Extraction of Web Content to Adapt Web Pages for Mobile Devices

Neha Gupta[1], Saba Hilal[2]

**[1] Department of Computer Application, Manav Rachna International University,
Faridabad, Haryana 121001, India**

**[2] Department of Computer Application, GNIT Group of Institutions,
Greater Noida, Uttar Pradesh, India**

Abstract

Now a day's mobile phones are replacing conventional PCs' as users are browsing and searching the Internet via their mobile handsets. Web based services and information can be accessed from any location with the help of these Mobile devices such as mobile phones, Personal Digital Assistants (PDA) with relative ease. To access the educational data on mobile devices, web page adaptation is needed, keeping in mind security and quality of data. Various researchers are working on adaptation techniques. Educational web miner aims to develop an interface for kids to use mobile devices in a secure way. This paper presents a framework for adapting the web pages as part of educational web miner so that educational data can be accessed accurately, securely and concisely. The present paper is a part of the project whose aim is to develop an interface for kids, so that they can access the current knowledge bases from mobile devices in a secure way and to get accurate and concise information at ease. The related studies for adaptation technique are also presented in this paper.

1. Introduction

The name says it all - a web page is a page on the web. [3]. A web page is a digital document or a computer file that can be viewed using a web browser program. World Wide Web is nothing but the interconnected network of these web pages which are the part of web sites. Hence, web pages are considered as the building blocks of the web. [3]

A. Existing data formats Supported by various web pages

Mostly a web page has only text and HTML code. The Multimedia files like audio, video, flash animation are included into a web page using various HTML tags. Web browser read the web page, parses it and display it to the user based on the instructions in HTML tags.

Web pages with their multimedia content make a web site. A webpage acts as information set and contains numerous types of information, which can be seen, heard or interacts by the end user. [1] There are varieties of data format available on web pages. Few of them are:-

- Textual information: with diverse render variations.
- Non-textual information.

Non-textual information can further be classified into many categories such as:-

1. Static images or raster graphics: - Format supported is GIF, JPEG or PNG; or vector formats as SVG or Flash.

2. Audio: - MIDI or WAV formats or Java applets.

3. Animated images: - Animated GIF and SVG format, but also may be Flash, Shockwave, or Java applet.

4. Video: - Formats supported are WMV (Windows), RM (Real Media), FLV (Flash Video), MPG, MOV (QuickTime)

5. Interactive illustrations: ranging from "click to play" image to games, Flash, Java applets, SVG, or Shockwave, Interactive text via DHTML

B. Web Pages on mobile devices

Use of Internet for educational purposes has grown to a great extent and handheld .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. School children and parents also use mobile devices for getting information related to their home assignment. 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, 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. 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]

C. Use of Web mining on mobile devices

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. [4]

2. KEWCM (Kids' Educational Web Content Miner)

The System Known as KEWCM i.e. Kids' Educational Web Content Miner) is being designed to provide the students specially the kids, the secure educational content from the web based material and helps in the improvement of quality and reliability of educational content and above all extract data from the current state of knowledge bases of WWW and display it on mobile devices. Some modules of this project are:-
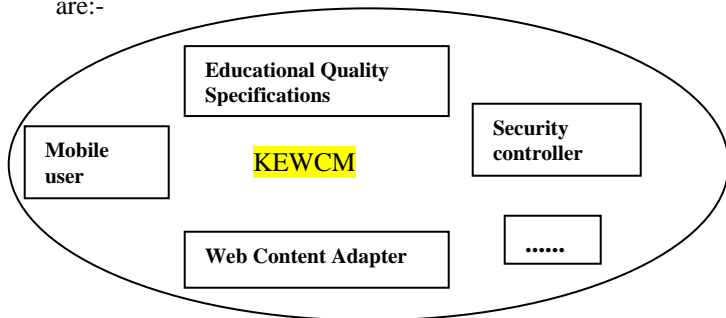

Figure 1

The study presented below is only a part of this project where we are only focusing on the mobile web adapter module and discussed various tools and techniques that can be used to develop this particular module. A framework has been designed for this module, which can be used as foundation stone for the development of this particular module.

3. Web Content Adaptation

Adaptation is a process in which the web pages are delivered to the mobile device according to the specifications of that device, considering various factors like Screen size, Language support (e.g. WML, XHTML, cHTML etc), Browser support etc.

The need of adaptation arises only because of the wide variety of Mobile browsers available on various devices.

4. Process and Modes of Adaptation

Adaptation process follows two simple steps:-
1. Content is analyzed according to the specification of devices
2. Content is transformed according to analysis made.

Adaptation can be done with either of the following modes but in current scenario automatic adaptation procedure is the most convenient method to be used.
   1. Designing of Alternative CSS
   2. Developing multiple versions of web pages
   3. Automatic content adaptation using adaptation engines

Designing alternative CSS helps in controlling the display of elements and images. By this method one can detect the device and the appropriate CSS file can be linked.

A second method creates multiple versions of the same web page and links the pages to the devices according to their specification.

Third method is the most widely used method and convert the content according to device specific version. It is most popular as it allows creating content in one format only. In the present study we are focusing only on the automatic content adaptation process and not on the above two methods.

5. Types of Adaptation

Web adaptation is broadly classified into three broad categories on the basis of the place where adaptation takes place.

1. Client Side
2. Server Side
3. Proxy based / Intermediate adaptation

In Client side adaptation, the require tools and algorithms to convert the content to user need takes place at client device i.e. Mobile Device. Client side adaptation give flexibility to designers to design the web page only once as same content is delivered to every device.

In server side adaptation, the adaptation process is completed at server thus reducing the computational time and the adapted page is send to the client via network. In this mode, several versions of the same web page are stored at the server and the page which meets user need is delivered according to preferences.

In Proxy based approach, a proxy server analyzes and transcodes the content on-the-fly, before sending the result to the client. [11]

6. Related Studies

An important paper [15] has focused on P2P Collaborative deployment scheme which works by dividing the web pages into small logical blocks, so as to display the existing web pages using web adaptation engines. Another paper [16] has proposed a model for adaptation engines with seventeen criteria for evaluating a web adaptation engine. The author of [17] has discussed the issues related to the differences between single-user browsing and co-browsing and proposed a content adaptation framework based on the concept of shared viewpoint and personal viewpoint. The paper [18] has referred the thumbnail view concept, VIPS method and AJAX, and proposed a dynamic Web page adaptation for mobile device.

7. Web Adaptation Tool features

1. Adaptation for almost each and every device.
2. Automatic content adaptation
3. Filtering of data
4. Should minimize bandwidth requirement.
5. Ease of navigation across pages
6. Automatic site awareness.
7. Automatic fault management and recovery.

8. Optimized access to flash and multimedia content
1. 9. Cost effective

8. Architectural view of Automatic Adaptation process of a web page on mobile devices using Proxy Based Adaptation Method
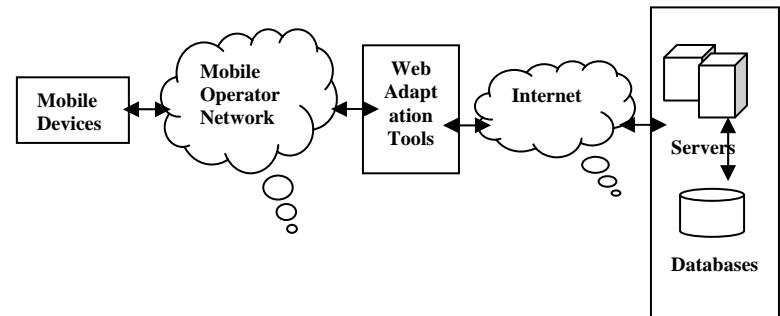

Figure 2

In the above figure, whenever a mobile device is trying to access a web page from the internet, adaptation is needed. Adaptation is achieved with the help of various FOSS (free and open source softwares) available on web. As we are focusing on developing a mobile web adapter for educational data, so security reliability and quality shall also be ensured. Keeping in mind all the above factors, we have tried to design a framework for kids' mobile adapter so that secure, easy and reliable environment can be provided to them.

The adaptation framework shown below has three steps:-
1. Web Adaptation Engines/Server
2. Client profile determining softwares
3. Processors

The working of these components is elaborated with the help of following diagram.
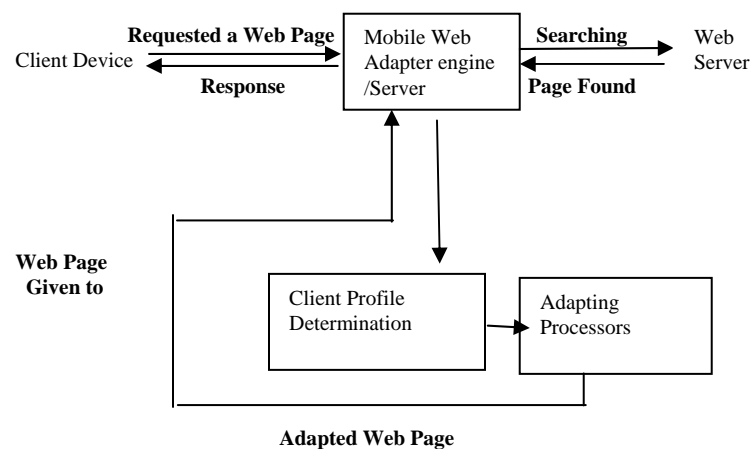
9. Proposed Framework

Adapted Web Page

Figure 3

A. Mobile Web Adapter Engines/Servers

The purpose of these servers are to accept the request from the client device and search that particular page on various web server, if found, pass that page to processor for adapting the content according to client's Profile. Most of these engines have built in processors and client profile determination softwares in it. While the engine passes the request to the web server for searching the requested page, the engine itself uses Client profile determination software to determine the device capabilities etc. Varieties of these engines are available as free and open source softwares. Few of them are:-

- Info Gin's Intelligent Mobile Platform [5]
- Byte mobile's Web Fidelity Suite [6]
- Novarra's Vision Server + Native Browser [7]etc.

B. Client Profile determination Softwares

As mentioned earlier, adaptation is needed because of the variety of browsers supported by the client devices. So determining the client profile before adaptation is most important step in adaptation process. The factors that must be determined before Adaptation is: -

1. Device Specification (Memory, Processor, Display statistics)
2. Bandwidth availability
3. Preference of user (Font, Images, Navigation etc)

Device Specification is a process in which all the mobile device related information is captured and stored in a file and the adaptation is done according to the information stored in file. This can be achieved with the help of various Free and open source softwares like

- WURFL [8]
- WALL [9]
- UAProf [10] etc.

C. Adaptation Processors

The final step requires a processor for selecting, modifying and generating the content so that the device specific result can be formed. Now days these processors are inbuilt and integrated part of web adaptation engines so separate processors may or may not be required. Examples of such Processors are: -

- Cocoon (Apache) [12]
- WEMP (IBM) [13]
- MyMobileWeb (Morfeo) [14].

Other tools available for Images and animation are:-
- GAIA Image Transcodes
- PHP Image Rendering Library

- Image Server
- HAWHAW

10. Conclusion

Web page adaptation is needed for providing the retrieved web content via the mobile interface. The Study of KEWCM is not yet completes therefore the implementation details for the proposed adaptation framework is not provided as part of this paper. However, this paper has put forward important details on web content adaptation with an architectural view of automatic adaptation process of web pages on mobile device using proxy based adaptation method. The designed framework includes mobile web adapter engine with client profile determination and adapting modules.

References

[1] Wikipedia the free encyclopedia, "Web page", http://en.wikipedia.org/wiki/Web_page#Elements_of_a_webpage , 12 July 2010
[2] User Vision, "Usable web pages for mobile devices", http://www.uservision.co.uk/resources/articles/2005/usable-web-pages-mobile-devices/, 2nd March 2005
[3] Web Developers Notes, "What is a web page?" http://www.webdevelopersnotes.com/basics/what-is-a-web-page.php, Retrieved on 28 July 2010

[4] Wen-Chen Hu, Jyh-Haw Yen, Hung-Ju Chu, "Adaptive web browsing using web mining technologies for Internet-Enabled Mobile hand held Devices" , http://www.igi-global.com/bookstore/TitleDetails.aspx?TitleId=653&DetailsType=Preface, 2006
[5] InfoGin, "Intelligent Mobile Platform", http://www.infogin.com/solutions.asp, retrieved on 15 June 2010
[6] Byte Mobile, "Web Fidelity suite", http://www.bytemobile.com/products-applications/wfs-widget-bar-overview.html, retrieved on 17 June 2010
[7] Novarra, "Vision Server + Native Browser", http://www.novarra.com/solutions/open-internet-services-platform/, retrieved on 18 June 2010
[8] Luca Passani, "WURFL- Wireless Universal Resource File", http://wurfl.sourceforge.net/, retrieved on 23 June 2010
[9] Luca Passani, "WALL - the Wireless Abstraction Library", http://wurfl.sourceforge.net/java/wall.php, retrieved on 24 June 2010
[10] Mark Butler, Steve Geach, Rainer Hillebrand, "UAProf profile repository", http://w3development.de/rdf/uaprof_repository/, retrieved on 25 June 2010
[11] TeliaSonera, "Web content Adaptation", http://www.medialab.sonera.fi/workspace/WebContentAdaptationWP.pdf, Aug-2004
[12] Apache, "Cocoon", http://cocoon.apache.org/, retrieved on 15 June 2010

[13] IBM, "Web sphere everyplace Mobile Portal Enable", http://www-01.ibm.com/software/pervasive/ws_everyplace_mobile_portal_enable/, retrieved on 15 June 2010

[14] Morfeo, "MyMobileWeb", http://mymobileweb.morfeo-project.org/, retrieved on 15 June 2010

[15] Yunpeng Xiao Yang Tao Qian Li Wenji Li, "Web Page Adaptation for Small Screen Mobile Device: A New P2P Collaborative Deployment Approach" , Published in ICINIS '08, Vol 1, 21 November 2008

[16] Anastasios A. Economides, "On evaluation of adaptation engines", Proceedings of the 10th IASTED International Conference on Computers and Advanced Technology in Education Pages: 102-106, 2007

[17] H. N. Chua, S. D. Scott, Y. W. Choi, P. Blanch field, "Web-Page Adaptation Framework for PC & Mobile Device Collaboration," AINA, vol. 2, pp.727-732, 2005

[18] Yunpeng Xiao Yang Tao Wenji Li, "A Dynamic Web Page Adaptation for Mobile Device Based on Web2.0", ASEA, vol 13, pp. 119-122, 2008

BIOGRAPHICAL SKETCH OF AUTHORS

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.