Volume: 6 Issue: 4 Pages: 2398-2403 (2015) ISSN: 0975-0290

Mobile Web Browsing Based On Content Preserving With Reduced Cost

Dr.N.Saravanaselvam
Head of CSE Department, Sri Eshwar College of Engineering, Tamil Nadu
Dr.K.Rajangam
Head of EEE Department,SCAD Institute of Technology, Tamil Nadu
Ms.L.Dhanam
Assistant Professor, Department of Computer Science & Engineering, SCAD Institute of Technology, Tamil Nadu
Ms.Anuradha Balasubramanian
Assistant Professor, Department of ECE,Info Institute of Technology, Tamil Nadu

ABSTRACT

Internet has played a drastic change in today's life. Especially, web browsing has become more exclusive in compact devices. This tempts the people to migrate their innovations & skills into an unimaginable world. With these things in mind, it is necessary for us to concentrate more on the techniques that how the web data's are accessed and accounted. Developed countries use a widely popular technique called Flat- rate pricing, which is solely independent on data usage. But whereas, developing countries are still behind the concept of "pay as you use", which leads to high usage bills.With an effort to resolve the problem of high usage bills, we propose a cost effective technique, which reduces the data consumption in web mobile browsing. It reduces the usage bills in the mechanism of usage-based pricing. The key idea of our approach is to leverage the data plan of the user to compute a *cost quota* for each web request and a network middle-box to automatically adapt any web page to the cost quota. Here we use a simple but effective content adaption technique that highly decides which image or data best fits the mobile display with low cost and high quality resolution. It also emphasis on the trendy technique," The Data Mining "which mines the requested & required data. The mined data's are filtered based on the content adaption technique and fit into the display effectively. Interesting and noticeable feature in this concept is that only important web contents requested by the user are exhibited. A feedback process involves in this concept to retrieve the required data alone and also to improve the best fit resolution. With this proposed system web mobile browsing becomes cheaper & contributes an enormous logic for the future project in the field of Mobile browsing.
***Keywords*— content adaptation, cost-aware web browsing, user experience, pricing, web usage, optimization.**

Date Of Submission: October 20, 2014 Date Of Acceptance: December 29, 2014

I. INTRODUCTION

The increasing pervasiveness of web browsing on portable devices calls for the careful consideration of how web pages are accessed by mobile users. While flat-rate pricing (independent of the usage) has been the conventional data pricing model in rich countries, mobile roaming users and users in developing regions are typically subject to usage-based pricing based on cost per megabyte downloaded. With the exceptional growth in both the size and complexity of web pages, usage-based pricing can result in high monthly bills even under nominal web usage [23]. That aims to substantially reduce the usage cost incurred for the end-user due to usage-based pricing. There is a large body of work on web content adaptation and also specifically in the space of content adaptation for mobile our system on popular web pages, we show that our system can significantly reduce the size of downloaded data, and therefore the cost. On average, our system is able to reduce the cost of mobile web browsing by a factor of 2, and our most aggressive level of content adaptation reduces the cost by a factor of 110. We also show that, for each page request, our solution is to provide the web page with the quality of content in minimized cost.

Current systems generate thumbnails by shrinking the original image. This method is simple. However, thumbnails generated this way can be difficult to recognize, especially when the thumbnails are very small. This phenomenon is not unexpected, since shrinking an image causes detailed information to be lost. An intuitive solution is to keep the more informative part of the image and cut less informative regions before shrinking. Some commercial products allow users to manually crop and shrink images. Burton et al. proposed and compared several image simplification methods to enhance the full-size images before sub sampling. They chose edge-detecting smoothing, loss image compression and self-organizing feature map as three different techniques in their work. In quite a different context, DeCarlo and Santella tracked a user's eye movements to determine interesting portions of images, and generated non-photorealistic, painterly images that enhanced the most salient parts of

Volume: 6 Issue: 4 Pages: 2398-2403 (2015) ISSN: 0975-0290

the image. Chen et al. use a visual attention model as a cue to conduct image adaptation for small displays.

II. RELATED WORK

Content adaptation techniques can be classified into two categories, general purpose and content-type-specific. Many current content adaptation techniques are content type- specific, i.e., designed for specific content types, such as those for text, image, and video objects. [4] For example, much research has been done on adaptation of images, such as color reduction, using perceptual hint to keep semantically important blocks, mesh modeling and generation, and other transcoding techniques. For video adaptation, schemes such as adaptive streaming, frame skipping, multidimensional bit-rate reduction, temporal filtering, and spatio-temporal resolution adjustments have been proposed. Info Pyramid has been designed for general multimedia content adaptation. It provides a multimodal, multi resolution representation hierarchy for multimedia content. A transcoding-enabled caching system is proposed for streaming media. [17] Content transcoding is performed at the edge servers to reduce the user-perceived latency and network traffic. For text contents, several summarization techniques, such as keyword extraction and summary sentence selection and compression, have been proposed. All these systems consider only individual types of contents.

Data Volume	Price (Ksh)	Price per MB (Ksh)
50MB	100	2.00
200MB	250	1.25
500MB	499	1.00
1.5GB	999	0.65
3GB	1,999	0.65
4GB	2,499	0.61
8GB	3,999	0.49
20GB	9,999	0.49
30GB	14,999	0.49

TABLE I
PREPAY DATA BUNDLES FOR SAFARICOM, KENYA'S LARGEST MOBILE NETWORK. ALL PLANS COST 8KSH PER MB AFTER THE BUNDLE VOLUME. 83.7KSH = 1USD (DECEMBER 2011)

Some general-purpose content adaptation systems have also been developed. In the Power Browser project, a proxy-based architecture about content adaptation for PDAs has been introduced. The proxy [11] fetches Web pages on the client's behalf and dynamically generates hierarchical views of the fetched pages.

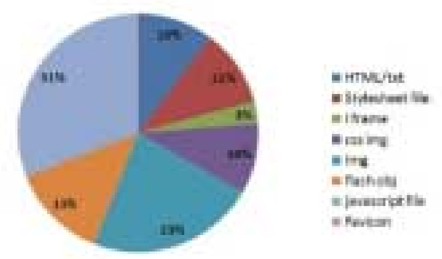

Fig. 1. Web page anatomy of Alexa top 100 pages broken down by bytes.

III.PROPOSED FRAMEWORK

Recent developments towards ubiquitous multimedia trigger the need for mobile users to access and interact with the multimedia content anywhere at any time. The pressing need of improving the user's media experience have resulted in the paradigm shifting from traditional device-centric multimedia to contemporary user-centric multimedia. In this new paradigm, users are placed in the center of ubiquitous multimedia environment and are provided with access to personalized multimedia content intelligently. Regardless of mobile device capabilities, seamless access is desired to maximize the quality of experience for mobile users, [8] based on the media content and user's intention. To facilitate portable usage, mobile devices are usually designed to be light weight and to serve the short viewing distance. On one hand, such a design benefits mobility and allows ubiquitous access of multimedia content; on the other hand, it has been the main reason for the small display size of mobile device, which is in turn the main limiting obstacle to affect user's viewing experience. To provide mobile user with high viewing experience, proper adaptation on high resolution media content is necessary to fit various limited display of mobile devices, heterogeneity of network links, and distinct customer intention of mobile users. To obtain adaptation results which are consistent with the media contents and mobile user intentions, the semantic gap and user intention gap are two critical challenges that need to be properly addressed.

Volume: 6 Issue: 4 Pages: 2398-2403 (2015) ISSN: 0975-0290

Fig. 2. (a) Original page. (b) Level 2 adaptations. Visually, the images are down sampled, and the page retains most of its original formatting.

Text and Image Adaptation

Most existing adaptation approaches utilizing cropping, warping, or seam carving techniques for various displays are completely based on low level features without consideration of real contents in images corresponding to the underlying concepts in the real world. They utilize saliency which can be computed through various low level features such as color, intensity, contrast, orientation, gradient, etc. to catch and preserve the important contents. Although various techniques were developed to detect the salient image parts, such adaptation schemes based solely on low level features may fail to catch some important contents in images. They lack intuitive connection to the semantics of images and user demands. As a result, the regions obtained for display may not be consistent with semantically important regions to users. Some significant content may be dropped or shrunk too much; hence, user viewing experiences are often not satisfactory. According to, humans tend to view and comprehend the images based on *semantics* which refers to the creatures and objects comprising the scene and the relations among the entities. Therefore, semantic based image adaptation is a desired research direction to provide the mobile user better perceptual experiences. To carry out semantic based image adaptation, semantic analysis to extract the key persons or objects and associate them with high level concepts is an indispensible step. Although a good portion of user generated media data have already been tagged with some keywords such as “wedding” and “baseball”, important objects have to be extracted and localized for content selection in adaptation.

We present a design and implementation for displaying and manipulating HTML pages on small handheld devices such as personal digital assistants (PDAs), or cellular phones. We introduce methods [22], [23] for summarizing parts of Web pages and HTML forms. Each Web page is broken into text units that can each be hidden, partially displayed, made fully visible, or summarized. A variety of methods are introduced that summarize the text units. In addition, HTML forms are also summarized by displaying just the text labels that prompt the use for input. We tested the relative performance of the summarization methods by asking human subjects to accomplish single-page information search tasks. We found that the combination of keywords and single-sentence summaries provides significant improvements in access times and number of required pen actions, as compared to other schemes. [5].

Most of the content adaptation work for mobile devices concentrates on content transformation to suit the display of smaller devices. Various content transformation processes like layout change and content format reconfiguration are used in [18] for adapting web content to mobile user agents. The idea of partitioning the web page into blocks (snippets), and processing the information only in the useful blocks by filtering out unnecessary ones is used in [19]. Some content adaptation solutions like those in [26] target only the multimedia content of the web pages. The task of the general parser is to build a structure tree to represent the hierarchical structure embedded in the HTML source, while the content-specific parser extracts specific, content related attributes. The general parser maintains a list of HTML tags. Each tag in the list corresponds to a content-specific parser. Then a tag is identified in the HTML source. If it is not in the list, it is ignored since there are no further adaptation rules for the corresponding content. If the tag is in the list, the general parser creates a new node in the appropriate location in the structure tree. After the creation of the node, the general parser passes the corresponding HTML source between the current tag and the next tag as well as the pointer to the newly created tree node to the corresponding content specific parser.

The content-specific parser extracts the attributes of the object from the HTML source and attaches the information to the tree node. The page facts specific to the object are generated by the content-specific parser. Once the content-specific parser completes its content processing, the control is sent back to the general parser. The general parser continues a similar parsing process on the next HTML tag until the whole structure tree is built. The pointer will be moved up in the hierarchy when an HTML close tag is met. The structure tree is successfully built. The general parser then traverses the tree and generates the page facts about the hierarchical structure of the HTML page, such as parent-child and sibling relationship.

Content Adaptation Ladder

Volume: 6 Issue: 4 Pages: 2398-2403 (2015) ISSN: 0975-0290

Given the anatomy of a web page, we introduce a six-level content adaptation hierarchy for a given web page. The first three levels of the Web page which form the *text-only ladder focus* on textual versions of the page while the final three levels which form the *advanced ladder* focus on adaptation and filtering techniques that preserve the basic structure of the page.

1) Text-only ladder: The three levels of the text-only ladder of a web page are:

Snippet page: In the snippet version of a page, we extract the most important textual snippet of the web page requested by the user. In the preprocessing step, we extract the main body of the web page by removing redundant data, advertisements and split the block-level HTML components into a sequence of paragraphs. To extract the main text snippet, we extend the basic mechanism proposed in [27], which uses a simple averaging mechanism to smooth the word count graph at a paragraph level. We extend this scheme to include the influence of a desired number of neighboring paragraphs. We normalize the word count of every paragraph by convolving the word count function with a Gaussian function with a fixed width, that is, if $\{ni\}^{N}i=1$ is the normalized word count as a function of the paragraph number, we do the following:

$$n_i = \frac{\sum_{k=-p}^{p} n_{i-k}\, e^{-\frac{2k^2}{p^2}}}{\sum_{k=-p}^{p} e^{-\frac{2k^2}{p^2}}}$$

Where p is the range: number of samples to be considered before and after the current sample. A salient feature of the Gaussian smoothing function is that the current sample is given the highest weight and the weight of the neighboring samples decays smoothly with distance. This feature helps in distinguishing the main text with higher probability.

1). Text-only version: In the text-only version, we strip all unessential HTML tags and present a condensed HTML (with a plain CSS) containing the textual content of a page. This is done by repeatedly using the above mentioned procedure of main text snippet extraction to extract the next level of main snippets. We then take the top 5 results of the main text snippet extraction and form a text-only version of the requested web page.

Page summarization (Level 0): At this level, we present a summarization of the web page where we show the important headings with little more information about them. This method filters out everything but the headings and the paragraphs next to each heading. This technique works very well for news web pages or other pages with textual data.

2) Advanced Ladder: The advanced ladder supports three levels as outlined below.

Level 1: HTML, CSS, I frame, Relevant Java scripts, Images in headings

Level 2: Level 1 + Images (compressed, down-sampled), I frame

Level 3: Level 2 + embedded objects

Level 1 extracts and presents only the critical components of the page while Level 2 is a filtered and adapted version of the page (removing ads and unnecessary scripts). Level 3 presents an adapted version of the page with no filtering. While generating these different versions of the pages, we performed a few critical optimizations to enhance load times and reduce the web page size.

HTML rewriting: Minification is the practice of reducing the size of the code thereby improving the load times. Minification usually involves removing white spaces (space, tab, newline) and comments. In the case of JavaScript, this improves response time performance because the size of the downloaded file is reduced. This contributed to over a 5 – 10 KB reduction in web page sizes. Rendering time optimization: Moving the Cascaded Style sheet to the header of the HTML allows progressive rendering. Similarly, moving the JavaScript's [24] to the end of the HTML code enhances user experience since it would enable the light weight components of the web page to be downloaded first.

Ads and favicon filtering: JavaScript's that contain particular keywords are potential signatures of advertisements.[6] To separate scripts that are related to a page and those that are not, we constructed a large list of keywords that we used for identifying scripts related to advertisements. Similarly, we distinguish images related to the web page from those that are not based on the name, location of the images and location of the image pointer in the HTML. We also remove favicons across all adaptation levels which save 2 - 10 KB.

Image and embedded objects filtering: We use the Standard approaches of down-sampling images and removing embedded objects at the appropriate levels of the adaptation ladder which may include multimedia content, like sounds, videos, flash etc., in web pages. The typical size of a flash object in web pages is 85.4 KB which contributes 13% to the average web page size.

Type	50%	75%	90%	95%
Snippet	110	200	300	350
Level 1	3	7	15	20
Level 2	2	5	10	16

TABLE II : CUMULATIVE DISTRIBUTION OF THE SIZE EDUCTION FACTOR FROM THE ORIGINAL PAGE.

Size reduction: To measure the size reduction, we ran our system across the top sites from Alexa for different levels of adaptation. Table II summarizes our results. The table shows the cumulative distribution of the reduction factor page size when using our system for different adaptation levels at 50%, 75%, 90% and 95% of the CDF. We observe that in the median case, Level 1 and Level 2 adaptations provide a size reduction by a factor of 3 and 2 respectively while snippet adaptation can provide a reduction by a factor of 110. The reduction factor of page size increases up to 20, 60, and 350 times for adaptation levels (1, 2, and snippet) at 95% of the CDF.

IV. SYSTEM DESIGN

A. Rule base

The rules in the Rule Base are classified into content analysis rules and content adaptation rules. The content adaptation rules define the manner in which various objects are adapted. The content analysis rules process the raw page facts and derive new facts of higher abstraction.

B. Content Analysis Rules

Inferred content facts can have many different varieties. Width information is very important in content adaptation, but it is not always provided in HTML documents. The width information in HTML documents will be directly utilized when it is available. Otherwise, some rules are designed to derive the width information. Some rules assign a width value to the object based on the analysis of many Web pages. For example, the widths of some commonly used form objects can be set to a fixed value, such as 25 pixels for a check box, or 20 pixels for a radio button. Other width information could be computed from facts. For example, the width of a button can be estimated from the number of characters in its label.

C. Content Adaptation Rules

Content adaptation rules derive adaptation instructions from facts. Each adaptation instruction corresponds to an adaptation procedure that may be implemented in different programming languages or simply using existing content adaptation components. The content adaptation rules are organized according to the Web object types that they are applied to. Many rules are specific to certain object types. For example, the rules to summarize the text only work on text objects. The rule to reduce image color depth is only applicable to image objects. Some rules may be specific to structure objects or pointer objects. The rules for structure objects generally adapt the structure and determine the layout of the adapted page. The rules for pointer objects generally extract and rearrange the pointer objects in a page to facilitate navigation between pages. The goal of rule classification is to improve the inference speed. Consider the rule is for reducing the color depth of an image object. Although the color depth can be automatically adjusted at the client device, implementing the rule will significantly reduce the size of the image file to be transferred over the network, which is an important concern due to the low network bandwidth and processing power of mobile devices. There are also some rules that are general and applicable to multiple object types. For example, the rule to change the object width can be applied to both image and structure objects.

V. CONCLUSION

The majority cell phone consumers in emergent counties are focus to practice found costly data price procedure in several type or a further consequential in high handling expenses. The Proposed system offered a content- sensitive mobile web accessing agenda that systematically adapts internet pages and content as a utility of a user’s data costing preparation and utilization altitudes to significantly decrease data expenses and increase the quality. Also to minimize the pricing, the proposed methodology improves consumer web accessing comfortable by forwarding the mainly tremendous potential adaptation of a web page for a provided data cost and clarity. The proposed assessment demonstrates the considerable cost savings with content preservation model which the current implementation can present for internet accessing cell phone consumers.

REFERENCES

[1]. Let’s make the web faster - google code. Http articles/web-metrics.html.

[2]. Opera mini. http://www.opera.com/mobile/download/.

[3]. [3] State of the mobile web (2011). http://www.opera.com/smw/.

[4]. [4]Websiteoptimization.com. http://www.websiteoptimization.com/.

[5]. P. Baudisch, X. Xie, C. Wang, and W. Ma. Collapse-to-zoom: viewing web pages on small screen devices by interactively removing irrelevant content. In Proceedings of the 17th annual ACM symposium on User interface software and technology, pages 91–94. ACM, 2004.

[6]. T. Bickmore, A. Girgensohn, and J. Sullivan. Web page filtering and re-authoring for mobile users. The Computer Journal, 42(6):534–546, 1999.

[7]. A. Blekas, J. Garofalakis, and V. Stefanis. Use of rss feeds for content adaptation in mobile web browsing. In Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A): Building the mobile web: rediscovering accessibility?, pages 79–85. ACM, 2006.

[8]. O. Buyukkokten, O. Kaljuvee, H. Garcia-Molina, A. Paepcke, and T. Winograd. Efficient web browsing on handheld devices using page and form summarization. ACM Transactions on Information Systems, 20(1):82–115, 2002.

[9]. J. Charzinski. Traffic Properties, Client Side Cachability and CDN Usage of Popular Web Sites. Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance, 2010.

[10]. J. Chen, L. Subramanian, and J. Li. Ruralcafe: web search in the rural developing world. Proceedings of the 18th International Conference on World Wide Web, 2009.

Volume: 6 Issue: 4 Pages: 2398-2403 (2015) ISSN: 0975-0290

[11]. X. Chen and X. Zhang. Coordinated data prefetching by utilizing reference information at both proxy and web servers. ACM SIGMETRICS Performance Evaluation Review, 2001.

[12]. Y. Chen, W. Ma, and H. Zhang. Detecting web page structure for adaptive viewing on small form factor devices. In Proceedings of the 12th international conference on World Wide Web, May, pages 20–24, 2003.

[13]. J. Domˇcnech, A. Pont, J. Sahuquillo, and J. Gil. A user-focused evaluation of web prefetching algorithms. Computer Communications Review, 30(10), 2007.

[14]. L. Fan, P. Cao, W. Lin, and Q. Jacobson. Web prefetching between lowbandwidth clients and proxies: potential and performance. Proceedings of the ACM SIGMETRICS International Conference on Measurement and modeling of computer systems, 1999.

[15]. A. Fox and E. Brewer. Reducing www latency and bandwidth requirements by real-time distillation. Computer Networks and ISDN Systems, 28(7-11):1445–1456, 1996.

[16]. Y. Hwang, J. Kim, and E. Seo. Structure-aware web transcoding for mobile devices. Internet Computing, IEEE, 7(5):14–21, 2003.

[17]. J. Korva, J. Plomp, P. M¨a¨att¨a, and M. Metso. On-line service adaptation for mobile and fixed terminal devices. In Mobile Data Management, pages 254–259. Springer, 2001.

[18]. T. Laakko and T. Hiltunen. Adapting web content to mobile user agents. Internet Computing, IEEE, 9(2):46–53, 2005.

[19]. E. Lee, J. Kang, J. Choi, and J. Yang. Topic-specific web content adaptation to mobile devices. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pages 845–848. IEEE Computer Society, 2006.

[20]. T. F. MediaLab. Web content adaptation - white paper, 2004.

[21]. P. K. Mishra, K. Baliyan, Deepshika, and S. Singh. User interactive web content adaptation for mobile devices. In 2010 Impact of Globalization and Privatization on meeting India's IT HR needs, March 6-7, 2010.

[22]. I. Mohomed, J. Cai, S. Chavoshi, and E. de Lara. Context-aware interactive content adaptation. In Proceedings of the 4th international conference on Mobile systems, applications and services, pages 42–55. ACM, 2006.

[23]. I. Mohomed, J. Cai, and E. De Lara. Urica: Usage-aware interactive content adaptation for mobile devices. *ACM SIGOPS Operating Systems Review*, 40(4):345–358, 2006.