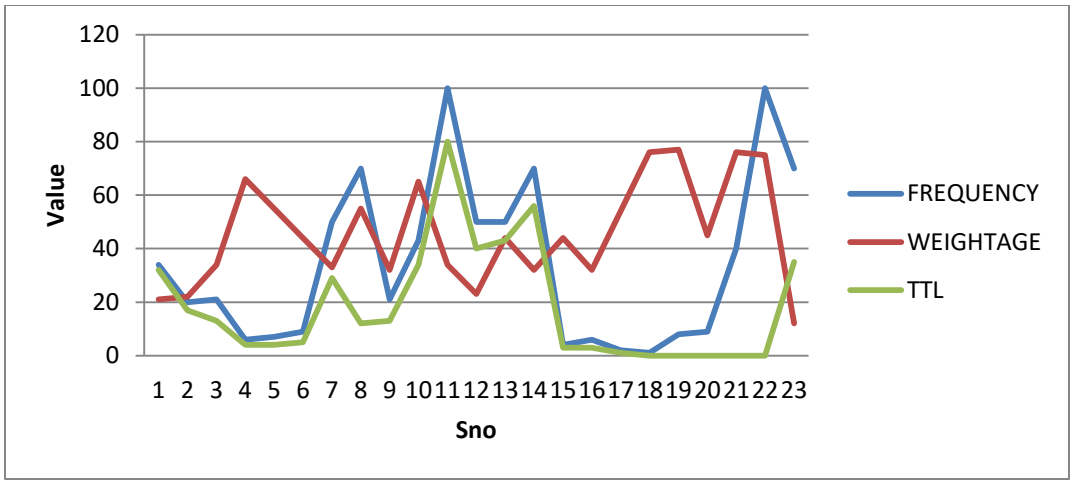


Fig 3 Graphical representation of line chart for dataset of visited sites

SQL QUERY TO CHECK THE RANK CONSIDERING FREQUENCY ONLY

SELECT * FROM CRAWLER ORDER BY FREQUENCY DESC

Table 2 Rank considering frequency

SNO	URL	FREQUENCY	WEIGHTAGE	TTL
11	http://trdinfotech.com/page11	100	34	80
22	http://trdinfotech.com/page22	100	75	0
8	http://trdinfotech.com/page8	70	55	12
14	http://trdinfotech.com/page14	70	32	56
23	http://trdinfotech.com/page23	70	12	35
7	http://trdinfotech.com/page7	50	33	29
12	http://trdinfotech.com/page12	50	23	40
13	http://trdinfotech.com/page13	50	44	43
10	http://trdinfotech.com/page10	43	65	34
21	http://trdinfotech.com/page21	40	76	0
1	http://trdinfotech.com/page1	34	21	32
3	http://trdinfotech.com/page3	21	34	13
9	http://trdinfotech.com/page9	21	32	13
2	http://trdinfotech.com/page2	20	22	17
6	http://trdinfotech.com/page6	9	44	5
20	http://trdinfotech.com/page20	9	45	0
19	http://trdinfotech.com/page19	8	77	0
5	http://trdinfotech.com/page5	7	55	4
4	http://trdinfotech.com/page4	6	66	4
16	http://trdinfotech.com/page16	6	32	3
15	http://trdinfotech.com/page15	4	44	3
17	http://trdinfotech.com/page17	2	54	1
18	http://trdinfotech.com/page18	1	76	0

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

Line diagram considering above chart

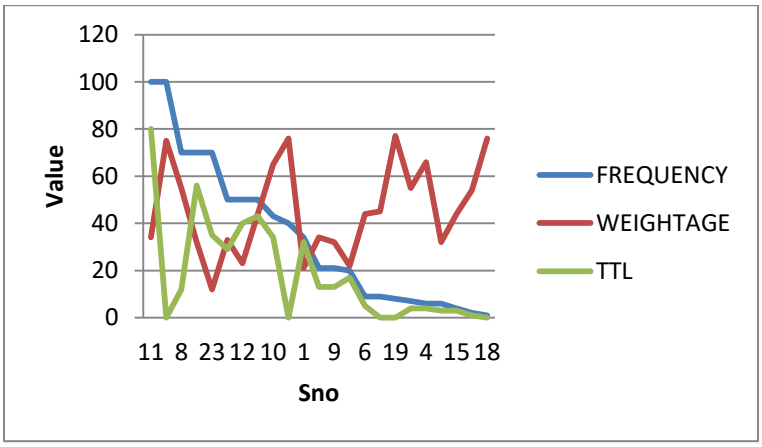


Fig 3 Graphical representation of line chart for dataset of visited sites consider only frequency

SQL QUERY TO CHECK THE RANK CONSIDERING WEIGHTAGE ONLY

SELECT * FROM CRAWLER ORDER BY WEIGHTAGE DESC

Table 3 Rank considering weight age

SNO	URL	FREQUENCY	WEIGHTAGE	TTL
19	http://trdinfotech.com/page19	8	77	0
21	http://trdinfotech.com/page21	40	76	0
18	http://trdinfotech.com/page18	1	76	0
22	http://trdinfotech.com/page22	100	75	0
4	http://trdinfotech.com/page4	6	66	4
10	http://trdinfotech.com/page10	43	65	34
8	http://trdinfotech.com/page8	70	55	12
5	http://trdinfotech.com/page5	7	55	4
17	http://trdinfotech.com/page17	2	54	1
20	http://trdinfotech.com/page20	9	45	0
13	http://trdinfotech.com/page13	50	44	43
6	http://trdinfotech.com/page6	9	44	5
15	http://trdinfotech.com/page15	4	44	3
11	http://trdinfotech.com/page11	100	34	80
3	http://trdinfotech.com/page3	21	34	13
7	http://trdinfotech.com/page7	50	33	29
14	http://trdinfotech.com/page14	70	32	56
9	http://trdinfotech.com/page9	21	32	13
16	http://trdinfotech.com/page16	6	32	3
12	http://trdinfotech.com/page12	50	23	40
2	http://trdinfotech.com/page2	20	22	17
1	http://trdinfotech.com/page1	34	21	32
23	http://trdinfotech.com/page23	70	12	35

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

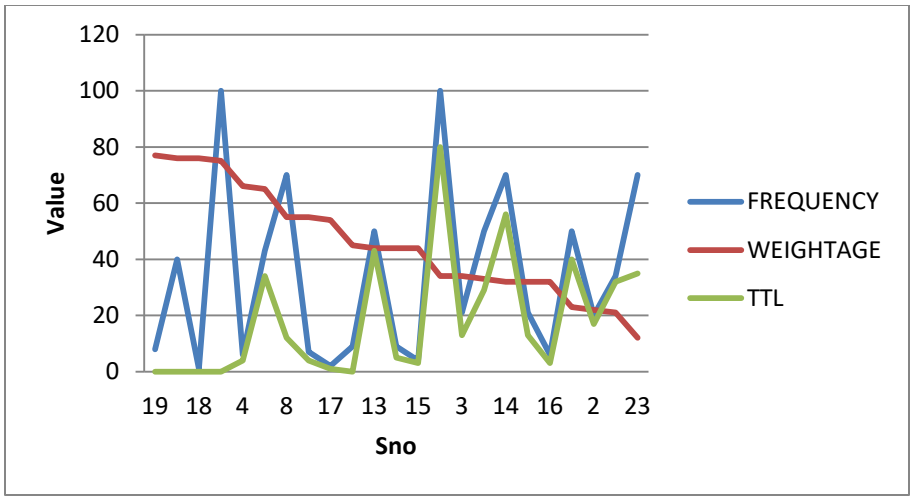


Fig 4 Graphical representation of line chart for dataset of visited sites consider only Weightage

SQL QUERY TO CHECK THE RANK AND FILTERING DURING WEB CRAWLING

SELECT * FROM CRAWLER WHERE TTL>(FREQUENCY/2) ORDER BY FREQUENCY DESC, WEIGHTAGE DESC, TTL DESC

The list of url is represented after eliminating TTL with less than half of frequency and ranking as per frequency and weightage.

SNO	URL	FREQUENCY	WEIGHTAGE	TTL
11	http://trdinfotech.com/page11	100	34	80
14	http://trdinfotech.com/page14	70	32	56
13	http://trdinfotech.com/page13	50	44	43
7	http://trdinfotech.com/page7	50	33	29
12	http://trdinfotech.com/page12	50	23	40
10	http://trdinfotech.com/page10	43	65	34
1	http://trdinfotech.com/page1	34	21	32
3	http://trdinfotech.com/page3	21	34	13
9	http://trdinfotech.com/page9	21	32	13
2	http://trdinfotech.com/page2	20	22	17
6	http://trdinfotech.com/page6	9	44	5
5	http://trdinfotech.com/page5	7	55	4
4	http://trdinfotech.com/page4	6	66	4
15	http://trdinfotech.com/page15	4	44	3

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

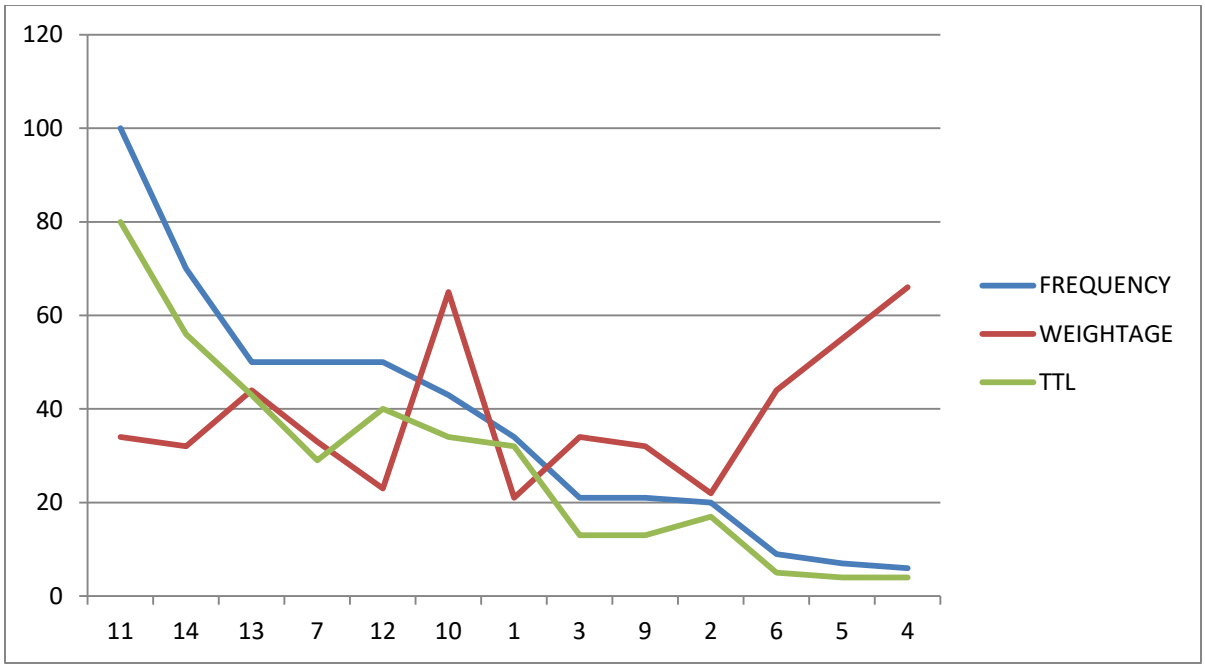


Fig 5 Graphical representation as per proposed work considering all parameters

[6] CONCLUSION

Research has successfully count the frequency of visiting the page along with weight age of the page that represents the time page visited by users. This work has used TTL value. This value has helped to find expiry time of the page to filter it. Proposed system is found capable to filter web page list during searching considering visiting frequency, weight age and TTL value. The intelligent web crawling system has successfully performed URL filter and URL ranking modules and eliminated the expired Url and set their rank. The system is intelligent as the database is updated as per user action to store the TTL value, weight age and frequency of page visiting.

[7] FUTURE SCOPE

However proposed intelligent extensible web crawling system is performing URL filter and URL ranking modules and eliminated the expired Url and set their rank. But there is need to integrate the security. As it is observed that several web sites are acting as virus carrier and influence the system of web user. In future the proposal would consider the integration of security by considering user feedback regarding website.

Reference

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

[1] J. Xu, C. Duan, L. Du, and M. Li, "The Application of Web Crawler in City Image Research," 2020 IEEE 5th Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2020, pp. 270–274, 2020, doi: 10.1109/ICCCBDA49378.2020.9095599.

[2] V. V. Mahale, M. T. Dhande, and A. V. Pandit, "Advanced web crawler for deep web interface using binary vector page rank," Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018, pp. 500–503, 2019, doi: 10.1109/I-SMAC.2018.8653765.

[3] Y. Wang, Z. Hong, and M. Shi, "Research on LDA Model Algorithm of News-oriented Web Crawler," Proc. - 17th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2018, pp. 748–753, 2018, doi: 10.1109/ICIS.2018.8466502.

[4] S. M. Nakashe and K. R. Kolhe, "Smart Approach to Crawl Web Interfaces Using a Two Stage Framework of Crawler," Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018, pp. 1–6, 2018, doi: 10.1109/ICCUBEA.2018.8697592.

[5] Z. Shi, M. Shi, and W. Lin, "The Implementation of Crawling News Page Based on Incremental Web Crawler," Proc. - 4th Int. Conf. Appl. Comput. Inf. Technol. 3rd Int. Conf. Comput. Sci. Appl. Informatics, 1st Int. Conf. Big Data, Cloud Comput. Data Sci. Eng. ACIT-CSII-BCD 2016, pp. 348–351, 2017, doi: 10.1109/ACIT-CSII-BCD.2016.073.

[6] M. Kumar and R. Bhatia, "Design of a mobile Web crawler for hidden Web," 2016 3rd Int. Conf. Recent Adv. Inf. Technol. RAIT 2016, pp. 186–190, 2016, doi: 10.1109/RAIT.2016.7507899.

[7] N. Kumar and M. Singh, "Framework for Distributed Semantic Web Crawler," Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN 2015, pp. 1403–1407, 2016, doi: 10.1109/CICN.2015.272.

[8] S. Sharma and P. Gupta, "The anatomy of web crawlers," Int. Conf. Comput. Commun. Autom. ICCCA 2015, pp. 849–853, 2015, doi: 10.1109/CCAA.2015.7148493.

[9] A. Gupta and P. Anand, "Focused web crawlers and its approaches," 2015 1st Int. Conf. Futur. Trends Comput. Anal. Knowl. Manag. ABLAZE 2015, pp. 619–622, 2015, doi: 10.1109/ABLAZE.2015.7154936.

[10] G. H. Agre and N. V. Mahajan, "Keyword focused web crawler," 2nd Int. Conf. Electron. Commun. Syst. ICECS 2015, pp. 1089–1092, 2015, doi: 10.1109/ECS.2015.7124749.

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

[11] A. Aghamohammadi and A. Eydgahi, "A novel defense mechanism against web crawlers intrusion," 2013 Int. Conf. Electron. Comput. Comput. ICECCO 2013, pp. 269–272, 2013, doi: 10.1109/ICECCO.2013.6718280.

[12] K. S. Shetty, S. Bhat, and S. Singh, "Symbolic verification of web crawler functionality and its properties," 2012 Int. Conf. Comput. Commun. Informatics, ICCCI 2012, 2012, doi: 10.1109/ICCCI.2012.6158649.

[13] W. Ma, X. Chen, and W. Shang, "Advanced deep web crawler based on Dom," Proc. 2012 5th Int. Jt. Conf. Comput. Sci. Optim. CSO 2012, pp. 605–609, 2012, doi: 10.1109/CSO.2012.138.

[14] W. Guo, Y. Zhong, and J. Xie, "A web crawler detection algorithm based on web page member list," Proc. 2012 4th Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2012, vol. 1, pp. 189–192, 2012, doi: 10.1109/IHMSC.2012.54.

[15] O. Jalilian and H. Khotanlou, "A new fuzzy-based method to weigh the related concepts in semantic focused web crawlers," ICCRD2011 - 2011 3rd Int. Conf. Comput. Res. Dev., vol. 3, pp. 23–27, 2011, doi: 10.1109/ICCRD.2011.5764237.

[16] F. Ahmadi-Abkenari and A. Selamat, "A clickstream-based web page significance ranking metric for web crawlers," 2011 5th Malaysian Conf. Softw. Eng. MySEC 2011, no. 1, pp. 223–228, 2011, doi: 10.1109/MySEC.2011.6140674.

[17] M. S. Zhao, P. Zhu, and T. C. He, "An intelligent topic web crawler based on DTB," Proc. - 2010 Int. Conf. Web Inf. Syst. Mining, WISM 2010, vol. 1, no. 1, pp. 84–86, 2010, doi: 10.1109/WISM.2010.155.

[18] S. Shekhar, R. Agrawal, and K. V. Arya, "An architectural framework of a crawler for retrieving highly relevant web documents by filtering replicated web collections," ACE 2010 - 2010 Int. Conf. Adv. Comput. Eng., pp. 29–33, 2010, doi: 10.1109/ACE.2010.64.

[19] H. C. Wang, S. H. Ruan, and Q. J. Tang, "The implementation of a web crawler URL filter algorithm based on caching," 2nd Int. Work. Comput. Sci. Eng. WCSE 2009, vol. 2, pp. 453–456, 2009, doi: 10.1109/WCSE.2009.851.

[20] P. Xiang, K. Tian, and Q. Huang, "A framework of deep web crawler," Proc. 27th Chinese Control Conf. CCC, pp. 582–586, 2008, doi: 10.1109/CHICC.2008.4604881.

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

[21] V. Kaewmarin, N. Arch-Int, and S. Arch-Int, "Semantic web service discovery and integration using service search crawler," 2008 Int. Conf. Comput. Intell. Model. Control Autom. CIMCA 2008, pp. 884–888, 2008, doi: 10.1109/CIMCA.2008.78.

[22] J. Akilandeswari and N. P. Gopalan, "An architectural framework of a crawler for locating deep web repositories using learning multi-agent systems," Proc. - 3rd Int. Conf. Internet Web Appl. Serv. ICIW 2008, pp. 558–562, 2008, doi: 10.1109/ICIW.2008.94.