Dynamic Multiplatform HTML to WML Mobile Converter

Ayma A. Issa, Munib A. Qutaishat

***Abstract*-The research work reported here is concerned with designing and developing a multiplatform dynamic mobile converter that transforms Hyper Text Markup Language (HTML) documents into their equivalent Wireless Markup Language (WML) documents. This system is a server side service that analyzes the client request upon reception, renders the response document to a suitable format, and finally transfers the produced document back to the client. In comparison to the currently available converters, the conversion process of the newly developed converter has been improved in two main perspectives. First, new HTML tags have been supported. Second, new image formats, such as Wireless Bitmaps (WBMP), have been supported. Moreover, more complex HTML structures have been handled. System validation and verification showed a successful transformation rate of more than 75%. Finally, to achieve better transformation results, it was recommended to use the HTML dialects that support mobility, and take the nature of the mobile users into considerations when deciding to enable any site wirelessly.**

***Index Terms*—Hyper Text Markup Language, Wireless Markup Language, Mobile Converter.**

I. INTRODUCTION

Wireless computing, m-commerce, and mobile databases present major new challenges for modern computer science. The use of traditional centralized database systems and traditional hardware is narrowing, while the use of mobile computing systems is spreading. Recently, it is estimated that there will be half a billion WAP-enabled mobile phones globally. This means that potential Wireless Application Protocol (WAP) [13] users will outnumber conventional internet users in the world. WAP users are not sitting in front of a PC. They are on the move, on their way to a meeting, in a crowded train, or sometimes they are under pressure. Therefore, the goal when designing a WAP application should be to build a usable application [5].

WML [14] is an eXtensible Markup Language (XML) based language that is designed to work with small wireless devices of limited resources. Several limitations necessitate the existence of WML instead of using the familiar HTML [12] such as: small screens, limited input facilities, lack of graphics, no or small data storage and processing capabilities, and limited bandwidth. Currently, WAP devices have very little bandwidth available when compared to PCs. In Europe, users can connect on at a maximum speed of 144,000 bits per second (bps) but typically yields data speeds between 40,000 and 60,000 bps which is one fifth of the ordinary home dial-up connections.

From the voice and data perspectives, the experts divided the mobile devices into five categories [11]:

1. Mobile phone: A voice-centric terminal that includes Short Messages Services (SMS) capabilities.
2. Feature phone: A voice-centric terminal with internet connection. SMS, email capabilities, and internet browsing are included, as well as simple games.
3. Personal Digital Assistant (PDA): A data-centric terminal with internet connection. SMS, email capabilities, internet browsing, Personal Information Management (PIM) applications and games are included but it lacks voice capabilities.
4. Smart phone: A voice-centric terminal with internet connection. SMS, email capabilities, internet browsing, PIM applications and games are included.
5. Multimedia terminal: A voice and data terminal with internet connection. SMS, email capabilities, internet browsing, PIM applications, games, audio and video capabilities are included.

Hence, the more powerful wireless device, the better-formatted WML documents. Figure 1 shows several examples of the above mobile devices categories.

The main objective of this research is to extend the capabilities of the available naïve HTML-WML converters by designing and developing a dynamic mobile converter that's capable of transforming most of the traditional static and dynamically generated HTML documents into their equivalent WML documents. This dynamic mobile converter receives the client request, analyzes it using a predefined configuration database, renders the response to a suitable format, and finally transfers the produced document back to the client. Therefore, conventional hypertext pages could be managed and viewed using WAP capable browsers. Although there will, inevitably, be a large amount of formatting lost on conversion due to the nature of differences between HTML and WML, the aim is to provide the user with the ability to browse any conventional web page without the need for a fully blown web browser.

In section 2, a brief summary is presented for documents transformation models available in the literature. Section 3 discusses the proposed system architecture and operation. In section 4, client's requests lifecycle is demonstrated. The validation and verification results of the new system are presented in section 5, and a conclusion including an outline of future research work is presented in section 6.

[1] Ayman A. Issa, Software Engineering Department, Faculty of Information Technology, Philadelphia University, Jordan, aissa@philadelphia.edu.jo
Munib A. Qutaishat, University of Jordam, maq@ju.edu.jo

Figure 1: WAP devices from different categories.

II. DOCUMENTS TRANSFORMATION MODELS

In this section the work of three leading academic and professional research centres in the field of documents conversion will be examined.

A. Kansas State University Model

The computer science department of Kansas State University, as cited in [5], has involved designing and implementing methods of XML to WML processing. In a project accomplished by grad student, entitled *Conversion of given XML data to* WML, Mr. Deep Kapadia discussed and implemented three different solutions to the problem.

The first is concerned with writing a Java program that reads the input, extracts the required data, adds WML tags where appropriate, and outputs a file with a (.wml) extension. The second involved using eXtensible Stylesheet Language Transformation (XSLT) [3], the Xerces XML parser, the Xalan XSLT processor, and a Java file to apply the conversion. Finally, the third method combines a Java servlet with Cocon, which worked like a web server that's only responding to Uniform Resource Locator (URL) requests by publishing files, transformed as specified.

In his conclusions, Mr. Kapadia concludes that his second implementation using XSLT gave the best speed, simplicity, and reusability results.

B. LazyWAPv0.5 Model

LazyWAP is a freeware Personal Home Page (PHP) HTML to WML converter written by an internet consultant [2]. It was classified as a lazy tool since it has fewer than 100 lines of code. It simulates the Java program and Aperghis-Tramoni's CGI/Perl utility developed by Mr. Kapadia's, and again, it has no routines for handling anything more than very simple input.

The methods used involve: (1) string replacement, and (2) highlighting the difficulty and sloppiness of implementations in Java, C, or PHP.

C. Valtion Teknillinen Tutkimuskeskus (VTT) Model

VTT is Finland's leading technical research center that highlighted two approaches for delivering internet data to WAP devices [6]. They have proposed methods of handling frames and complex HTML conversion. Briefly they explain how a tree structure is created and manipulated according to adaptation rules using Document Type Definitions (DTD) [4].

Also, they highlighted (1) how users felt uncomfortable navigating the often-meaningless names given to frames in the output, and (2) how their delivery of tables to the small output screens jumbled up the sorting of links items and confused users.

Most notably from analyzing, evaluating and testing the above literature, problems have arisen in accepting raw HTML input, failing to deal efficiently with HTML nesting constructs, lack of or poor support for frames, and hard and narrow rules of which tags can and cannot be accepted.

Therefore, the proposed new converter has been designed and implemented to solve the failings made in the literature. In particular, this work:

- aims to accept more HTML tags (e.g. <td>, <tr>, <frame>, etc.) in different combinations so as to cover more types of HTML documents,
- introduces efficient routines to handle the nesting constructs,
- implements a new design for presenting frames,
- allows unprecedented ease of code reuse and upgradeability through template modularization, and
- supports dynamic HTML to WML conversion on the fly for all the WML Multipurpose Internet Mail Extension (MIME) types. The WML MIME Types are: plain WML document, compiled WML document, wireless bitmap image, WML Script, and compiled WML Script.

III. SYSTEM ARCHITECTURE AND OPERATION

The architecture of the wired/wireless internet requests consists of four tiers as depicted in figure 2. First, the client tier that sends his request to the server and waits for the response. Second, the server that serves all the clients. Third, the Internet Information Server (IIS) that filters and processes all the coming requests. The developed HTML-WML converter resides in this tier so as to interfere, as detailed in section 4, whenever a wireless request is detected. And, the last tier is the target file itself that will be picked out, processed, and finally sent back to the requesting client.

Several wireless applications architecture classifications have been proposed in the literature. Robertson [10] defined four types of wireless applications in his classification: (1) web repurposer, (2) divergent deployment path, (3) integrated deployment path, and (4) native solution architectures.

Web repurposer includes those solutions that parse and redeploys HTML documents. Divergent deployment path solutions develop parallel paths for deployment to mobile and wired devices. Integrated deployment path allows a single application suite to control business and presentation logic and to deliver content to both wired and unwired devices. Native solution provides programs that are usually proprietary with limited flexibility [10].

Analyzing the design of the proposed converter [5] using Robertson classification, it has been concluded that it can be viewed from different perspectives and can be classified accordingly to any of the above categories as detailed in [5]. In other words, it can be concluded that the system classification into one or more design approaches is a user configuration dependent and the system could be a member of more than one design category concurrently.

a. Compile the WML response before sending it back to the client. This will reduce the response size considerably.
b. Split the HTML response page sections into a reasonable multiple cards response within the WML deck. Figure 4 shows an example of HTML page splitting into multiple WML cards that complies with the upper limit of the WAP packet size.

One main difficulty in the implementation process was designing and implementing an effective handler for documents containing frames. Since WML cannot display conventional frames, it was proposed to offer the users with frames names as hyperlinks, as shown in figure 5, so they can choose which file they wish to view/translate.

Finally, when the WML response is generated, the job of the developed filter is completed by returning it to the web server and from it directly to the ISP, which returns it back to the client. Figure 3 depicts WML converter lifecycle.

V. SYSTEM VALIDATION AND VERIFICATION

The system has been validated and verified using as wide a variety of HTML documents as possible. These HTML documents were gleamed from live websites. The process of selection was a mixture of random retrievals through www.random.com/all/, and internet hyperlinks following.

The selected HTML pages were of different types in order to cover as much as wireless application types as possible. And though cover as many different structures and tag sets as possible regardless of how common or impractical these may be.

The browsers used for retrieving and saving the input data were alternated between Netscape 4.7, IE5.0 and Opera 3.60. However, theoretically this would not affect the software's operation but will help in looking to the HTML documents from different perspectives since the browsers differ in their forgiving percentage.

The generated WML documents have been validated and verified using one of the following three methods:

1. Manual method using either a file editor or some online WML validators.
2. Browse the WML documents using WML browser emulators to show if the output was generated and displayed correctly. Three emulators platforms has been selected for this purpose; M3gate [8]**,** Nokia [7], and OpenWave [5].
3. Third, an online real time testing was performed by: (1) installing and hosting the system on an online available exposed server, (2) directing real requests to the configured serer using different mobile types as detailed in section 1.

The rationale behind using more than one WML browser emulator and device is the lack of standardized behaviour among wireless infrastructure components. Therefore, the

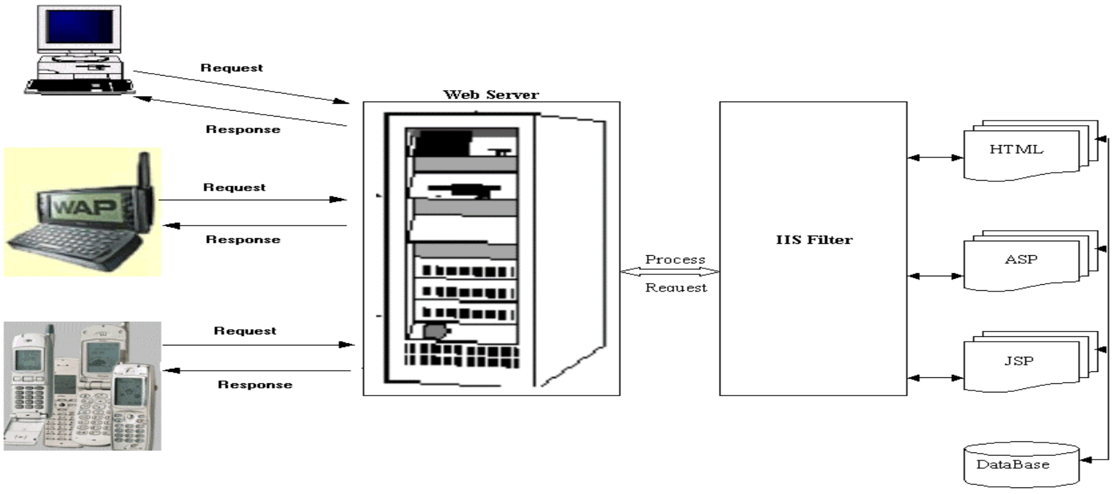


Figure 2: Architecture of internet requests.

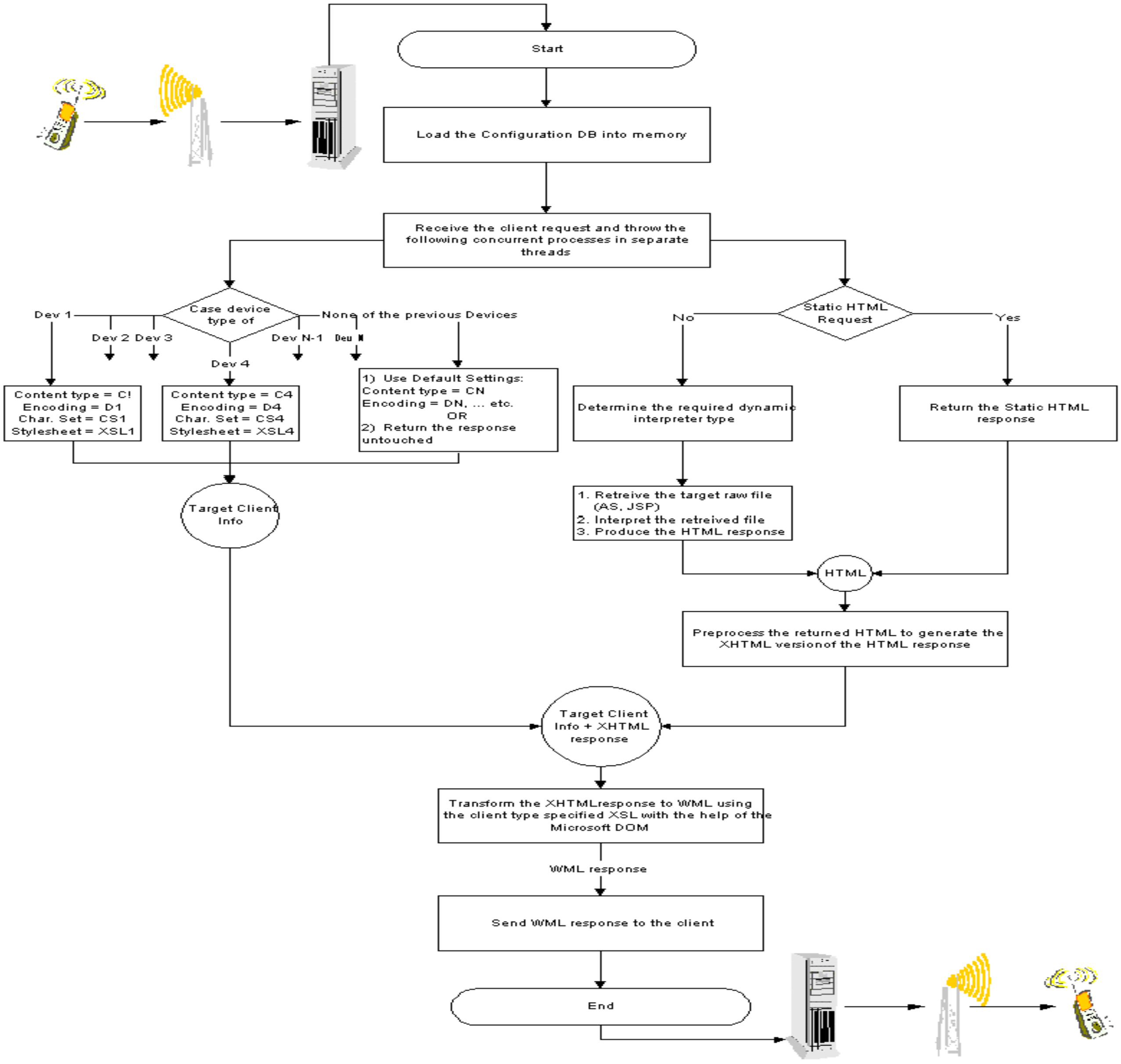

Start
Load the Configuration DB into memory
Receive the client request and throw the following concurrent processes in separate threads
Case device type of
Dev 1
Dev 2
Dev 3
Dev 4
Dev N-1
Dev N
None of the previous Devices
Content type = C!
Encoding = D1
Char. Set = CS1
Stylesheet = XSL1
Content type = C4
Encoding = D4
Char. Set = CS4
Stylesheet = XSL4
1) Use Default Settings:
Content type = CN
Encoding = DN, ... etc.
OR
2) Return the response untouched
Target Client Info
Static HTML Request
No
Yes
Determine the required dynamic interpreter type
Return the Static HTML response
1. Retreive the target raw file (AS, JSP)
2. Interpret the retreived file
3. Produce the HTML response
HTML
Preprocess the returned HTML to generate the XHTML versionof the HTML response
Target Client Info + XHTML response
Transform the XHTMLresponse to WML using the client type specified XSL with the help of the Microsoft DOM
WML response
Send WML response to the client
End

Figure 3: Client's requests lifecycle.

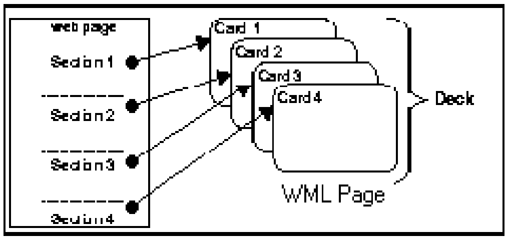
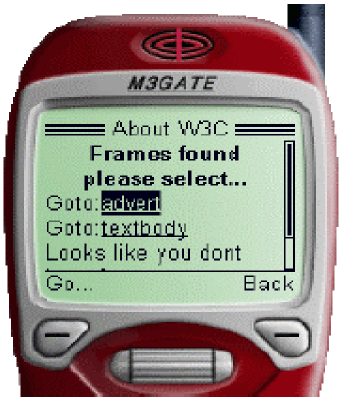


Figure 4: HTML pages sections splitting to WML cards.



About W3C
Frames found please select...
Goto:advert
Goto:textbody
Looks like you dont

Figure 5: Frames handling example.

decision was to test as many device, browser, character set, encoding, and gateways combinations to discover any undocumented idiosyncrasies and peculiarities.

Over fifty pages were chosen, processed, and tested. Although fifty documents represent a very small fraction of the available HTML files, their basic structure is represented by the cross section of the tested pages, which encompassed as wide as possible mix of HTML branching constructs. Coupled with this, the time involved in running test data and checking every output in each browser was considerable and sufficient for almost accurate validation and verification. Accuracy has been supported by employing a specialised testing team for this purpose.

A successful transformation is defined as valid well formed WML document that is displayed in all testing browsers and maintaining a reasonable level of formatting.

Figures 6 and 7, and 8 and 9 present two pairs of sample HTML pages and their corresponding WML transformation for a classic car web page and forms usage, respectively, produced using the developed system.

The validation and verification phase passed through several iterations. Namely, there was three distinct iterations that have been targeted to different HTML documents types with different complexity levels.

Table 1 clarifies the results of the main three validation and verification iterations. The failed transformations are marked with ‘X’ in the corresponding cell.

Using the above testing results, the failure percentage could be calculated as:

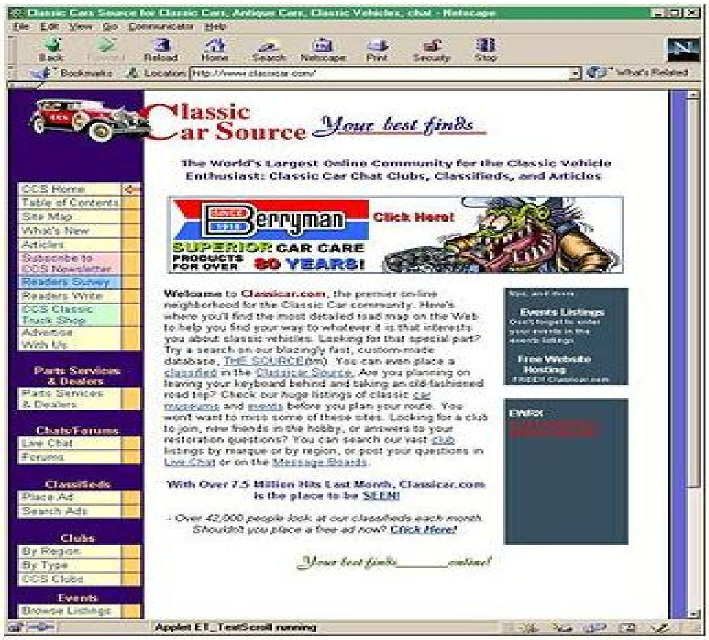
$$\frac{\text{Number of Failed Transformations}}{\text{Total Number of Transformation Samples}} = \frac{11}{50} \times \frac{100}{100} = 22\,\%$$

which implies that a success rate of 78% has been achieved. The main reason for some of the failed transformations was related to quite unexpected HTML tags nesting which outlines future upgrades and refinements to the developed system.

VI. CONCLUSION AND FUTURE WORK

The design and implementation for a new dynamic HTML to WML converter have been presented. The new converter does not cover the entire HTML to WML conversions due to some unexpected HTML tags nesting. However, a wider range of HTML tags and structures are supported compared to the currently available converters. In addition, a successful rate of more than 75% of the input HTML files has been achieved.

The system can be considered as a device independent if the user defines one XSL transformation for all the device types, the matter that will sacrifice many individual device features and use the lowest common denominator. Alternatively, it can be considered as a device dependent if the user configures an XSL for each device type.


Classic Cars Source for Classic Cars, Antique Cars, Classic Vehicles, chat - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http://www.classicar.com/ What's Related
Classic Car Source Your best finds
CCS Home
Table of Contents
Site Map
What's New
Articles
Subscribe to CCS Newsletter
Readers Survey
Readers Write
CCS Classic Truck Shop
Advertise With Us
Parts Services & Dealers
Parts Services & Dealers
Chats/Forums
Live Chat
Forums
Classifieds
Place Ad
Search Ads
Clubs
By Region
By Type
CCS Clubs
Events
Browse Listings
Berryman Click Here! SUPERIOR CAR CARE PRODUCTS FOR OVER 80 YEARS!
Events Listings
Free Website Hosting
Applet E1_TextScroll running

The World's Largest Online Community for the Classic Vehicle Enthusiast: Classic Car Chat Clubs, Classifieds, and Articles

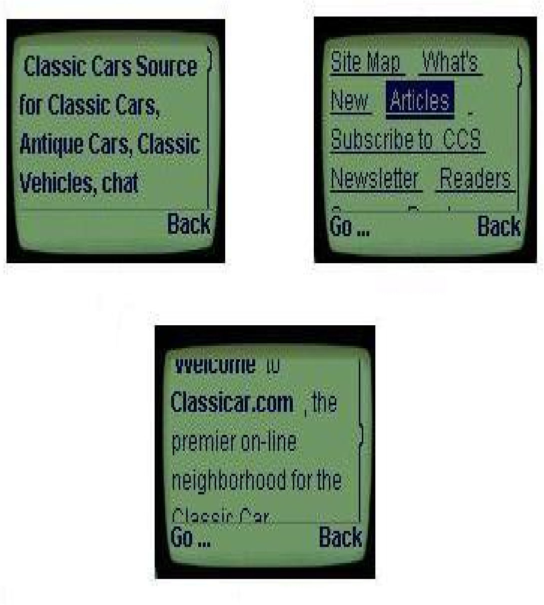
Welcome to Classicar.com, the premier on-line neighborhood for the Classic Car community. Here's where you'll find the most detailed road map on the Web to help you find your way to whatever it is that interests you about classic vehicles. Looking for that special part? Try a search on our blazingly fast, custom-made database, THE SOURCE(tm). You can even place a classified in the Classicar Source. Are you planning on leaving your keyboard behind and taking an old-fashioned road trip? Check our huge listings of classic car museums and events before you plan your route. You won't want to miss some of these sites. Looking for a club to join, new friends in the hobby, or answers to your restoration questions? You can search our vast club listings by marque or by region, or post your questions in Live Chat or on the Message Boards.

With Over 7.5 Million Hits Last Month, Classicar.com is the place to be SEEN!

- Over 42,000 people look at our classifieds each month. Shouldn't you place a free ad now? Click Here!

Your best finds______online!

Figure 6: Classic car example HTML source document.

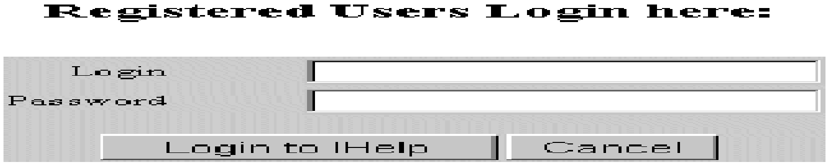


Classic Cars Source for Classic Cars, Antique Cars, Classic Vehicles, chat

Site Map What's New Articles Subscribe to CCS Newsletter Readers

Welcome to Classicar.com , the premier on-line neighborhood for the Classic Car

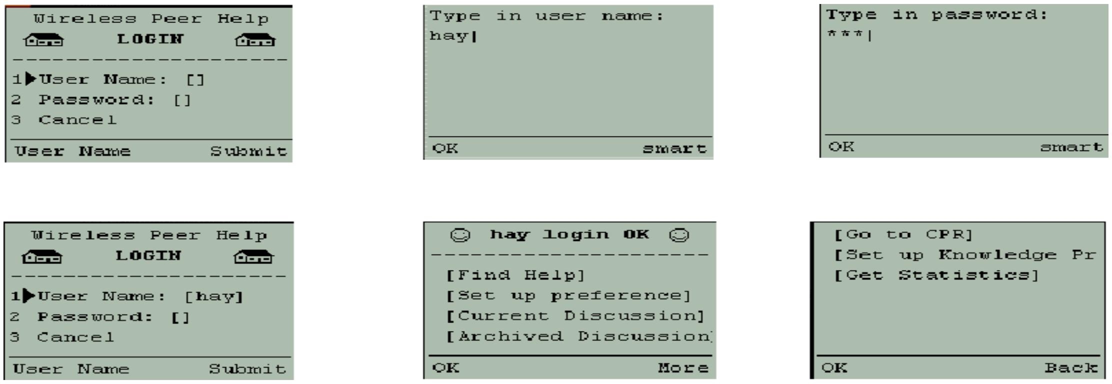
Figure 7: Classic car example converted HTML document displayed using WAP emulator.



Registered Users Login here:

Login	
Password	

Figure 8: Forms usage example HTML source document.



Wireless Peer Help
LOGIN
1 User Name: []
2 Password: []
3 Cancel
User Name Submit

Type in user name:
hay|
OK smart

Type in password:
***|
OK smart

Wireless Peer Help
LOGIN
1 User Name: [hay]
2 Password: []
3 Cancel
User Name Submit

☺ hay login OK ☺
[Find Help]
[Set up preference]
[Current Discussion]
[Archived Discussion
OK More

[Go to CPR]
[Set up Knowledge Pr
[Get Statistics]
OK Back

Figure 9: Forms usage example converted HTML document displayed using WAP emulator.

Table 1: Test Data Transformation Results.

Test Data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Test Iteration I		X	X				X		X		X			X			
Test Iteration II	X									X		X					
Test Iteration III									X		X						

The main recommendations that will help in increasing the above successful conversion rate are:

1. Use HTML dialects that are defined to support a wide range of target devices in designing new HTML sites.
2. Take the WAP users nature and the conditions they are facing during accessing WAP sites through their WAP devices into consideration when designing wireless enabled sites.

In addition to the system performance and throughput, the research team is studying several possible extensions to the developed converter including:

1. The XSL transformer can be extended to handle more dynamic HTML resources such as PHP, CGI [1], DHTML, and New Markup Languages.
2. More work can be appended as new XSL templates that handle any new WML tags when added in any prospective WML releases.
3. The wireless communications for the next generation of mobile phones and how more multimedia can be supported in the mobile phones is a research topic that can be combined to the developed mobile converter in order to support HTML sites with more visual effects [12,13].

REFERENCES

[1] Aperghis, S., (2001). HTML2WML CGI Converter Version 0.4.1 [Online]. POD Translation. Available from http://maddingue.free. fr/techie/ html2wml/test.html [Accessed 10/8/2002].

[2] Baikov, M., (2001). Lazy WAP v0.5. [Online]. Available from http://www.tourbase.ru/zink/ [Accessed 20/8/2002].

[3] Cagle, K., and Corning, M., (2001). *Professional XSL*. Wrox Press Ltd. 2nd edition. Canada.

[4] Eckstein, R., (1999). *XML Pocket Reference*. 2nd edition. O'Reilly & Associates, Inc.UK.

[5] Issa, A., (2003). *Supporting Mobile Databases By Translating Traditional Web Pages Into Wireless Markup Language (Wml) Pages*. MSc Thesis. University of Jordan.

[6] Kaasinen E, Aaltonen M, Kolari J, Melakoski S, and Laakko T., (2000). Two approaches to bringing Internet services to WAP devices [Online]. Elsevier Science BV, Amsterdam. Available from http://www9.org/w9cdrom/228/ 228.html [Accessed 10/1/2003].7] Kapadia, D., (2000). *Conversion of given XML data to WML*. MSc Thesis. Kansas State University, USA.

[7] Nokia Group, (2002). SDK and its WAP phone emulator [Online]. Available from http://www. nokia.com [Accessed 15/1/2003].

[8] Quick Algorithm Laboratories, (2001). M3Gate WAP phone emulator. USA

[9] Raggett, D., (2001). HTML Tidy [Online]. Available from http://www.w3.org/people/Raggett /tidy/ [Accessed 10/1/2003].

[10] Robertson, A., (2001). Wireless Application Architectures for Business: Key Implementation Consideration. *Cutter IT Journal*, March 2001, Vol.14, No. 3.

[11] Skoglund, E., (2002). The OSE Wireless Platform, Making 3G Handsets a Reality [Online]. OSE Systems. Available from http://www. ose.com/ [Accessed 2/12/2002].

[12] W3C Note, (1999). HTML 4.0 Guidelines for Mobile Access [Online]. Available from http:// www.w3.org/ [Accessed 15/12/2002].

[13] WAP Forum, (2002). Wireless Application Protocol 2.0 [Online]. Available from http://www.wapforum.org [Accessed 15/12/ 2002].

[14] WAP Forum, (2002) WML Tag Reference. 2002 [Online]. Available from http:// www.w3schools.com/wap/wml _reference.asp [Accessed 15/12/2002].