4th WSEAS Int. Conf. on COMPUTATIONAL INTELLIGENCE, MAN-MACHINE SYSTEMS and CYBERNETICS Miami, Florida, USA, November 17-19, 2005 (pp123-129)

Enabling the Mobile Web Through Auto-generating Multimodal Web Pages

KHALED KHANKAN, ROBERT STEELE
Faculty of Information Technology
University of Technology, Sydney
PO Box 123 Broadway, NSW 2007
AUSTRALIA

Abstract: - Multimodal interaction refers to interactions involving a combination of voice, text, stylus etc. Such multimodal interaction can provide advantages for mobile device input/output given the limitations of mobile device keypads and screens and the limits on user's availability to use such traditional inputs and outputs while mobile. One important resource that can be accessed from mobile devices is the World Wide Web, but Web pages, as they are visual and may require typed input, may not be readily suited to mobile device access. While there are technology initiatives addressing the authoring of new Web pages that are multimodal-enabled, the generating of multimodal versions of existing Web pages has not been addressed. To help address this problem, we propose in this paper an architecture and approach to auto-generating multimodal representations of existing Web pages, thereby helping to enable the Mobile Web.

Key-Words: - Multimodal, mobile Web, auto-generation, XForms

1 Introduction

With the increasing availability of Internet-enabled mobile devices, including mobile phones, personal digital assistants (PDAs), and in-car systems, accessing the Web anywhere anytime is no longer an idea, it is happening.

However, mobile Web access is far from the proliferation stage for many reasons, including, the nature of the Web contents, the hardware and software limitations of the mobile devices, and the diversity of these devices.

Firstly, most of the existing Web pages have been authored without the mobile access probability in mind, but rather with the assumption that the Web access will be through powerful desktops or laptops, which is increasingly not the case anymore. Obviously, it is not practical to re-author the whole Web contents in order to support the currently available (or yet to be developed) range of mobile device.

Secondly, the majority of mobile devices have limited input-output capabilities and mobile users usually have to input data through a small keypad and receive the output through small LCD. As such, filling Web forms and / or browsing ordinary Web pages involves significant effort in scrolling and keeping track of the screen contents.

Thirdly, it is a real challenge for mobile Web developers to support and provide an appropriate user interface for the ever-increasing number of mobile devices.

For the above-mentioned reasons, facilitating the mobile Web access proliferation is greatly facilitated by a mechanism that automatically presents the Web contents through a convenient, natural, yet efficient multimodal user interface and relieves both the device user and UI developer from the burden of dealing with the mobile device limitations.

In this paper, we introduce a novel approach supported by a suggested architecture for automatic generation of device independent multimodal enabled versions of existing Web pages.

The paper begins with the motivation behind the proposed architecture, followed by a brief description of the involved technologies as well as other research efforts relevant to our work. Next, a presentation of the architecture that supports our approach and shows how it allows generating a transparent multimodal enabled interface for existing Web pages followed by a discussion about alternate architectures before we conclude.

2 Motivation

Mobile users might need to access the Web virtually anywhere anytime to accomplish time critical tasks. For instance, they might need to communicate with their enterprises or home offices, check for emails or send a new one, or check for real-time information such as stock exchange prices or flight arrival. However, using mobile devices for this purpose is not as convenient as using desktop computers and for those users accessing the Web through a convenient, fast, and efficient user interface might be long expected.

Knowing that multimodal interaction (MMI) could be substantially faster than traditional GUI-based interaction as suggested by the studies [2] and [4], we believe facilitating the above multimodal dialog as part of a convenient and usable representation of the Web contents to the mobile users can have a great impact on the proliferation of mobile Web access. Moreover, to support the vast number of mobile devices, we consider the auto-generation approach of device independent multimodal user interface as a faster approach of user interface development.

As indicated in [1], mobile users are usually interested in reaching specific Web pages through the minimum possible links. However, in order to reach the destination page and access certain information, mobile users usually have to deal with Web forms to enter data and see results traditionally as text and or graphics.

Assume for a moment that, during a meeting, a mobile user wants to check the current exchange rate of a foreign currency before committing to a certain business decision. Within the vicinity of a local access point and using a PDA with Wireless connectivity capabilities, the user connected to the Internet and typed the URL of a foreign exchange currency converter Web page such as the one shown in figure 1. Typically, some of the page contents or Web form contents are not immediately visible as the case when the user accesses the same Web page using a desktop PC and the user has to scroll horizontally and vertically few times to fill that form. Moreover, the user has to scroll two long lists of hundreds of currency types to select the currencies involved in the query.

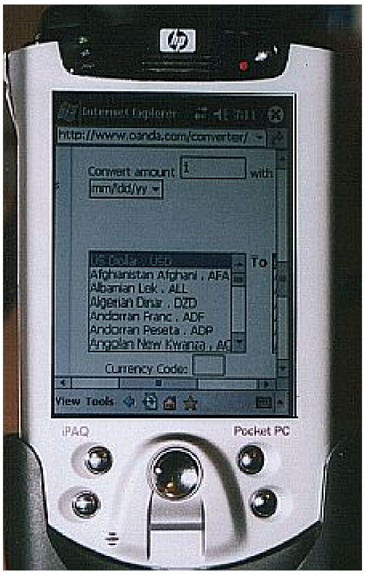

Fig 1: A currency converter Web page

Alternatively, a multimodal interaction enabled Web interface enables the user to perform the same task more in a more convenient and faster approach.

User: { using a stylus} WWW.xyz.com
System: {voice} Welcome to the …currency converter service; please select the currency to be converted.
User: {voice} Australian dollar
System: {voice} Please select the desired currency
User: {voice} US dollar
System: {voice} How many dollars would you like to convert?
User: {voice} 1
System: {voice} The Australian dollar is currently = … US dollars. Would you like to check other currencies?
User: {voice} yes
System: {voice} Please select the currency to be converted
…

Fig 2: A multimodal dialog for currency exchange query

Figure 2 shows is an example for a dialog between the user and multimodal enabled interface during the session of conducting the above-mentioned task. More details about this example will be discussed in section 4.

3 Background

In addition to the multimodal interaction, and among other technologies involved in the proposed architecture, are XForms [8] and VoiceXML [10]. In this section, we briefly highlight these technologies and review some of the research efforts related work.

3.1. Related Technology

In principle, the multimodal interaction is giving input by means of various synchronized modalities, such as text, speech, keypad, digital ink (i.e. the motion of a stylus), pointing device, lip movement, or gaze tracking, and receiving output in a number of synchronized modalities as well, such as text, speech, audio, graphics, or video. In this regard, the W3C Multimodal Interaction Working Group has published a multimodal interaction framework specifications [5] followed by two proposed markup languages specifications, namely, InkML [7] and EMMA [9] to support any architecture implementing the proposed framework.

The second related technology is an emerging Web forms technology called XForms. According to W3C, XForms is a specification of Web forms that can be used with desktops, handheld devices, and others. Among other advantages of XForms is decoupling the data model from the interface and binding them at run time. In fact, that makes XForms very appropriate for developing device and modality independent interfaces and paves the way for adaptive mobile UI, which can respond to the environment changes and be updated on the fly.

Furthermore, XForms is an extension of XHTML and its widgets are abstract in nature and the XForms processor is responsible for translating these abstract widgets into specific concrete interface components based on the client device profile.

The third related technology is VoiceXML. VoiceXML is a standalone markup language that has its own control flow and execution environment. Furthermore, VoiceXML has the ability to be smoothly integrated with XForms and XHTML and be attached to user interface controls as voice handlers using XML Events to produce multimodal interfaces.

3.2. Related Work

The related efforts that have been investigated can be categorized in two groups:

1. Realization of MMI-enabled systems
2. Generating of user interface for mobile devices to facilitate Web browsing on mobile devices

In terms of multimodal interaction and the realization of multimodal interaction-enabled systems, a few trends of research efforts are evident. First, projects such as [13], [14], and [19] focused on the multimodal integration, fusion, and fission of modalities. Second, projects such as [11] and [13] attempted to introduce multimodal interaction layer for map-based systems in order to produce interactive maps. Third, researches such as [12] and [15] focused on providing multimodal frameworks to facilitate developing multimodal enabled applications.

We contributed to these efforts by introducing an architecture that automatically generates abstract multimodal user interface for mobile Web services in [18] and in this paper we extend these efforts to address the Web contents in the context of Web pages.

In terms of facilitating Web browsing on mobile devices, there are a number of efforts, which can be categorized in two mean trends: The first trend is based on automatically generating mobile user interface with or without multimodal capability. However, each of these proposed approaches has its limits that make it inappropriate for the task at hand of presenting the Web contents in multimodal fashion. For instance, Reitter and colleagues has an attempt to generate multimodal adaptive user interface [16], which might be appropriate for mobile applications, however, it assumes an unambiguous, language-and-mode independent input, and obviously the existing Web contents can not fulfill this requirement. Another example is [3], which attempts to adapt the existing Web pages by splitting them into smaller logically-related units and automatically transform them into mobile devices friendly uni-modal (graphical) views, that is, it *does not* address multimodality.

The second trend is influenced by the authoring techniques identified by the W3C Consortium in [6], namely, multiple authoring, single authoring and flexible authoring. The attempts in this direction as in [17] are based on creating new abstracted Web pages with adaptation capabilities to various devices. While this approach might be acceptable for new Web pages that are yet to be authored, it is not practical for many existing Web pages.

We attempt to automatically generate a multimodal mobile Web interface for the existing Web pages without assuming any particular input as in [16] or having the Semantic Web existing as in SmartWeb [20]. In this paper, we focus on addressing the Web forms component of Web pages as part of our ongoing research.

4. Architecture

Web pages, usually, have one or more of the following components: pure text, graphics, hyperlinks, forms, and embedded components such as java applets or ActiveX components. These components present a certain content structure through a designated layout structure. Extensive analysis of these structures is needed before adapting any Web page and representing that page using multiple modalities. As such, the process of generating a multimodal representation for existing Web page has three phases: First, analyzing the semantic structure of the Web page. Second, extracting the Web forms and generating correspondent abstract XForms. Finally, generating a concrete Web multimodal interface in X+V format

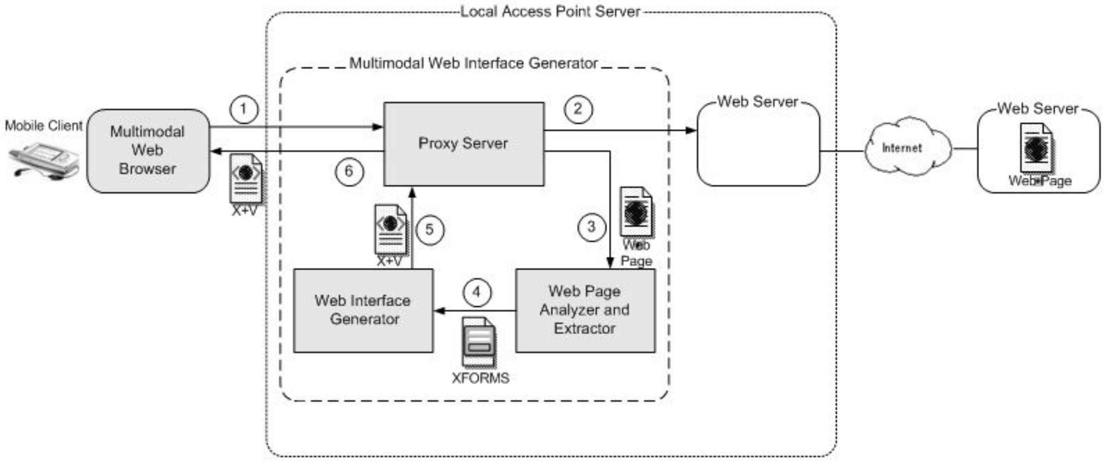

Fig 3: The Architecture overview and data flow

out the XForms-based interface. These stages can be modeled in separate components and deployed in a client-based, server-based, or proxy-based architecture. However, we recommend a proxy-based architecture as the one illustrated in figure 3 for the reasons discussed later in section.

The proposed architecture has three main components, namely, Web page analyzer, MMI Web interface generator, and MMI rendering component, typically, a multimodal browser. These can all be situated at a proxy Web server. The roles of these components are illustrated through the following example scenario:

1. In the vicinity of a local access point, a mobile user starts the multimodal browser installed on his/her PDA and enters the URL of a currency converter Web page. As a result, a http request is intercepted by the proxy server connected to the access point.
2. The proxy server looks up the cashed Web sites repository, and upon failing to find a cashed copy of the requested page, it passes the request to the Web server, which responds by retrieving the requested Web page or giving an error message in case of failing to do so.
3. Once the proxy server receives the requested Web page, the proxy server sends it to the Web page analyzer. In turn, the Web page analyzer extracts the <Form> block out of the html file and starts analyzing the form block to identify the input fields. Figure 4 shows the main HTML <Form> block of a currency converter Web page.
4. The extracted information is then transcoded into XForms and passed to the MMI Web interface generator component. Figure 5 illustrates part of the generated XForms code corresponding to the retrieved HTML form.
5. Based on the client device profile, context and user preferences, the MMI Web interface generator processes the abstracted XForms and generates a concrete ready-for-rendering

```
…
…
<form method="post" action="/convert/classic" ENCTYPE= "x-www-
      form-encoded" id=form1 name=form1>
      <h1>Currency Converter </h1>
      To get the exchange rates for any of the 164 currencies,
      select the desired currencies from the lists below and amount
      for which you would like to conduct the currency conversion.
      Click on &quot;Convert Now&quot;
      to get the results of your currency conversion. <br>
      <br><p><nobr> Convert amount
      <input name="value" value="1" size=7 maxlength=15>
      <br><br></nobr></p><br>
      <table cellpadding=0 cellspacing=0 border="0">
        <tr> <td align=center width="20" > &nbsp; <big>
       <strong> To</strong></big>&nbsp;</td>
       <td align=center width="20" >&nbsp;<big>
       <strong>From</strong></big>&nbsp;</td> </tr>
        <tr><td align=center width="100" >
        <select name="exch" size=7>
         <OPTION VALUE="USD" SELECTED>US Dollar . USD
         <OPTION VALUE="AFA">Afghanistan Afghani . AFA
         <OPTION VALUE="ALL">Albanian Lek . ALL
         <OPTION VALUE="DZD">Algerian Dinar . DZD
         <OPTION VALUE="ADF">Andorran Franc . ADF
         <OPTION VALUE="ADP">Andorran Peseta . ADP
         <OPTION VALUE="AON">Angolan New Kwanza . AON
         <OPTION VALUE="ARS">Argentine Peso . ARS
         <OPTION VALUE="AWG">Aruban Florin . AWG
        <OPTION VALUE="AUD">Australian Dollar . AUD
          …
        </select> <br> </td>  <td align=center width="100" >
        <select name="expr" size=7>
        <OPTION VALUE="EUR" SELECTED>Euro . EUR
          …
        <OPTION VALUE="USD">US Dollar . USD
        <OPTION VALUE="UGS">Uganda Shilling . UGS
        <OPTION VALUE="UAH">Ukraine Hryvnia . UAH
        <OPTION VALUE="UYP">Uruguayan Peso . UYP
        <OPTION VALUE="AED">Utd. Arab Emir. Dirham . AED
        <OPTION VALUE="VUV">Vanuatu Vatu . VUV
          …
        </select> <br></td> </tr>
        <tr> <td width="20%" > </td><td align="center" width="41%">
       <strong><big><input type="SUBMIT" value= "Convert Now"
        bgcolor=YELLOW name="SUBMIT"></big></strong>
        </td></tr></table>
 </form>
…
```

Fig 4: Retrieved HTML form

information in X+V format and send it to the proxy server.

6. The proxy server sends this X+V formatted information to the client browser as the response for its http request. Finally the multimodal browser renders the received X+V and the interaction dialog with the user.

```
…

  <xforms:model id="mdlMessage">
    <xforms:instance>
      <query>
            <fromC>US Dollar</fromC>
             <toC>Australian Dollar<toC>
             <amount>1</amount>
      </query>
    </xforms:instance>
    <xforms:submission action="http://localhost/currencyconverter/
                              currencyconverter.aspx?Handler=Convert"
     method="post" id="submit" includenamespaceprefixes=""/>
  </xforms:model>
 …
  <xforms:select1 ref="fromC"> <!--usually a combo box -->
         <xforms:label>From</xforms:label>
           <xforms:hint>
             Choose the currency that you would like to convert
           </xforms:hint>
 …
         <xforms:item>
          <xforms:label>Afghanistan Afghani . AFA </xforms:label>
           <xforms:value>Afghanistan Afghani</xforms:value>
         </xforms:item>
         <xforms:item>
           <xforms:label>Australian Dollar . AUD</xforms:label>
           <xforms:value>Australian Dollar</xforms:value>
         </xforms:item>
 .
         <xforms:item>
           <xforms:label>US Dollar . USD</xforms:label>
           <xforms:value>US Dollar</xforms:value>
         </xforms:item>
  </xforms:select1>
  <xforms:select1 ref="toC">
         <xforms:label style="width: 150px;">To:</xforms:label>
         <xforms:hint>
            Choose the currency that you would like to convert to
         </xforms:hint>
         …….
         <xforms:item>
           <xforms:label>Australian Dollar . AUD</xforms:label>
           <xforms:value>Australian Dollar</xforms:value>
         </xforms:item>
            …
  </xforms:select1>
  <xforms:input ref="amount">
      <xforms:label>Amount:</xforms:label>
      <xforms:hint>
            Enter the amount that you would like to convert to
      </xforms:hint>
  </xforms:input>
  <xforms:submit submission="submit">
        <xforms:label>Submit</xforms:label>
```

Fig 5: A sample for a generated XForm

4.1. Alternate architectures

As an alternate for the suggested proxy-based architecture is a client-based architecture. In this architecture, a mobile application deployed to mobile device will include all the above-mentioned components (i.e. Analyser, extractor, MMI Web interface generator, and MMI browser). On one hand, this architecture neither needs a dedicated proxy server to perform the Web page processing and generating the MMI equivalent interface nor needs modifying the Web servers of the content providers. On the other hand, this architecture requires a client mobile device with considerable computational power; which is not the case for most of the mobile devices. Moreover, this architecture requires deploying the client application to every client device, causing maintenance and updating overhead. Requiring client side adoption and change generally represents a greater barrier to adoption.

Another alternate would be a server- based architecture. In this architecture, a multimodal browser is installed on the mobile client device and the multimodal Web interface generator component is deployed to the server side and can be integrated with the Web server. Obviously this will require updating and maintaining the deployed components on the Web server of every content provider. It also needs supporting the various types of Web servers.

Obviously, in a proxy- based architecture there is no need to modify, update, or maintain either the client devices or Web servers. In addition, it utilizes the computational power and resources that the proxy servers usually have. With this architecture, the generated XForms can be cashed on the proxy server for better system responsiveness, and manual refinement if needed. Furthermore, in this architecture, the client multimodal browser is required to process only X+V and the XForms processing is delegated to a proxy component.

5. Future Work

As we mentioned earlier in this paper, we attempt to auto generate MMI enabled Web interface for mobile devices out of existing Web pages. We started by addressing the Web forms as a pivotal component of any Web page. Further research will address the other components of the Web page contents structure as well the layout structure.

Also, a testbed will be setup to evaluate the generated multimodal Web interfaces and investigate the limits and issues that may arise during a full auto-generation cycle of modalities.

6 Conclusion

Multimodal mobile Web access of existing Web pages has the potential to significantly change the mobile devices pattern of use, leading to better user experience and productivity.

In this paper, we proposed an approach for automatically generating multimodal representation of existing Web pages adaptable to mobile devices.

References:

[1] Anderson, C.R., Domingos, P. & Weld, D.S. 2001, 'Adaptive web navigation for wireless devices', *The 17th International Joint Conference on Artificial Intelligence*, Seattle, USA, pp. 879-884.

[2] Chai, J., Lin, J., Zadrozny, W., Ye, Y., Budzikowska, M., Horvath, V., Kambhatla, N. & Wolf, C. 2000, 'Comparative Evaluation of a Natural Language Dialog Based System and a Menu Driven System for Information Access: a Case Study', *The International Conference on Multimedia Information Retrieval (RIAO 2000)*, Paris, France.

[3] Chen, Y., Xie, X., Ma, W.-Y. & Zhang, H.-J. 2005, 'Adapting Web Pages for Small-Screen Devices', *IEEE Internet Computing*, vol. 9, no. 1.

[4] Cohen, P.R., Johnston, M., McGee, D., Oviatt, S.L., Clow, J. & Smith, I. 1998, 'The efficiency of multimodal interaction: A case study', *The International Conference on Spoken Language Processing*, Sydney, pp. 249-252.

[5] Consortium, W.W.W. 2003, Multimodal Interaction Framework, viewed 29 September 2005 <http://www.w3.org/TR/2003/NOTE-mmi-framework-20030506/>.

[6] Consortium, W.W.W. 2004a, Authoring Techniques for Device Independence. W3C Working Group Note 18 February 2004, viewed 29 September 2005 <http://www.w3.org/TR/2004/NOTE-di-atdi-20040218/>.

[7] Consortium, W.W.W. 2004b, Ink Markup Language, W3C Working Draft 28 September 2004, viewed 29 September 2005 <http://www.w3.org/TR/2004/WD-InkML-20040928/>.

[8] Consortium, W.W.W. 2004c, XForms-The Next Generation of Web Forms, viewed 29 September 2005 <http://www.w3.org/MarkUp/Forms/>.

[9] Consortium, W.W.W. 2005a, EMMA: Extensible MultiModal Annotation markup language W3C Working Draft 16 September 2005, viewed 29 September 2005 <http://www.w3.org/TR/emma/>.

[10] Consortium, W.W.W. 2005b, Voice Extensible Markup Language (VoiceXML) 2.1, viewed 29 September 2005 <http://www.w3.org/TR/voicexml21/>.

[11] Corradini, A., Wesson, R.M. & Cohen, P.R. 2002, 'A Map-based System Using Speech and 3D Gestures for Pervasive Computing', *ICMI 2002*, Pittsburg, USA, pp. 191-196.

[12] Flippo, F., Krebs, A. & Marsic, I. 2003, 'A framework for rapid development of multimodal interfaces', *The 5th international conference on Multimodal interfaces*, Vancouver, British Columbia, Canada, pp. 109-116.

[13] Johnston, M., Bangalore, S., Vasireddy, G., Stent, A., Ehlen, P., Walker, M., Whittaker, S. & Maloor, P. 2002, 'MATCH: An Architecture for Multimodal Dialogue Systemsa', *The 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, USA, pp. 376-383.

[14] Johnston, M., Cohen, P.R., McGee, D., Oviatt, S.L., Pittman, j.A. & Smith, I. 1997, 'Unification-based multimodal integration', *The 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, pp. 281-288.

[15] Krahnstoever, N., Kettebekov, S., Yeasin, M. & Sharma, R. 2002, 'A Real-Time Framework for Natural Multimodal Interaction with Large Screen Displays', *Fourth International Conference on Multimodal Interfaces (ICMI 2002)*, Pittsburgh, PA, USA.

[16] Reitter, D., Panttaja, E.M. & Cummins, F. 2004, 'UI on the Fly: Generating a Multimodal User Interface', *Human Language Technology conference 2004 / North American chapter of the Association for Computational Linguistics (HLT/NAACL-04)*, pp. 45-48.

[17] Simon, R., Wegscheider, F. & Tolar, K. 2005, 'Tool-Supported Single Authoring for Device Independence and Multimodality', *The 7th international conference on Human computer interaction with mobile devices & services*, Salzburg, Austria, pp. 91-98.

[18] Steele, R. & Khankan, K. 2005, 'Auto-Generation of Abstract Multimodal Interfaces for Mobile Web Services', *The International Journal of Wireless and Mobile Computing*.

[19] Wahlster, W. 2002, 'Smartkom: Fusion and fission of speech, gestures, and facial expressions', *The 1st International Workshop on Man-Machine Symbiotic Systems*, Kyoto, Japan, pp. 213-225.

[20] Wahlster, W. 2004, 'SmartWeb: Mobile Applications of the Semantic Web', *Informatik 2004,Annual Conference of the German Association of Computer Science*.