Algorithm to Filter & Redirect the Web Content for Kids'

Neha Gupta
Research Scholar, FET Manav Rachna International University
Faridabad, Haryana, India
Neha.fbc@mriu.edu.in

Dr. Saba Hilal
Professor, School of Computer Science
Lingayas University
Faridabad, Haryana, India
Saba21hilal@gmail.com

***Abstract*—searching the content for the kids is different from searching for adults. The needs of the kids are different and the way they search the content on the web is also different. There are a lot of issues that needs to be taken care of while searching for the kids. Our aim in this paper is to develop an interface that can help parents, teachers and kids to search the web in a secure way and make the interaction session with the web a pleasurable experience. The Algorithm discussed will keep the kid focused by not blocking or filtering the websites but by redirecting the interest of the kid towards the educational side of his interest. The interface designed using the algorithm will not block any website but has redirected the query to the educational part of kids' interest.**

Keywords—Content Filktering, Redirection, Kids, Education, Algorithm

I. INTRODUCTION

Although there is a lot of material available on the web for the kids like Educational software, encyclopedias, notes, ppt, Educational Content Websites, School Websites, Virtual courses, Learning Management Systems, Collaborative Learning Environments etc, still to access the content there is no central controlling authority who can guide the innocent kids about what to search and open and what not. We can imagine web like a big library of millions of web pages not arranged in any order, so to extract the content, the need and the help of search engines are required but no search engine can give the result as per the need of the kid and can take care of all the below mention aspects.

1. Grammar, spelling and query formation
2. Understanding of topic
3. Vulnerable environment of web
4. Need of secure content
5. Filtration of valid or relevant content
6. Age appropriate content
7. Quality of content
8. Focus on educational or secure content
9. Security from cyber bullying, phishing, internet crimes etc

II. CLASSIFICATION OF CONTENT SEARCHED BY THE KIDS

To make searching a pleasurable and secure experience for kid it is necessary to understand the search need of the kid and his interests. During the course of the study, students of various schools have been interviewed about their interest on the web using mobile devices. Also the teachers and the parents of various kids are also interviewed regarding the use of internet on mobile devices for the school assignments of their child. Based on the results of these interviews the content searched by the kids on the web can be classified into two categories:

1. For doing their home assignments and the related projects
2. For the entertainment or for social networking purpose

So, broadly we can say the content searched by the kids is either

1. Educational
2. Non- educational

Educational content can be any content related to their home assignment, project, searching a topic of subject, class presentation or any other work related to their academic growth.

Non-educational content can be any content related to entertainment like games, movies, songs or for social networking like Face book, Twitter, and Orkut or for general interest like roadies, mobile wallpapers, new updates etc.

III. CLASSIFICATION OF KIDS ACCORDING TO THEIR AGE

During the course of the study, a lot of interaction and brainstorming sessions were conducted with the kids to understand their search needs. During these sessions kids of age group 6 to 15 have participated. While understanding their search needs it has been observed that search need of the kid change as he grows. The kids between the ages of 6 to 10 have different search need and different perspective towards the type of content they are looking for their assignments, while kids between ages of 11 to 15 have entirely different needs and different perspective towards web. The kids between age 6 to 10 need assistance while typing the queries, selecting the content, navigating the web while kid of age 10+ is independent enough to browse the web, select the content on his own and have the decisive capabilities The table below can explain the difference in the mental level of the kids of different age groups:

IV. RECHANNALIZATION OR REDIRECTION OF CONTENT SEARCHED TOWARDS EDUCATIONAL CONTENT

Kids often enjoy searching non-educational content. Facebook is the most preferred destination of every teen. Web games or mobile games can be the second choice. There is no harm in doing all these things but vulnerable environment of the web and the inappropriate content can cause damage. As the aim of the study is to provide secure and age appropriate content to the kid, so blocking the interest of the kid will not help. Blocking the website or the content can be easy way and has been tried by many researchers but eventually kids lose their interest in the interface and start looking for other alternatives.

Our aim is to keep the kid focused by not blocking or filtering the websites but to redirect the interest of the kid towards the educational side of his interest. The interface designed has not blocked any website but has redirected the query to the educational part of kids' interest

Parameter	**Ages between 6 to 10**	**Ages between 11 to 15**
Query formation	*Needs assistance*	*Independent*
Selection of content	*Needs assistance*	*Independent*
Navigation of web	*Needs assistance*	*Independent*
Type of content searched	*Educational*	*Educational and non-educational*
Usage of web/ Time spent	*When required*	*Very often*
Possibility of security threat	*Less*	*High*
Need of secure content	*High*	*Very high*
Health problems	*Low*	*High*

V. RELATED RESEARCHES

The paper [1] discusses various characteristics of kids that one should keep in mind while developing any information retrieval system for kids. The experiments and results are shown using adult based search interfaces. The author of [2] has focused on information need of kids by examining their search behavior. The author has tried to find the answer of few questions that needs to be answered by any IR system of kids. Another paper [3] analyses the session and queries for kids information need and compare them with general queries and sessions. The author has enriched the AOL query log by implementing the result of kid's queries. The paper [4] presents the use of query assistance and search moderation techniques for kids so that kids have a better experience searching the web. The author has also focused on interface design for kids. The author [5] has analyzed a large query log from a commercial search engine and identifies the problems related to child search behavior. The target audiences of his work was child from age 6 to adult of 18. They have also worked on search difficulties based on query matrices.

The paper [6] presents a study of 64 fifth grade students who were using science library catalogue for searching the content on the web. The study highlights the problems of kids while searching and the possible solution. The paper [7] presents an automatic way of identifying the web page suitable for kids. The focus is on child psychology and cognitive science. The author has investigated the potential of combining topical and non-topical aspect of identifying age appropriate content for kids. The paper [8] discuss cognitive specifies of children and the way they can be encoded for classification. The author has worked on two dimensions: child friendliness and focus toward child audiences. The author [9] discuss project Gutenberg to make available classic literature to children in a secure way. The paper [10] presents an interaction based information filtering system for kids. This system focuses on user interaction modelling, user evaluation, automatic detection of child friendly information etc. The paper [11] presents a system named Tad Polemic which will assist children in searching the web for difficult topics and also provide filtering of content based upon child interest and age.

The author of [12] has conducted a study to gather the quantitative and qualitative data about children interaction with web search engines. They identified that kids perform poorer on metaphorical interfaces and good on Google.

The paper [13] presents a paradigm to identify the suitable videos for kids on youtube on the basis of various features like people reviews, comments, author information, community information etc. The paper [14] tries to uncover methods and techniques that can be used to automatically improve search results on queries formulated by children.

VI. THE ALGORITHM USED

To cope up with the requirement of the kids and for ensuring the security and relevancy of content for them an algorithm has been designed to filter the content of the web as per the user requirement and above all redirect the content searched towards the educational aspect. Restricting the content will not resolve the issue as the child will start searching for the alternatives, so instead of restricting, the algorithm will redirect the result towards the educational aspect intelligently.

The algorithm is as follows:

Step 1:

Defining various variables that can be used:
Let W be the web from where the results will be retrieved.
Let Q be the string to capture the query entered by the kids
Let NE[n] is the array of terms which are categorized as non educational among kids.
Let A be the age of the kid specified by the parent.
Let S[a] is the array of sieve coupon used.
Let G be 9.
Let EQ be the enhanced query to be submitted.
Let R be the relevant link and NR be the non relevant link
Let L is the link of Res, i.e. L Res
Let Status = Accept/ Reject w.r.t. L Res
Let Dis is the array of links to be displayed to the user.
Let n, m be the integer variables

Step 2:

User enter the query and the query string is stored in Q
Removal of stop words from the Q and the words of Q are stored in character array B[m].
Compare B[m] with NE[n]

```
If match found
        Check if A<=G
        Then
        EQ= (Q) U S[a1] U S[a3]
        Else
        EQ= (Q) U S[a2] U S[a4]
Else
        If A<=G
        Then
        EQ= (Q) U S[a1]
        Else
        EQ= (Q) U S[a2]
```

Step3:
Submit the EQ via the search engine to W. The search engine will return the result in the variable Res.
Res = R + NR

Step 4:
Extract top 10 links L[b] from Res.
n=1, m=10
Evaluate Res (Link[b]) against SC
If Status = = Accept
Then
Dis = L[bn]
Else
Res= Res - L[bn]
Res= Res + L [bm+1]
m=m+1
SC= SC+ L[bn]
Repeat until 10 links are there in the variable Dis.

Step 5:
Present the Dis to the user.
If user quit
Then
Take user feedback
Else
If user reenter the query in Q
Then
Goto step 2 and repeat until step 5 till user find the correct result.

VII. ALGORITHM IMPLEMENTATION AND THE RESULT PRESENTATION

The algorithm discussed above has been implemented in Java and the results have been analyzed for the kids' requirements. An example has been taken for the illustration purpose. For example if a kid want to login to Facebook, the situation can be handled in two ways:

Way1 will burn the fire for searching alternative options to open the Facebook which may lead to jumping over to sites or the software which may contain inappropriate content or may cause cyber bullying. While way 2 will have the probability that they will open one or the either link, explore and understand other aspect of Facebook as well.

Way 1:
Block Facebook using monitors or filters available in the market as shown below:

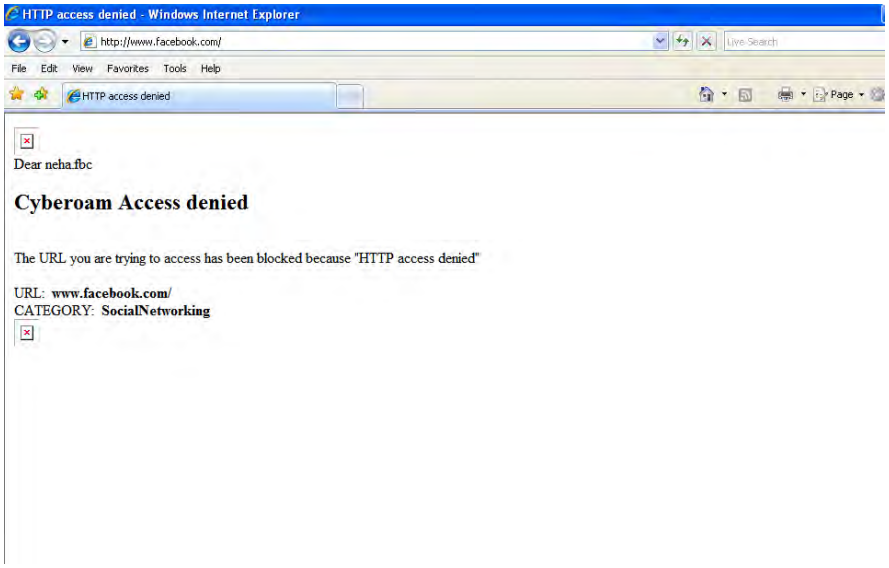

Figure 1

Way 2:
Display the result of the Facebook by redirecting them using the algorithm as shown below: or like this

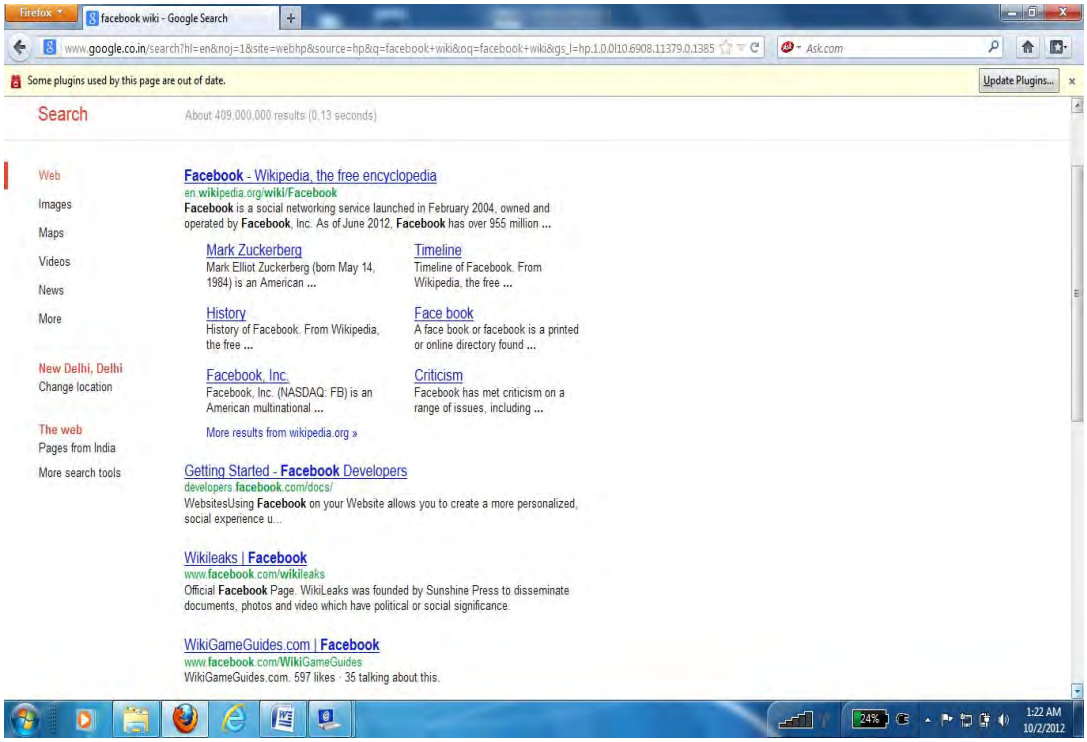

Figure 2

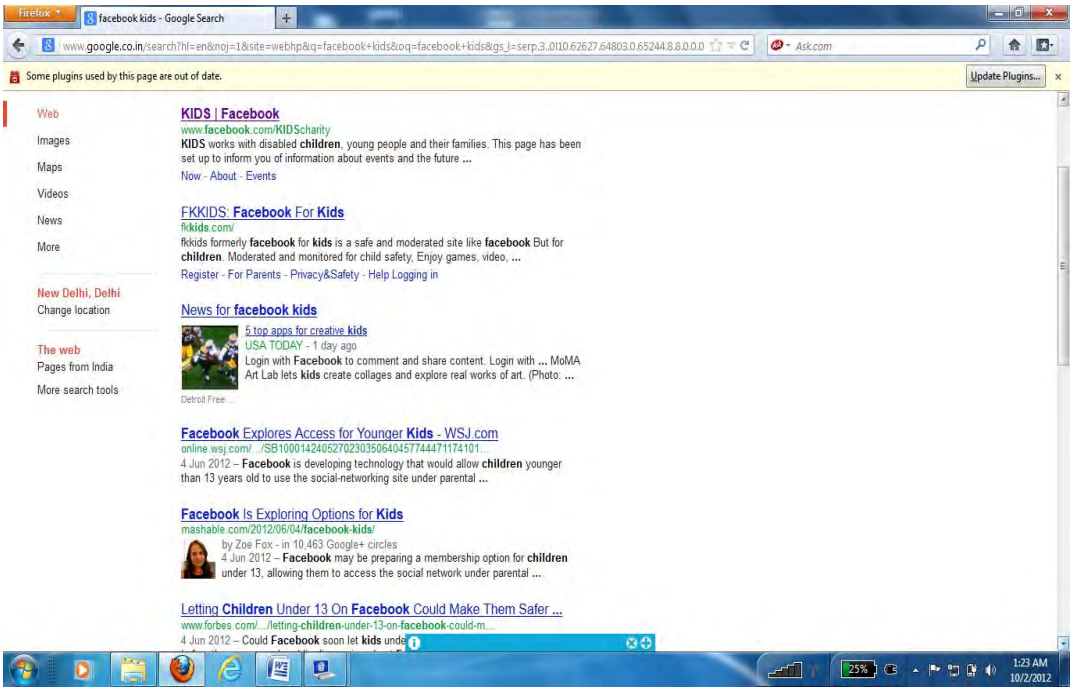

Figure 3

VIII. ADVANTAGES OF USING THE ALGORITHM

The Algorithm discussed above can have following benefits:

1. No need to block the website or the soft wares
2. Feeling of restriction on access to internet is not there.
3. No blocking still the secure content.
4. Safety from various security threats
5. No need to buy expensive filters, monitors or locks
6. Can keep kid busy while not giving full access.
7. Can help kid explore his creative side without let him feel cheated.
8. Can be used with both educational and non-educational queries.

The most important benefit of the algorithm is to let you play psychologically with your child. By redirection parents can assure their child that they are not blocking anything for them and yet don't give them the full access of the web.

IX. CONCLUSION

In this paper, the focus is on implementing the algorithm discussed above for kids searching so that the kids can access the filtered content in secure way and the inappropriate content will be filtered or the interest of the kid is redirected to educational side. The algorithm is implemented using java technology and the results have been analyzed.

REFERENCES

[1] Jochmann-Mannak, H.E. and Huibers, T.W.C. and Sanders, T.J.M. (2008) Children's information retrieval: beyond examining search strategies and interfaces. In: The 2nd BCS-IRSG Symposium: Future Directions in Information Access, 22 Sept 2008, London.

[2] S.Duarte,"Information retrieval for children based on aggregate search paradigm", University of Twente (Internal Report), 2011.

[3] Duarte Torres, Sergio and Hiemstra, Djoerd and Serdyukov, Pavel (2010) Query log analysis in the context of information retrieval for children. In: 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, July 19-23, 2010, Geneva, Switzerland.

[4] Eickhoff, Carsten and Polajnar, Tamara and Gyllstrom, Karl and Duarte Torres, Sergio and Glassey, Richard (2011) Web Search Query Assistance Functionality for Young Audiences. In: 33rd European Conference on Information Retrieval, ECIR 2011, 18-21 April 2011, Dublin, Ireland.

[5] Duarte Torres, Sergio and Weber, Ingmar (2011) What and how children search on the web. In: 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, 24-28 October 2011, Glasgow, UK.

[6] Sandra G. Hirsh, "How Do Children Find Information on Different Types of Tasks", Children's Use of the Science Library Catalog. Library Trends 45(4): (1997)

[7] Carsten Eickhoff et.al, "A combined topical/non-topical approach to identifying web sites for children", WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining Pages 505-514, 2011

[8] Carsten Eickhoff Et.al, " Web page classification child suitability", CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management Pages 1425-1428,2010

[9] Hauff, C. and Trieschnigg, R.B., “Enhancing Access to Classic Children’s Literature.” BooksOnline'10 Workshop at CIKM 2010, 26 Oct 2010
[10] R. Glassey, L. Azzopardi, D. Elliott and T. Polajnar., "Interaction-based Information Filtering for Children", In Proceedings of the 3rd Information Interaction in Context Symposium (IIiX '10) New Brunswick, NJ, USA, 2010.
[11] Karl Gyllstrom, Marie-Francine Moens, “Clash of the Typings: Finding controversies and children's topics within queries”, ECIR'11 Proceedings of the 33rd European conference on Advances in information retrieval, pg 80-91, 2011
[12] Jochmann-Mannak, H., Huibers, T., Lentz, L., Sanders, T. (2010) Children searching information on the Internet: Performance on children’s interfaces compared to Google. In Proceedings of the Workshop on Accessible Search Systems, SIGIR ’10, July 23, 2010
[13] Carsten Eickhoff, Arjen P. de Vries, Identifying Suitable YouTube Videos for Children, Proceedings of the 3rd Networked & Electronic Media Summit (NEM),2010
[14] M.G. van Kalsbeek , J.J. de Wit, “Automatic Reformulation of Children’s Search Queries”, University of Twente (Internal Report), 2010.)