Role of Web Content Mining in Kids' based Mobile Search

Neha Gupta
Research Scholar, FET
Manav Rachna International University
Faridabad, Haryana, India

Saba Hilal, PhD.
Professor, School of Computer Science
Lingayas' University
Faridabad, Haryana, India

ABSTRACT

Kids are using web more frequently on mobile devices. The need to have mobile compatible web pages and mining the content of the web for these mobile devices is also necessary. Many researchers are focusing the issue. We have tried to elaborate this issue by conducting a detail study on the kids of various schools of Faridabad, tried to understand kids' requirement, their interest on mobile devices, the difficulties they face, parents and the teachers point of view, the various problem associated with kids usage of mobile and above all the role of web content mining in extracting the content from the web using these mobile devices.

Keywords

Mobile Web, Kids, Security, Information Retrieval

1. WEB AND THE MOBILE DEVICE

Use of Internet for educational purposes has grown to a great extent and handheld devices are being used to access Internet by school children and by their parents [1]. School children and parents also use mobile devices for getting information related to their home assignment. This means the users can access web using a PC at home or at office and can access the same information on their mobile phone while traveling.

Although there are advancements in technical and bandwidth aspects [1], still mobile devices are limited by small screen sizes which limit the amount of information that can be displayed at one time. Mobile browsers display the content on mobile devices using two main transformation methods: direct migration and linear [1]. In direct migration, no transformations are made to the original web page. While in columnar (or linear) approach, page areas are presented one after another in a single column.

The presentation of information available on the website is changed to a long linear list that can easily fits within the small screen constraint of the mobile device. The major advantage of this approach is that horizontal scrolling is not required. [2]

2. INTRODUCTION TO MOBILE WEB

The Mobile Web refers to access to the world wide web, i.e. the use of browser-based Internet services, from a handheld mobile device, such as a Smartphone, a feature phone or a tablet computer, connected to a mobile network or other wireless network [11]. Mobile websites are specially designed for viewing mobile contents on Mobile Phones [12]. Earlier websites were seen only on computers that contained large screens for viewing, and logical computational power &

higher Internet speed. But mobile phones are smaller in screens and less powerful than a computer. The Mobile phones that support Internet connection has also very slower speed. So websites for computers cannot function well on mobile phones, it needs a special structure for mobile phones and for running on mobile phone; there is the need to create special websites for mobile phones [12].

Mobile web browsing is being very popular and widely accepted due to widespread proliferation of iPhones and other mobile appliances with surf-friendly interfaces. As a result mobile websites are being more important to offer a mobile-friendly alternative to current website [12]. The mobile site can be accessed by BlackBerry, iPhones, Treo, Smart phones, WAP phones, iPod Touch and most PDAs with various operating systems and wireless Internet connections.

2.1 Features of Mobile Web

The Mobile Web refers to browser-based web services such as the World Wide Web, WAP and i-Mode (Japan) using a mobile device such as a cell phone, PDA, or other portable gadget connected to a public network. Such access does not require a desktop computer, or a fixed landline connection [3]. Mobile web is part of www and carry almost the same features as the web contains like rich user experience, user participation, dynamic content, metadata, web standards, scalability, openness, freedom and collective intelligence by way of user participation. [4]. Mobile web picks the pages of www and publish the data on mobile screen. Mobile Web access today still suffers from interoperability and usability problems. This is partly due to the incompatibility of the format of much of the information available on the Internet with mobile devices and partly due to the small physical size of the screens of mobile devices and other device limitations. [5]

2.2 Importance of Information Retrieval (IR) on Mobile

Retrieving the information on mobile is very important because of the following factors:-
   1. Flexibility of access anywhere anytime
   2. Personal
   3. Portable
   4. Always connected
   5. Increasingly multi-functional beyond the original purpose of voice communications.

In addition to these factors, the advantages of mobile devices will increasingly include:
   1. Location awareness
   2. One-handed operation
   3. Always on

4. Universal alerting device

By way of illustration of some of these factors: the Web can go where you go. You do not have to remember to do something on the Web when you get back to your computer. You can do it immediately, within the context that made you want to use the Web in the first place. [6]

Moreover, with mobile devices appearing in all shapes and forms [13], and with a growing variety of features like location technology, cameras, voice recognition, touch screens etc, the Web can reach a much wider audience, and at all times in all situations. It has the opportunity to reach into places where wires cannot go, to places previously unthinkable [13] (e.g. providing medical info to mountain rescue scenes) and to accompany everyone as easily as they carry the time on their wristwatches.

Finally, today, many more people have access to mobile devices than access to a desktop computer [13]. This is likely to be very significant in developing countries, where Web-capable mobile devices may play as similar a role in deploying wide-spread Web access as the mobile phone has played for providing "plain old telephone service"[13].

2.3 Difficulties of IR on Mobile

World Wide Web is a popular and interactive medium to publish information today. The web is huge, varied and dynamic and thus raises the scalability, multimedia and temporal issues respectively. Due to those situations we are currently drowning in information and facing information overload. [7] Following problems arises when retrieving the information from the web for mobile devices:

1. Volume of data is very large
2. Fast rate of change and growth of data
3. Redundant data
4. Small screen size
5. On most mobile devices, the speed of service is very slow, often slower than dial-up Internet access
6. The access and bandwidth charges levied by cell phone networks are much higher than those for fixed-line internet access
7. Most devices do not support client-side scripting and storage of cookies (Smartphone's and iPhones excluded), which are now widely used in most Web sites for enhancing user experience, facilitating the validation of data entered by the page visitor, etc. This also results in web analytics tools (like Google Analytics) not being suitable for uniquely identifying visitors using mobile devices.[5]

2.4 Relation between Web Mining and Mobile Web

Most of the HTML web pages are not supported by internet enabled mobile handheld devices because the web pages may not be properly and speedily displayed on the micro browsers of the devices due to low memory capacity, small screen size, limited computing power ,narrow network bandwidth, & resources etc. Web usage mining a branch of web mining can be helpful in summarizing the web pages for these devices. Web usage mining helps in data gathering, navigation pattern discovery, pattern analysis etc & hence helps in improving the readability and download speed of mobile web pages. [10]

Web mining is important as it helps us in [10]: -
1. Finding relevant information
2. Creating new knowledge out of information available on web
3. Personalization of information.
4. Learning about consumers or individual users.

Web Mining and Personalization requires modeling of an unknown number of overlapping sets in the presence of significant noise and outliers, (i. e., bad exemplars). Moreover, the data sets in Web Mining are extremely large. [8]

The essential feature that led to the explosive growth of the web is - decentralized   content publishing with essentially no central control of authorship. Because publishing was now open to tens of millions, web pages exhibited heterogeneity at a daunting scale, in many crucial aspects. [9]. Mobile web uses the essential features of web for its working.

The mobile web consists of web pages that are designed specifically for display on mobile devices. Due to their limited capabilities (relative to standard computers), mobile devices access and render web content using the specialized Wireless Application Protocol (WAP). [3]. WAP gateways are used which convert the HTML pages to WML pages and vice versa. Basically Mobile web uses the content of WWW using various web-mining techniques so that web data will be displayed in a suitable form on mobile device. Now a days researches are going on to convert the HTML based web pages to WML based mobile pages or to develop a separate web for mobile itself.

2.5 Mobile as a handy tool for searching the web

As per the features discussed above for the mobile devices, it can be very clearly depicted that mobile is a handy tool for searching the web. The reason for the statement is:
1. People usually carry their mobiles with themselves.
2. Easy to operate, friendly and easily understandable.
3. Flexibility of access anywhere anytime
4. Personal
5. Portable
6. Always connected
7. Increasingly multi-functional beyond the original purpose of voice communications.
8. Screen size is also increasing day by day.
9. Touch applications and android technology have further simplified the process of web access.

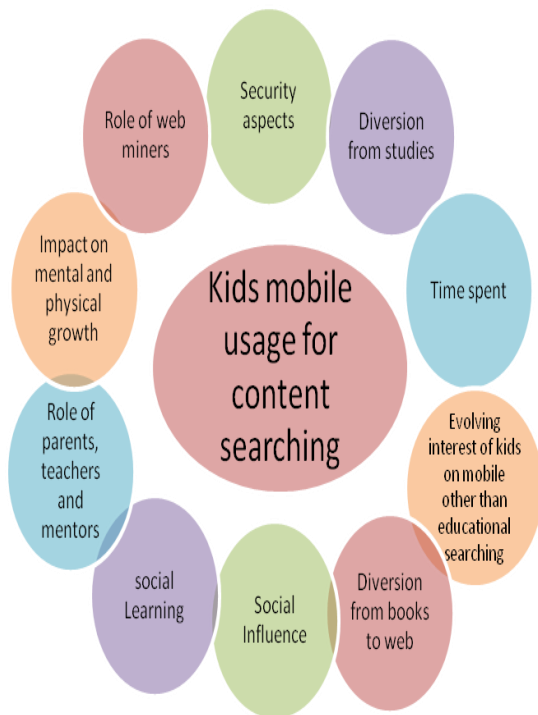## 3. KIDS' MOBILE USAGE FOR CONTENT SEARCHING

Figure 1. Kids' mobile usage for content searching

The above figure touches various aspects of usage of web by the kids using mobile devices. All of these issues and their impacts are discussed below under various sub-headings. Although usage of mobile device to access the web has various advantages but the risks involved are much greater and our aim is to highlight both sides of the coin.

3.1 Evolving interest of Kids on Mobile

As kids are the active user of the web, so a questionnaire was filled by the kids of Sector 30, Faridabad, who study in various schools to know, what are their areas of interest, when they search the web using the mobile devices? According to the answers obtained, following the major areas of interest among the kids:

1. Accessing social networking websites like Face book, Orkut etc.
2. Downloading or playing online games
3. Downloading ringtones, songs or movies
4. Searching cartoons on YouTube
5. Searching latest bikes, cars
6. Places to hangout
7. Latest mobile phones
8. Fashion trends and accessories
9. MTV Roadies/ Big Boss contestant information, their lifestyle
10. Doing assignments

4. FACTORS THAT NEED ATTENTION WHILE KIDS ARE SEARCHING CONTENT USING MOBILE

As discussed previously, kids are an active user of the web and the most frequently used medium for accessing the web is the mobile device. Although accessing the web via the mobile device will lead to many advantages but it will have certain concerns as well. The various factors that need attention when kids are searching the web are:
1. Security threats
2. Impact of usage
3. Time spent

4.1 Security Threats

Because of no rule and guideline on the data that is placed on the Internet, data may be correct or incorrect. While searching a topic on web the search engines on the Internet provide links to many web pages. Some web pages are relevant and some are not related to the topic. So it is very easy for the innocent kids to jump to a page which may contain data that could be harmful to them.

Some of the security threats are:
A. Access inappropriate information
B. Sexually explicit material,
C. Violent and disturbing images,
D. Buying an selling goods or wagering (A credit card number is enough to do all the bad things)
E. Incorrect and inaccurate information (Because of no controlling authority, anyone can upload any type of data on internet and that may be correct or incorrect.)
F. Making friends innocently without verifying their intentions
While chatting, child may form friendship with a stranger, who may have wrong intensions. Because of their credulity they may pass their personal information to the, who may use it for wrong purposes.
G. Full prey to cyber bullying
H. Unsolicited advertising& spam pressures
I. Excessive use may lead to various health problems
J. Putting privacy in danger by giving personal details.
K. Apart from these perils other major risks include:
--- Internet Hoaxes:
--- Personal Risks: These include frauds, cyber crime, pornography, etc.
--- Viruses.
--- Spam Mails

4.2 Impact of Usage

The popularity of internet is growing at a very fast pace and the most frequent user of web are kids. As the content on the web is growing exponentially, kids are also spending more time on searching and surfing the web. Although using the web can be beneficial but excessive use of web can cause a lot of concern. It has been observed that kids are keener to play games on mobiles and laptops then outdoor. The major concerns for excessive use of the web can be:
1. Diversion from studies
2. Social Influence
3. Social learning
4. Diversion from books to web
5. Impact on mental and physical growth.

4.3 Time Spent

According to online survey conducted by Nielson Company New York," TIME SPENT ONLINE AMONG KIDS INCREASES 63 PERCENT IN THE LAST FIVE YEARS". In May 2009, children aged 2-11 comprised nearly 16 million, or 9.5 percent, of the active online universe • Since 2004, the number of kids online has increased 18 percent, as compared

to 10 percent for the total active universe, with a fairly even split between boys and girls. The growth of children online outpaces the overall growth of children in the U.S., where kids under 14 are projected decrease by 1 percent from 2004 to 2010 (according to the U.S. Census Bureau, from 7/04 – 7/10 projections)

5. AVAILABLE SOLUTIONS TO THE THREATS

As discussed above, there are many risks involved while kids are searching the content through the web. Although 100% security cannot be guaranteed but kids can be protected using certain safety measures. They are:

1. Parents and mentors can play a major role by guiding and counsel them about the use of web.
2. Kids based web content miners can be implemented to protect the kids and filter out the content.
3. Giving access to web via mobile by implementing educational miners on mobile devices.

The above mentioned points are discussed below.

5.1 Role of parents and mentors

Child's security while accessing internet starts at home. The Primary objective of security is achieved if the parents take proper control of the material accesses by their child. There is no doubt that internet helps kids in improving their reading skills by providing interesting materials to read, helps them in sending e- mails to their friends, discover about cultures, access libraries and reference materials for school or homework. There are various other benefits as well but while accessing these materials, they may accidentally jump to a site which may contain porn or indecent data.

5.2 Role of kids based Web Content Miners

There are n numbers of web content miners available as freeware that help to ensure child security over internet. These miners are classified according the purpose they serve. For example if I need to block certain sites or chat rooms I need a child lock , if I want to monitor my child activity while he was online , I need a monitoring software etc. on the basis of requirement these miners are classified into 4 categories :-

- Educational browsers/ Tailored browsers
- Child locks
- Filtering software
- Monitoring Software

5.3 Educational Browsers/Tailored Browsers

These are the browsers which can filter the websites and IM for safe downloading and uploading of data. These browsers mainly have the web filters and content filters attached to them so that child gets a secure web for their school assignments.
Various other educational browsers are Kid coaster, kidzui.com, Glubble, Buddy Browser, webkinz etc.

5.4 Child Locks

As the child grows, the desire to explore the World Wide Web grows. Spending too much time at the computer can be quite harmful for his health as well as harms his mental and physical growth. Child locks helps you protects the child from

this danger by restricting the time that the child spends at PC. Child locks let your children play with the computer for a limited time a day. After the time is over, the computer will be locked. Some child locks can even helping locking USB drives, CD-ROMs and floppy disks etc.
Child lock can even help in defining the log time of the user. It helps in limiting the PC usage time of their child.

5.5 Filtering Soft wares

Internet filter software is the additional layers of defense, giving parents another medium to control their kid's activities and monitor them. Internet filter software helps to control the content that is being displayed and also let parents set up passwords. Internet Filtering software provide various powerful tools like filtering the email, blocking popup and chat room monitoring, each designed to protect against and neutralize the strategy of aggressive online porn companies. [1]

5.6 Monitoring Software

These are software which helps the parents in monitoring their child activities while they are online. The monitoring software should keep track of all the activity that occur on your computer, by whom, and when. Some of the features that monitoring software provide are:-
Password capturing and keystroke Recording
Chat monitoring
URL monitoring

1. Screen shot monitoring
2. Online search monitoring
3. Website visited monitoring etc
4. Email sending and receiving
5. File transferred
6. Invisible mode monitoring

Although all the above features are more or less available in various paid software available but various free software are also available.

6. MOBILE BASED EDUCATIONAL WEB MINERS

There are various mobile based web content miners which are used to protect the kids online. The various MBEWCM are:

1. BSecure online for Android
2. Net Nanny 2.0 for Android
3. AVG Family safety for iPhone
4. K9 web protection browser for Android
5. Cell safety for Android and iPhone
6. eBlaster Mobile for Android
7. Mobichip for Android
8. McAfee Safe eyes for iPone
9. Halascape 2.0 for Android and iPhone
10. NQ Family for Android and iPhone

We have listed some of the famous and the most efficient Web content miners for Mobile devices. A comparative study of the above listed along with many other are discussed in the next section.

6.1 Comparative study of various EWCMs

The table below demonstrates the comparison of various popularly used EWCM which work either on Phone, or on desktops.

Name of the Miner	Filtering Algorithm used	Filtering capabilities	Reporting Capabilities	Support Options	Supported Configuration
Net Nanny	URL Based, Keyword Based, Dynamic Categorization	Excellent	Alert by Email, Alert by Text, Remote Reporting	E-Mail, Telephone	Window 7, Window Vista, Window XP, MacOS, Android
Mc-Cafee Safe Eyes	URL Based, Keyword Based, Dynamic Categorization	Excellent	Alert by Email, Alert by Text, Remote Reporting	E-Mail, Telephone	Window 7, Window Vista, Window XP, Window 8, MacOS, iOS
Cyber Sitter	URL Based, Keyword Based, Dynamic Categorization	Very Good	Alert by Email, Remote Reporting	Telephone	Window 7, Window Vista, Window XP,
Cyber Patrol	URL Based, Keyword Based, Dynamic Categorization	Very Good	None	Telephone	Window 7, Window Vista, Window XP,
Profile parental filter	URL Based, Keyword Based, Dynamic Categorization	Very Good	Log reports by Email, Remote Reporting	Telephone	Window 7, Window Vista, Window XP,
imView	URL Based, Keyword Based, Dynamic Categorization	Very Good	Alert by Email, Remote Reporting	Telephone	Window 7, Window Vista, Window XP,
Pure Sight	URL Based, Keyword Based, Dynamic Categorization	Very Good	Alert by Email, Remote Reporting	Telephone	Window 7, Window Vista, Window XP,
Kidswatch	URL Based, Keyword Based, Dynamic Categorization	Excellent	Alert by Email, Remote Reporting, Log report by Email	E-Mail, Telephone	Window 7, Window Vista, Window XP,
inetprotector	URL Based, Keyword Based, Dynamic Categorization	Good	Remote Reporting	Telephone	Window 7, Windows XP
Mobichip	URL Based, Keyword Based, Dynamic Categorization	Very Good	Alert by Email	Telephone	Android
eBlaster Mobile	URL Based, Keyword Based, Dynamic Categorization	Very Good	Remote Reporting	Telephone	Android
Cell Safety	URL Based, Keyword Based,	Excellent	Alert by Email	Telephone	Android, iOS
AVG family safety	URL Based, Keyword Based, Dynamic Categorization	Excellent	Alert by Email, Alert by Text, Remote Reporting	E-Mail, Telephone	Window 7, Windows XP
K9 web Protection	URL Based, Keyword Based, Dynamic Categorization	Excellent	Alert by Email, Alert by Text, Remote Reporting	E-Mail, Telephone	Window 7, Window 8, Android

6.2 Limitations of EWCM

Although EWCM are designed to provide the secure environment to the web, apart from this various EWCM are lacking in certain areas. These are:

1. Problem of under blocking and over blocking
2. Age Restriction/ Blocking according to age
3. Authorization and authentication
4. Providing quality data
5. Real time protection
6. Automatic updating of malicious websites
7. Continuous Protection and speed of computer
8. Monitoring features
9. Filtering features
10. Login time and Logout time
11. Tailored browsers

7. REFERENCES

[1] Neha Gupta, Saba Hilal, "Extraction of Web Content to Adapt Web Pages for Mobile Devices", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.

[2] User Vision, "Usable web pages for mobile devices", http://www.uservision.co.uk/resources/articles/2005/usable-webpages-Mobile-devices/, 2nd March 2005.

[3] Banerjee, Syagnik & Dholakia, Ruby Roy, "Does Location Based Advertising Work?" (International Journal of Mobile Marketing, Dec, (2008).

[4] Best, D., Web 2.0 Next Big Things or Next Big Internet Bubble? Lecture Web Information Systems. Techni sche Universiteit Eindhoven., 2006

[5] Wikimedia Foundation, Inc., "Mobile Web", http://en.wikipedia.org/wiki/Mobile_Web #cite_ref-0 , 2009

[6] W3C, "Mobile Web Best Practices 1.0", W3C Recommendation, http://www.w3.org/TR/mobile-bp/, 29 July 2008

[7] Masand and M spilipoulou, Webkdd 99, "Workshop on usage analysis & user Profiling", SIGKDD explorations 1(2000).

[8] Anupam Joshi, "Web Mining", http://www.cs.umbc.edu/~ajoshi/web-mine/, 15 May 2005

[9] 2008 Cambridge University Press, "Web characteristics", http://nlp.stanford.edu/IR-book/html/htmledition/web-characteristics-1.html, 2009-04-07.

[10] Raymond Kosala, Hendrik Blockeel, "Web Mining Research: A Survey", SIGKDD Explorations, Vol 2 Issue 1, July 2000, pp 1-15

[11] Wikipedia, "Mobile web", http://en.wikipedia.org/wiki/Mobile_Web, Retrieved on 15th Sep 2009

[12] Rose India Technologies, "What is Mobile Website", http://www.roseindia.net/services/webdesigning/corporatewebsitedesign/What-is-Mobile-Website.shtml, Retreived on 15th Sep 2009

[13] W3C, "Mobile Web Best Practices 1.0", http://www.w3.org/TR/mobile-bp/, W3C Recommendation 29 July 2008C. Chang and S. Lui, " IEPAD: Information extraction based on pattern discovery" , In Proc. of 2001 Intl. World Wide Web Conf., pages 681–688, 2001

AUTHOR'S PROFILE

Prof. Saba Hilal is currently working as Director in GNIT – MCA Institute, Gr. Noida. She has done Ph.D (Area: Web Content Mining) from Jamia Millia Islamia, New Delhi, and has more than 16 years of working experience in industry, education, research and training. She is actively involved in research guidance/ research projects/ research collaborations with Institutes/ Industries. She has more than 60 publications/ presentations and her work is listed in DBLP, IEEE Explore, Highbeam, Bibsonomy, Booksie, onestopwriteshop, writers-network etc. She has authored a number of books and has participated in various International and National Conferences and Educational Tours to USA and China. More details about her can be found at http://sabahilal.blogspot.com/

Ms. Neha Gupta is presently working with Manav Rachna International University, Faridabad as Assistant Professor and is Pursuing PhD from MRIU. She has completed MCA from MDU, 2006 and Phil from CDLU, 2009. She has 6 years of work experience in education and industry. She is faculty coordinator and has received many appreciation letters for excellent teaching.