Web Page Adaptation for Large Screens

Neetu Narwal
Asst. Prof., Maharaja Surajmal Institute
Affiliate College of GGSIP University
New Delhi

Dr. Saba Hilal
Research Supervisor,HOD,
Department of Computer Applications,
Lingaya's University,Faridabad

Abstract—**With advancement in the technology numerous display devices are available ranging from small screen i.e., palm topm, androids, mobile phone etc., to the large screen devices such as 1200x1800, LCD display etc. In the small screen dimension lots of research had already been conducted but other end has received very little focus. We propose a methodology to adapt the web page on large screen context. We would be utilizing the native web technologies i.e., HTML5 and CSS3 to adapt the web page to large screen context and vision based methodology to extract the content of the web page**

Index Terms—**Web Page Segmentation, HTML5, CSS3.**

I. INTRODUCTION

With the advancement in the technology in the last decade the use of Internet on the small scale device has gained momentum. There is a lot of research directed towards these category of devices i.e., small screen size, less bandwidth, limited processing power and restricted memory etc. However, recent years have also witnessed a drastic technological advancement in the large display devices and they have come within the reach of individuals. But the websites available on Internet do not utilize the large screen space available in these large screen devices.

The developers have produced several variations of the same web site to adapt to specific device category mainly desktops and smart phones. However, there is no remarkable advancement in the large display device category, still most of the websites do not utilize the available large screen space. Most of the websites are designed using fixed size position as the developer can't predefine the look of web site on different Web Browser so the use of fixed size positioning give them opportunity to provide consistent look to the web page across various browsers. This leads to wastage of space when these websites are accessed in large display devices.

II. PREVIOUS WORK

A. Research done in the area of Web Page Segmentation

Web data extraction system is a sequence of steps that extracts the content of the web page. There is lot of work done in the area of extraction of content from web pages.

There are several algorithms used for Web Page Segmentation, the most popular of which are
- DOM-based algorithms
- Layout-based algorithms
- Vision Based algorithms

DOM based web extraction method, make use of the Document Object Model [14] described by W3 Consortium. DOM Document is a collection of nodes arranged in a hierarchy, which allows a developer to navigate through the tree to extract the information.

Some approaches make used of HTML tags to divide the web page content [9,10] and some other rely on information like content [11] and link [12].

As this method is dependent on HTML tags and it has been observed that Web Designer does not follow very strict grammar rules for designing the web page so the DOM based method normally does not give accurate results.

The layout-based segmentation method uses layout information of the web page to relate the similar content blocks based on location, shapes and other parameters. Hiroyuki Sano et.al. [13] make use of layout template of a Web page for segmentation. Layout template is useful in terms of speculating where the main content of the page is located.

Cai et.al. [4, 5, 6] use layout information such as "font," "color," and "size" to restructure a Web page in a content block tree.

Cai, et.al. [5] have introduced a vision-based page segmentation (VIPS) algorithm. This algorithm segments a Web-page based on its visual characteristics of the web page, they segment the page based on colors, font, background etc. to identify the coherence of each blocks and identify the implicit and explicit separators to divide the web page into semantic blocks.

Nwe Nwe Hlaing, et.al., [8] used VIPS algorithm to extract the blocks from the web page and then analysed the data regions to extract the data records proposed Vision based Extraction of data Record (VER) algorithm.

B. Some of the research has been done in the area of Web Page adaptation for Large Screen Context

Michael Nebeling et.al.[1], have developed a set of metrics that can be used understand the utilization of space in different widescreen settings. These metrics can be used by developer to

design a web page for large and high resolution screens.

Michael Nebeling[2] et.al, suggested an adaptive layout template that can accommodate a range of viewing situations and because it is based on only native web technologies, can be easily applied to existing web sites. They made use of HTML5 and CSS3 specification such as advanced media queries and multi-column layout for adaptive viewing.

Nobuo et.al.[3], suggested an extension of Web-page layout optimization method for multimodal browsing sizes, it dynamically changes box locations and font sizes by switching CSS files for different browsing sizes, so that the main content can be accessed without the screen scroll operation even at a small sizes whereas the blank space is avoided at a large size.

III. PROPOSED METHODOLOGY

We propose a methodology to adapt the web page according to the large screen display by dynamically repositioning of the elements from the web page to make use of available space. This methodology would avoid redesigning of the websites for different web browsing devices. The plugins can be attached with the web browser for rendering the webpage blocks to adapt themselves according to screen context.

We propose the following steps in adapting the web page and re-authoring for large screen devices context.

Step 1: The first step would be Web Data Extraction, according to literature survey these are broadly categorized as DOM based, Layout based and Vision based. We are using vision based method for our work.

Visual based method is based on visual cues of the web site i.e, color, font, background etc. and also make use of DOM tree for segmentation of the web page. It generates visual block tree as the output.

Vision based method may incorporate three aspects for measuring the distance of elements for grouping them in blocks

i)   Geometric measure
ii)   DOM measure
iii)   Semantic measure

Our proposed method makes use of Vision based method for web page segmentation and using all three measures for calculating the degree of coherence between blocks and then dividing them into visual blocks.
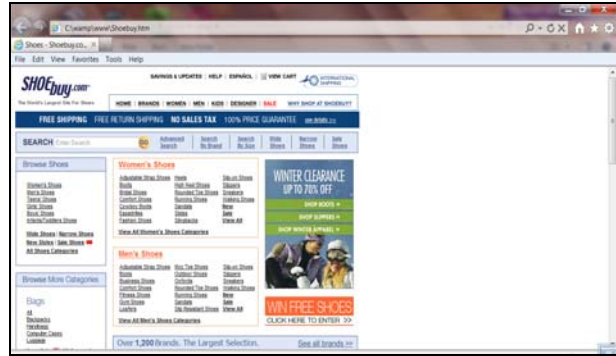

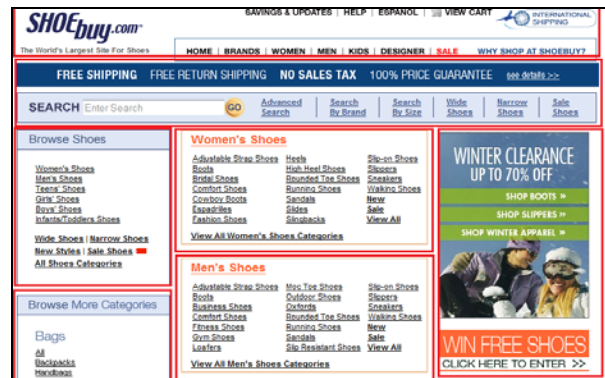Fig 1 : Original Web site displayed in the Web Browser on large screen display

Fig 2 : Visual Blocks Captured from the Web Site
.

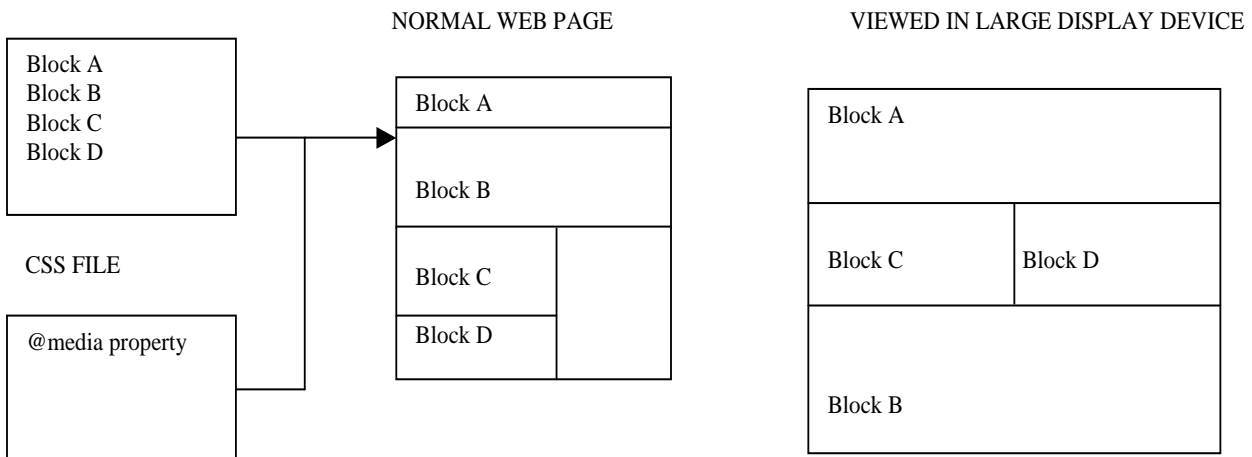HTML FILE   WEB PAGE VIEW

Fig3. Block Ordering and display format changed by accordingly in large display device

Step 2: The next step is to redesign the style sheet so as to accommodate the web page according to the window viewport size and the window size can be obtained by using CSS3 media queries. We will be making use of CSS3 and HTML5 property i.e, multi-column layout, float property, liquid layout etc. for rearranging the blocks of element to utilize the available extra screen space

New style rules will be added to already existing style sheet if no style sheet is found then new .css file will be created and linked to the current web page, the rules will be generated by keeping in mind the semantic similarity of element blocks.

The media style rules, normal flow, the liquid layout and the float property of CSS will be used for adjusting the blocks according to the size of display device

Step 3: After adding the style rules and rearranging the web page element blocks, the web page will be reloaded utilizing the extra screen space of large display device.
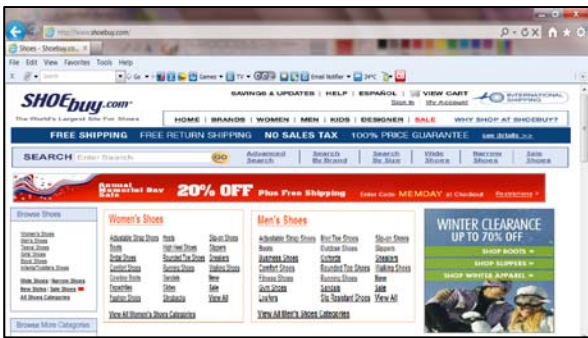

Fig 4 : Web Site after rearranging the visual blocks

CONCULSION

This paper presents a proposed scheme for Web page adaptation for large screen context. The method makes use of vision based techniques for web data extraction and also make use of media queries of CSS3 for modifying the web page content based on the window viewport size. The result has not been presented in this paper as the work is still in progress.

REFERENCES

[1] Michael Nebeling, Fabrice Matulic, Moira C. Norrie , Metrics for the Evaluation of News Site Content Layout in Large-Screen Contexts, , Vancouver, BC, Canada, 2011.

[2] Michael Nebeling, Fabrice Matulic, Lucas Streit, and Moira C. Norrie Adaptive Layout Template for Effective Web Content Presentation in Large-Screen Contexts,

[3] Nobuo Funabiki,Junki Shimizu, Megumi Isogai, Toru Nakanishi , An Extension of the Web-page Layout Optimization Method for Multimodal Browsing Sizes, 13th International Conference on Network-Based Information Systems, 2010

[4] Deng Cai, Shipeng Yu, Ji-Rong Wen and Wei-Ying Ma, Extracting Content Structure for Web Pages based on Visual Representation, , Microsoft Research Asia

[5] Deng Cai, Shipeng Yu, Ji-Rong Wen and Wei-Ying Ma, VIPS: A Vision based Page Segmentation Algorithm, , Microsoft Research Asia, Beijing, China, 2004

[6] Deng Cai1, Xiaofei He, Ji-Rong Wen, Wei-Ying Ma, Block-level Link Analysis, , ACM 2004

[7] Suhit Gupta, Gail Kaiser, David Neistadt, Peter Grimm, DOM-based Content Extraction of HTML Documents, Budapest, Hungary, 2003.

[8] Nwe Nwe Hlaing, Thi Thi Soe Nyunt, An Approach for Extraction Data Record from Web Page based on Visual Features, IJAMS, Aug 2011.

[9] Lin, S.-H. and Ho, J.-M., Discovering Informative Content Blocks from Web Documents, In Proceedings of ACM SIGKDD'02, 2002.

[10] Wilkinson, R., Effective Retrieval of Structured Documents, In Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, 1994, pp. 311-317.

[11] Embley, D. W., Jiang, Y., and Ng, Y.-K., Record-boundary discovery in Web documents, In Proceedings of the 1999 ACM SIGMOD international conference on Management of data, Philadelphia PA, 1999, pp. 467-478

[12] Chakrabarti, S., Joshi, M., and Tawde, V., Enhanced topic distillation using text, markup tags, and hyperlinks, ACM Press, 2001, pp. 208-216.

[13] Hiroyuki Sano, Shun Shiramatsu, Tadachika Ozono, and Toramatsu Shintani, A Web Page Segmentation Method based on Page Layouts and Title Blocks

[14] Web Site Citation : http://www.w3.org/DOM/