International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2017 www.irjet.net p-ISSN: 2395-0072

TO ENHANCE LEARNERS' PERFORMANCE IN THEIR ACADEMICS THROUGH MODIFIED GOOGLE SEARCH ENGINE

Mr. Lovkesh Singh Vermani, Dr. Akash Saxena

Mr. Lovkesh Singh Vermani, Research Scholar, Tania University, Sri Ganganagar, INDIA
Dr. Akash Saxena, Professor, Compucom Institute of Technology & Management, Jaipur, INDIA

Abstract: As you know that internet is plentiful source of information. The increased rate of the web recourses is exponential. I have explored different aspects of internet search behavior of North India school learners, in terms of user's background and experience on Internet, purpose of use of internet, searching skills, query formulation, frequency of use, favorite search engine, etc. All these factors affect the way of searching over the internet. Data searched by the learners collected through questionnaire, then we need to apply Page Rank algorithm to reach to productive result. Now a days Cluster analysis mining techniques become popular tool to solve statistical problems which involves objects, groups of respondents or cases that have similar properties to one another in the same group and different properties from those presents in other groups. But by this way, we can make group of contents although collected contents related to education, due to large amount of data a learner will not able to use it in appropriate manner therefore we need to devise a methodology/ideology. With the help of that methodology a learner will get suitable data for their learning purpose within stipulated time. Every educator has keen interest in developing educational contents and provide to target learners. To produce robust and consistent solutions, it is imperative that researcher select desirable approach therefore upcoming learner will get educational contents in order of their research. It is critical thing if a learner try to search on the internet for a particular query, because while searching a learner will get a huge amount of data and to reach to the particular contents it takes lotz of time sometime it takes away the interest of the learner. Therefore with the help of forthcoming research data by the learner will give benefits to the upcoming learner, and for this we need to modified Page Rank Algorithm.

Keywords: Internet surfing, Subjective information, Learner search behavior, Clustering, educational data mining, Search frequency, page rank, algorithm

1. INTRODUCTION

A multistage random sampling method was used as follows: All High Schools in North India formed the primary sampling frame and were clustered according day and boarding school. I purposively selected mixed (co-educational) schools and a total of 3 schools were sampled. The schools were The Punjab Public School, Nabha, Mayo College, Ajmer, Welham School, Dehradun.

One hundred twenty learners were selected from each school. Individual learners were selected or recruited into the study using systematic random sampling, whereby every stream in a school formed a secondary sampling frame.

Thereafter, the sampling interval was calculated based on a 10-15% of the school population. The result was then divided by the number of streams in the school to determine the sample size per class. The sampling interval (n) was then calculated by dividing the number of learners in a stream by the sample size. The first learner was selected blindly using a table of random numbers after which the remaining learners were selected at regular intervals (10) from the secondary sampling frame. This process was continued per stream until the required school sample size of 120 learners was achieved from each school.

Questionnaire was used as research instrument in collecting the data and SPSS version 20.0 was used for the analysis. All respondents were given serial numbers to facilitate coding and analysis. Frequency tables and chi-square were also used in presenting the data. Conclusions from relevant related literature were captured along to authenticate the findings of the study.

2. SOURCES USED IN ACQUIRING RELEVANT INFORMATION ON THE INTERNET

In second survey, we have provided equal opportunity for all 360 learners, we have provided Computer system and internet connection with 24 X 7 service to them. Then we have asked set of learners which denoted with 'A' to explore educational contents on internet according to their subjective knowledge, while exploring most of the learners went to Wikipedia and Youtube websites. Some learners visited educational websites like khan academy, Tataedge, Jagranjosh.com, shiksha.com, India education etc. we have conducted test based on their surfed educational contents.

But only 5% of them scored more than 50%. Therefore we have checked all the contents which were explored by the learners. More than 70% time they have wasted to reach towards the exact educational data.

Relevant information	Frequency	Percent
Acquired relevant information	372	30
Not Acquired relevant information	868	70
Total	1240	100

Assignment Score	Frequency	Percent
Got less than 50%	1178	95
Got more than 50%	62	5
Total	1240	100

Second time again we have asked to explore the same contents but this time we have make a list of educational website where they can surf easily about their educational contents related to their subject. This time learners explore subjective contents better than before.

Relevant information	Frequency	Percent
Acquired relevant information	963	77.7
Not Acquired relevant information	277	22.3
Total	360	100

Assignment Score	Frequency	Percent
Got less than 50%	1040	83.8
Got more than 50%	200	16.1
Total	1240	100

Third time we have given to learners different topics and asked them to search on internet and complete online assignment. This time again, we have given the list of educational website where they have to search on. But most of the time they have explore all kinds of given educational websites and explore topics and subtopics related to given problem, topic and concept. But after search experienced this time 40% learners search relevant information.

Relevant information	Frequency	Percent
Acquired relevant information	496	40
Not Acquired relevant information	744	60
Total	1240	100

Assignment Score	Frequency	Percent
Got less than 70%	1153	93
Got more than 70%	87	7
Total	1240	100

After third approach, we have conducted the test and but only 17% learners got more than 70% marks. We have collected all the educational contents explored by those learners those who have got more than 70% marks.
Forth time we have given same mathematical problem, biology topic and physics concept to other 'B' set of learners and asked them to explore but from the collected educational contents which were explored at the time of third approached by those who have scored more than 70% marks.

Assignment Score	Frequency	Percent
Got less than 70%	69	5.6
Got more than 70%	1171	94.4
Total	1240	100

Then we asked 'B' Set of learners to complete the assignment, B set of learners, 94.4% learners scored more than 70%. With the analysis of these testing for all the cases we reached to final result, and final result given a complete idea that when we collect the educational contents and arrange it according to learner's requirement then learners attain more attention towards these contents and scored better.
To judge relevant information on the Internet by using five point semantic differential scales. The results indicated that to some extent learners used reference URLs acquired by Insturctor and learners, friends and family and search history respectively

3. BROWSING HITS:

Different studies of the use of internet found that almost all users look at the first page of results only. Most users are satisfied that these initial ten or so results are good enough to answer their information need (Craven & Griffiths, 2002; Sullivan, 1998, 2002). Again the results of this study correspond to global trends as most of the

learners (202, 56.3%) generally browsed through first ten results. 95 (26.5%) learners marked the second option (20 hits)

Browsing hits

	Frequency	Percent
10 hits	698	56.3
20	326	26.3
30	134	10.8
40	54	4.4
50	14	1.1
More than 50	14	1.1
Total	1240	100.0

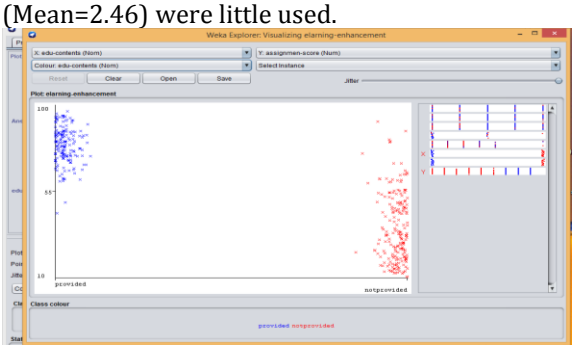
Table 11. Criteria of judging relevancy of information

	Mean	SD
Title	3.86	1.315
Highlighted words	3.46	1.155
Descriptions	3.35	1.168
URLs	2.46	1.219

Note: 1=Rarely----5=Often

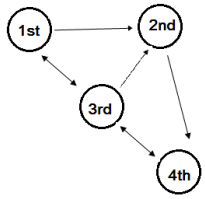
Relevancy of results

The World Wide Web contains tremendous amount of information, however it should be admitted that the problem of discovering the relevant resources has become intense, as the quantity of its information grows, people's ability of finding relevant materials has decreased dramatically as Safari (2005) called it "a needle in the haystack." The learners were asked whether they usually found the required information on the Web. The results state that to moderate extent they usually found the requiring information (Mean=3.74). They were also inquired about their relevancy judging criteria. The responses demonstrate that learners usually identified the relevancy by title (Mean=3.86) and highlighted words (Mean=3.46). Descriptions (Mean=3.35) and reading URLs (Mean=2.46) were little used.



It depicts that Assignment scored is depends upon if we provide educational contents to learners. There is significance relationship between specific provided educational contents and assignment score.

According to Google Page Rank algorithm I have calculated the page rank PR for each page then further enhance the Page Rank algorithm so that every upcoming learner take benefits from previous explored page.



Page	Iteration 0	Iteration 1	Iteration 2	$PR_{(page)}$
1st	1/4	1/12	1.5/12	1
2nd	1/4	2.5/12	2/12	2
3rd	1/4	4.5/12	4.5/12	4
4th	1/4	4/12	4/12	3

$$PR_1 = \sum_{O \in L(1)} \frac{PR(O)}{L(O)}$$

Page Rank Algorithm is based upon above formula. i.e. the PageRank value for page 1st is dependent on the PageRank values for each page O contained in the set L(1) (the set containing all pages linking to page 1st), divided by the number L(O) of links from page O.O stands for other pages. Dumping factor including in page rank algorithm to give some rank to the page even it is not linked by other pages. But I have not included dumping factor.

Calculate final Page Rank for the page 1

Step 1: Set the loop to the First Page

Step 2: Initialize page 1 with the page rank PR1,

$$PR_1$$

Step 3: Calculate number of users visited page 1 i.e. U1 divide by total number of users i.e. N

$$\frac{U_1^n}{N}$$

Step 4: Calculate the average marks got by the users in given assignment after visited page 1. e.g:- M1+M2……..+Mi is marks percentage of users U visited first page U1.

$$\frac{1}{U_1^n} \sum_{i=1}^{i=U_1^n} Mi$$

Step 4: Calculate the Final page rank PR1 for the Page 1

$$PR_1 = P_{R1} + \frac{U_1^n}{N} + \frac{1}{U_1^n} \sum_{i=1}^{i=U_1^n} Mi$$

PR_1	Final Page Rank for First Page Visited by the User
N	N is total Number of pages visited by the users
U_1^n	First Page visited by the Users U^n for n users
P_{R1}	Page Rank for the first page, suggest by search engine
Mi	Marks percentage got by the users in assignment, after visited First Page

Step 5: Page Number will be incremented to the next page
Step 6: Repeat step 2 to 5 until loop will reach to the last page i.e. N

Page	Iteration 0	Iteration 1	Iteration 2	$P_{R(page)}$	$PR_{(page)}$	$FPR_{(page)}$
1st	1/4	1/12	1.5/12	1	83.5	1
2nd	1/4	2.5/12	2/12	2	32	4
3rd	1/4	4.5/12	4.5/12	4	74.4	2
4th	1/4	4/12	4/12	3	43.1	3

$$PR_1 = P_{R1} + \frac{U_1^n}{N} + \frac{1}{U_1^n} \sum_{i=1}^{i=U_1^n} Mi$$

4. CONCLUSION

I have stats data uploading on World Wide Web by the users. They send 204 million emails, share 2.5 million pieces of content on Facebook, send 277,000 Tweets, post 216,000 photos on Instagram every hour and upload 300 hours of video on YouTube every minute.
We have proof that sufficient amount of data is available on WWW but question arises is it fruitful for our upcoming generation in their studies we need to either upload only sense data of we need to generate an algorithm or search medium with lots of filter with the use of which upcoming generation will get only fruitful data according to their requirement. We need to apply powerful technology to form structured data from unstructured data. Data may be in any form need to reform which is to be useful for our education community. That's often the case with online

Step 7: PageRank PRi will be sorted out with the help of Bubble sort in ascending order where N stands for total number of pages explored by the users.

a) for i=1 to N
b) for j=2 to N-i
c) if PR[j-1] > PR[j]
d) Swap(PR[j-1]: PR[j])
e) end if
f) end for
g) end for
h) return list
i) end Bubble_Sort

Now Array address value specified the page rank for the pages explored by the users, maximum Page Rank specified with the maximum array address value, means if Page Rank has maximum array address value, it means we take this page is very important page than other pages for the users point of view. PR[i] has more importance than PR[i-1] and so on.
Frequency of importance of page or Final Page Rank According to the learner point of view i.e. FPRi is directly proportional to PRi
FPRi α PRi

textual data in the form of news articles, social media posts, forums comments, and video uploading, much more. In fact, we made complex environment ourselves with realizing the future generation we are keep uploading data on the internet without the powerful technology again saturation level will come where we have the data of all kind but without having knowledge how to search and from where to search this data becomes unruly and useless for all.

Some Companies already develop searching platform on the basis of contents required by the young children like they have developed Kiddle.co. which is a web search engine and encyclopedia emphasizing that it is a safe search engine for the young children. Kiddle is powered by Google Custom Search and imply SafeSearch, with additional filters. The user enters topics in the search toolbar, and Kiddle returns and ranks its findings, with its first three results being suitable for children to read as well as of an easy reading level.
The domain was registered in 2014.The designers chose the .co domain name in order to emphasize that is for "children only." Kiddle became very popular in social media in 2016, and even became a meme due to blocking of certain words for a short period of time
Kiddle uses an outer space theme for their pages. It appears like a Google Custom Search window. After the user enters the subject, Kiddle presents search results, with the first three results being deemed safe and written specifically for kids and "checked by Kiddle editors", the

next four being safe sites not written specially for kids, but presented in kid friendly language. The eighth result and anything else beyond are safe sites written for adults but harder to understand for kids. Results presented are filtered through Google SafeSearch.
If the user enters what are considered "bad words" a picture of a robot is displayed, telling the user to try again.

In 2016, Kiddle was criticized for blocking the keywords "gay," "lesbian," and the phrase "sex education," claiming that it was OK for a twelve year old to know about, but not a five year old.They had also banned the words "transgender", and "bisexual". Due to public outcry, Kiddle announced they would unblock the word. Another similar websites are KidRex, KidzSearch etc.

To remove such type of complexity of data finding process we need to put a backlash against unruly data. There are now thinking developing in the big companies that how to structure the unstructured data. In a result this point of view is correct without structure, collection of all kind of data is not usable. It is just a mass of unrelated information that would take years to comprehend, and even then, may not yield any insights. But if structure can be overlaid effectively and analyzed, that is when unstructured data starts to become structured data.
This can only be possible when we infuse recommendation for every data which is to be uploaded on internet or after uploading we can categories those data according to user requirement by using data mining algorithms. Most appropriate algorithm is K- means clustering algorithm or instance based algorithm which is being applied on user logs.

After applied algorithm on data, it should be filtered structured wise and unstructured data, needs to be separated, therefore upcoming learner or user will get appropriate data according to their search. Which later on gives productive result everywhere.

REFERENCES:

- Aitken, W. (2007). Use of web in tertiary research and education. Webology, 4 (2), article 42. Retrieved February 16, 2009, from http://www.webology.org/2007/v4n2/a42.html
- Aula, A. (2003). Query formulation in web information search. In Isaias, P. & Karmakar, N. (Eds.), Proceedings of IADIS International Conference WWW/Internet 2003 (pp. 403-410). IADIS Press.
- Bashir, S., Mahmood, K., & Shafique, F. (2008). Internet use among university learners: a survey in University of the Punjab, Lahore. Pakistan Journal of Library and Information Science, 9, 49-65.
- Batthini, G., & Madnani, A. (2003). Web search behaviour of LIS professionals of selected libraries of Ahmedabad and Gandhinagar: a study. Retrieved January 9, 2009, from http://ir.inflibnet.ac.in:8080/jspui/bitstream/1944/217/1/cali_48.pdf
- Chang, N. C., & Perng, J. H. (2001). Information search habits of graduate learners at Tatung University. International Information & Library Review, 33 (4), 341-346.
- Cmor, D., & Lippold, K. (2001). Surfing vs. searching: the Web as a research tool. Retrieved January, 16, 2009, from http://staff.library.mun.ca/~DCmor/stlhe/
- Craven, J., & Griffiths, J. R. (2002). 30,000 different users . . . 30,000 different needs? Design and delivery of distributed resources to your user community. In Brophy, P., Fisher, S. & Clarke, Z. (Eds.), Libraries without walls 4: The delivery of library services to distant users (pp. 173-186). London: Facet.
- Ebersole, S.E. (2005). On their own: learners' academic use of the commercialized Web. Library Trends, 53, 530-538.
- Griffiths, J. R., & Brophy, P. (2005). Learner searching behavior and the web: use of academic resources and Google. Library Trends, 53, 539-554.
- Holscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. Computer Networks, 33, 337-346.
- Hsieh-Yee, Ingrid. (2001). Research on Web search behavior. Library & Information Science Research, 23, 167-185.
- Jansen, B. J., & Pooch, U. (2001). A review of web searching studies and a framework for future research. Journal of the American Society for Information Science and Technology, 52, 235-246.
- Johnson, F., Griffiths, J. R., & Hartley, R. J. (2001). DEVISE: a framework for the evaluation of Internet search engines. Retrieved February 16, 2009, from http://www.cerlim.ac.uk/projects/devise/index.php
- Lenhart, A., Madden, M. & Hitlin, P. (2005). Pew Internet and American life project: teens and technology. Retrieved January 25, 2009, from http://www.pewinternet.org/pdfs/PIP_Teens_Tech_July2005,web.pdf

- Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive strategies in web searching. Proceedings of the 5th conference on human factors & the Web. Retrieved January 9, 2009, from http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html
- Nicholas, D., Huntington, P., Jamali, H. R., Rowlands, I. & Fieldhouse, M. (2008). Learner digital information seeking behaviour in context. Journal of Documentation, 65 (1), 106-132.
- Poulter, A. (2003). World Wide Web. In Feather, J. & Sturges, P. (Eds.), International encyclopedia of information and library science. London: Routledge.
- Safari, M. (2005). Search engines and resource discovery on the Web: Is Dublin Core an impact factor? Webology, 2 (2), article 13. Retrieved February 16, 2009, from http://www.webology.org/2005/v2n2/a13.html
- Safdar, M. (2008). Internet use behavior and attitude of college learners: A survey of leadership colleges network. Unpublished master's thesis. University of the Punjab, Lahore.
- Spink, A., & Jansen, B. J. (2004). A study of web search trends. *Webology*, 1 (2), article 4. Retrieved February 16, 2009, from http://www.webology.org/2004/v1n2/a4.html
- Sullivan, D. (1998). *Counting clicks and looking at links*. Retrieved January 16, 2009, from http://searchenginewatch.com/2166431
- Sullivan, D. (2002). *Search engine placement tips*. Retrieved February 7, 2009, from http://searchenginewatch.com/2168021
- Ullah, M. (2007). Use of Internet by medical postgraduate trainees. *Pakistan Library and Information Science Journal*, 38 (3), 16-20.
- Warraich, N. F., & Ameen, K. (2008). Perceptions of library and information science professionals about a National Digital Library programme. *Library Hi Tech News*, 25 (8), 15-19.
- Zhang, X., Anghelescu, H. G. B., & Yuan, X. (2005). Domain knowledge, search behaviour, and search effectiveness of engineering and science learners: an exploratory study. *Information Research*, 10 (2), paper 217. Retrieved January 25, 2009, from http://informationr.net/ir/10-2/paper217.html

AUTHOR'S BIOGRAPHY:

Mr. Lovkesh S Vermani, Computer Teacher, IT Administrator, The Punjab Public School, Nabha
Pursuing Ph.D. From Tantia University in Computer Science and Engineering. In his research, he is exploring methods to enhance E-learning in the context of educational data.

Mr. Akash Saxena, Professor, Compucom Institute of Technology & Management, Jaipur, India. He is truly a Mentor & Guide for upcoming researcher in computer science and engineering field.