International Journal of Information Processing, 8(1), 11-22, 2014
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

Challenges and Issues in Adapting Web Contents on Small Screen Devices

Krishna Murthy A [a], Suresha [b], Anil Kumar K M[c]

[a]Department of Computer Science, University of Mysore, Mysore, 570 006, India
Contact: krishnarjun.research@gmail.com

[b]Department of Computer Science, University of Mysore, Mysore, 570 006, India
Contact: sureshabm@yahoo.co.in

[c]Sri Jayachamarajendra College of Engineering, Mysore, India Contact: anilkmsjce@yahoo.co.in

In general, Web pages are intended for large screen devices using HTML technology. Admittance of such Web pages on Small Screen Device's (SSD's) like mobile phones, palmtops, tablets, PDA *etc.*, is increasing with the support of the current wireless technologies. However, SSD's have limited screen size, memory capacity and bandwidth, which makes accessing the Website on SSD's extremely difficult. There are many approaches have been proposed in literature to regenerate HTML Web pages suitable for browsing on SSD's. These proposed methods involve segment the Web page based on its semantic structure, followed by noise removal based on block features and to utilize the hierarchy of the content element to regenerate a page suitable for Small Screen Devices. But World Wide Web consortium stated that, HTML does not provide a better description of semantic structure of the web page contents. To overcome this draw backs, Web developers started to develop Web page(s) using new technologies like XML, Flash *etc.*. It makes a way for new research methods. Therefore, we require an approach to reconstruct these Web pages suitable for SSD's. However, existing approaches in literature do not perform well for Web pages erected using XML and Flash. In this paper, we have emphasized a few issues of the existing approaches on XML, Flash Datasets and propose an approach that performs better on data set comprising of Flash Web pages.

1. INTRODUCTION

The rapid expansion of internet has made Web a popular place for dissemination of information and also provided avenues for research in various fields related to the web. In last few decades, research on Web is increasing at rapid rate. For example, improving the quality of Web by Analyzing Usability Test, Web Information Extraction, Tracking Product Opinions by analyzing user reviews, Browsing Web on SSD's. In general it is called as Web Mining.

According to analysis targets [1], Web mining is divided into three different types namely Web Usage Mining, Web Structure Mining and Web Content Mining. Web usage mining is the process of determining the patterns of users on the internet. It describes how a page is accessed, date and time, the page was accessed, IP address of the browser and page references *etc.* [2][3]. Web Structure Mining is the process of using graph theory to analyze the node and connection structure of the Web site. Web Structure Mining can be divided into two kinds; extracting patterns from hyperlinks in the Web, a hyperlink is a structural component that connects the Web page to a different location, and mining the document structure, analysis of the tree link structure of page structures to describe HTML or XML tag pages [4]. Web Content Mining aims in extracting useful information or knowledge from Web page content [5].

Currently surfing the Web on SSD's such as Mobile phones, Personal Digital Assistants (PDA) *etc.*, is becoming very popular. Delivering Web pages to SSD's have become possible with the advanced wireless technology.

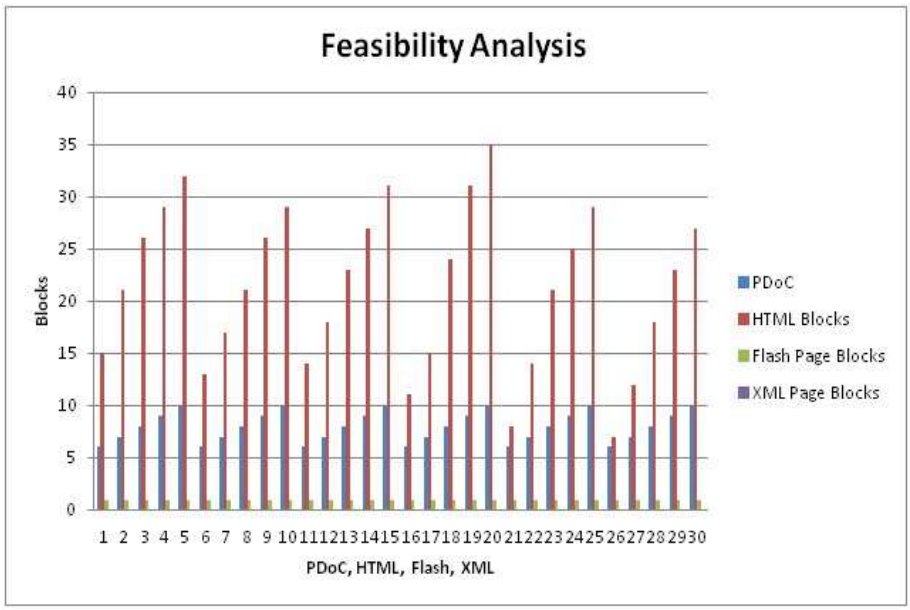

Figure 11. Feasibility Analysis on Existing System.

the better performance level based on response time and content coverage analysis.

REFERENCES

1. Magdalini Eirinaki, WEB MINING: A ROADMAP.
2. Jaideep Srivastava and Robert Cooley. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, *ACM SIGKDD,* 1(2):12–21, 2000.
3. Anil Kumar K M and Suresha. Identifying Subjective Phrases From Opinionated Texts Using Sentiment Product Lexicon, *In International Journal of Advanced Engineering and Applications,* 2:263–271, 2010.
4. Lin Fang and Rui J P de Figueiredo. Energy-Efficient Scheduling Optimization in Wireless Sensor Networks with Delay Constraints, *In IEEE ICC Proceedings*, pages 3734–3739, 2007.
5. Krishna Murthy A and Suresha. Comparative Study on Browsing on SSD's, in *IJMI*, 3:354–358, 2011.
6. http://www.w3c.org/DOM/
7. Kohlschutter C and Nejdl W. A Densitometric Approach to Web page Segmentation, in *WWW*, October 26-30, California, USA, 2009.
8. Ed Tittel. Complete Coverage of XML, *Tata McGraw-Hill Edition.*
9. Jing Li and Ezeife C I. Cleaning Web Pages for Effective Web Content Mining, in *Springer*, LNCS 4080, pages 560–571, 2006.
10. Deng Cai *et al.*. VIPS: A Vision Based Page Segmentation Algorithm, in *Technical Report MSR-TR-2003-79*, 2003.
11. Xiang P F, Yang X and Shi Y C. Web Page Segmentation Based on Gestalt Theory, *in Conference on Multimedia and Expo,* pages 2253–2256 IEEE, 2007.
12. D Chakrabarti, R Kumar and K punera. SmallA Graph-Theoretic Approach to Webpage Segmentation, *the* 17^{th} *International Conference on WWW*, 2008.
13. Xinyue Liu, Xianchao Zhang, Ye Tian and Hongfei Lin. Webpage Segmentation Based on Gomory-Hu Tree Clustering in Undirected Planar Graph, in *NSFC*, 2010.
14. Ruihua Song, Haifeng Liu, Ji-Rong Wen and Wei-Ying Ma. Learning Important Models for Web Page Blocks Based n on Layout and Content Analysis, in *ACM SIGKDD Explorations*, 2004.
15. Christian Kohlschutter, Peter Fankhauser and Wolfgang Nejdl. Boilerplate Detection using

HTML URL's	PDoC	BLOCK's	Flash URL's	PDoC	BLOCK's	XML URL's	PDoC	BLOCK's
	6	15		6	1		6	0
	7	21		7	1		7	0
uni-mysore	8	26	www.isim.	8	1	w3future.com/	8	0
.ac.in	9	29	ac.in/mlw	9	1	html/download	9	0
	10	32		10	1	.xml	10	0
	6	13		6	1		6	0
	7	17		7	1		7	0
www.isim	8	21	asual.com/swf	8	1	lisnews.org/	8	0
.ac.in	9	26	address/samples	9	1	rss.xml	9	0
	10	29	/flash	10	1		10	0
	6	14		6	1		6	0
	7	18		7	1		7	0
www.kiet	8	23	urbansurvi	8	1	xml.sergeant	8	0
conference	9	27	vours	9	1	.org/xpath	9	0
.com	10	31	/flash	10	1	.xml	10	0
	6	11		6	1		6	0
	7	15		7	1		7	0
ugc.ac.in	8	24	noleath.com	8	1	xml.sergeant	8	0
	9	31		9	1	.com/xml/bb/	9	0
	10	35		10	1	.xml	10	0
	6	8		6	1		6	0
	7	14		7	1		7	0
w3schools	8	21	sensisoft	8	1	roben.pl/cz	8	0
.com	9	25	.com	9	1	main.xml	9	0
	10	29		10	1		10	0
	6	7		6	1		6	0
	7	12		7	1		7	0
vtu.ac.in	8	18	marvismint	8	1	theworld.org	8	0
	9	23	.com	9	1	/tech.xml	9	0
	10	27		10	1		10	0

Figure 12. Existing system Feasibility Analysis on our Data set.

Sl.No	Websites	Sony Xperia X10		Sony Ericsson WT13i		LG Optimus Net	
		T-Web pages (Sec)	C-Web pages (Sec)	T-Web pages (Sec)	C-Web pages (Sec)	T-Web pages (Sec)	C-Web pages (Sec)
1	www.isim.ac.in/mlw	1.06	32.47	14.59	ND	6.21	FP
2	www.comunicacion.com	1.3	30.77	59.56	HP+FP	2.95	HP+FP
3	www.urbansurvivours.org	1.22	170	7.57	ND	3.14	FP
4	www.asual.com/swfaddress/ samples/flash	1.28	27.87	4.77	JS+FP	3.32	HP+FP
5	www.noleath.com	2	100	5.47	ND	2.61	23.12
6	www.sensisoft.com	1.28	389.82	5.96	FP	3.24	FP
7	www.marvismint.com	1.27	107.88	5.96	68.58	2.95	4.71

Figure 13. Response Time Analysis on Various SSDs.

Shallow Text Features, in *ACM*, pages 441–450, 2010.

16. Thanda Htwe and Khin Haymar Saw Hla. Noise Removing from Web Pages Using Neural Network, in *IEEE*, 281–285, 2010.
17. Yin X and Lee W S. Using Link Analysis to Improve Layout on Mobile Devices, in *13th*

Sl. No	Websites	Sony Xperia X10		Sony Ericsson WT13i		LG Optimus Net	
		T-Web pages (HV)	C-Web pages (SV)	T-Web pages (HV)	C-Web pages (SV)	T-Web pages (HV)	C-Web pages (SV)
1	www.isim.ac.in/mlw	1	1	1	0	1	0.1
2	www.comunicacion.com	1	1	1	0.2	1	0.2
3	www.urbansurvivours.org	1	1	1	0	1	0.1
4	www.asual.com/swfaddress/samples/flash	1	1	1	0.2	1	0.2
5	www.noleath.com	1	1	1	0	1	1
6	www.sensisoft.com	1	1	1	0.2	1	0.2
7	www.marvismint.com	1	1	1	1	1	1

Figure 14. Content Coverage Analysis on Various SSDs.

WWW, pages 338–344, 2004.

18. Chen Y, Xie X, Ma W Y and Zhang H J. Adapting Web Pages for SSD's, in *IEEE*, 9(1):50–56, 2005.
19. Hattori G, Hoashi K, Matsumoto K and Sugaya F. Robust Web Page Segmentation for Mobile Termilnals, in *WWW*, pages 361–370, 2007.
20. Xin Yang and Yuanchun Sh. Enhanced Gestalt Theory Guided Web Page Segmentation for Mobile Browsing, in *IEEE/ACM*, 3:46–49, 2009.
21. Krishna Murthy A and Suresha. XML URL: Data Set Creation for Future Web Mining Avenues, in *IJCA*, 2012.
22. http://david.abcc.ncifcrf.gov/helps/linear search.html
23. Lan Yi, Bing Liu and Xiaoli Li. Eleminating Noisy Information in Web pages for Data Mining, in *ACM SIGKDD*, 03:296–305, 2003.
24. Ed Tittel, Complete Coverage of XML, *Tata McGraw-Hill Edition.*

Suresha is a Professor, Dept. of Computer Science, University of Mysore, Mysore, India. He obtained his M.Sc from University of Mysore. M.phil from Devi Ahilya Viswa Vidyalaya, Indore, India. He received his M.Tech from Indian Institute of Technology Kanpur, India and Ph.D in Computer Science from Indian Institute of Science, Banglore, India. His area of research intrest is Web Technologies, E-Governance, DBMS and Distributed Systems.

Krishna Murthy A is currently the Research Student working on Web Data Mining at University of Mysore, Mysore. He obtained his Bachelor of Science in Mathematics from Periyar University, Salem, Tamil Nadu India. He received his Master degree in Computer Science from University of Mysore, Mysore, India.

Anil Kumar K M is currently working as Associate Professor, Dept of Computer Science, Sri Jayachamarajendra College of Engineering, Mysore, India. Completed Master of Technology in Computer Network Engineering in 2006 from National Institute of Engineering, Mysore India. And he obtained his Doctoral of Philosophy in Computer Science in 2012 from University of Mysore, Mysore, India. His area of intrest is Web Tecchnologies, Analysis and Design of Algorithms, Network and Internet Technologies.