Chapter 3.11
A Voice-Enabled Pervasive Web System with Self-Optimization Capability for Supporting Enterprise Applications

Shuchih Ernest Chang
National Chung Hsing University, Taiwan

ABSTRACT

Other than providing Web services through popular Web browser interfaces, pervasive computing may offer new ways of accessing Internet applications by utilizing various modes of interfaces to interact with their end-users, and its technology could involve new ways of interfacing with various types of gateways to back-end servers from any device, anytime, and anywhere. In this chapter, mobile phone was used as the pervasive device for accessing an Internet application prototype, a voice-enabled Web system (VWS), through voice user interface technology. Today's Web sites are intricate but not intelligent, so finding an efficient method to assist user searching is particularly important. One of these efficient methods is to construct an adaptive Web site. This chapter shows that multimodal user-interface pages can be generated by using XSLT stylesheet which transforms XML documents into various formats including XHTML, WML, and VoiceXML. It also describes how VWS was designed to provide an adaptive voice interface using an Apache Web server, a voice server, a Java servlet engine, and a genetic algorithm based voice Web restructuring mechanism.

INTRODUCTION

Mobile phone and Internet brought us to a new era by offering a new way for person to person communication and facilitating companies and their customers in conducting business through electronic commerce (Gulliver, Serif & Ghinea, 2004; Toye, Sharp, Madhavapeddy & Scott, 2005; Roussos, Marsh & Maglavera, 2005). Because

DOI: 10.4018/978-1-60566-146-9.ch008

of the pervasive nature of empowering people to use it anywhere and anytime, mobile phone is becoming one of the most pervasive devices in the world (Chang & Chen, 2005; Ballagas, Borchers, Rohs & Sheridan, 2006). With the rapid spread of mobile phone devices and the convergence of the phone and the personal digital assistant (PDA), there is an increasing demand for a multimodal platform that combines the modalities of various interface devices to reach a greater population of users. While there is a growing demand for technologies that will allow users to connect to the Internet from anywhere through devices that are not suitable for the use of traditional keyboard, mouse, and monitor (Zhai, Kristensson & Smith, 2005), the constraints of a typical mobile device, such as small screen size, slow speed, and inconvenient keyboard, make it cumbersome to access lengthy textual information (Anerousis & Panagos, 2002). In Taiwan, the penetration rate of mobile phone (104.6%)[1] is much higher than the penetration rates of other major telecom services, including local telephone: 58.2%, Internet: 71.3%, and broadband Internet: 68.7% (Institute for Information Industry, 2007). However, the same survey also shows that the utilization rate of accessing Internet from wireless devices is relatively low, with a penetration rate slightly lower than 50%, mainly because the text-based interaction between mobile devices and Web sites is very limited. However, voice interface does not have these limitations, because voice interaction could escape the physical limitations on keypads and displays as mobile devices become ever smaller and it is much easier to say a few words than it is to thumb them in on a keypad where multiple key presses may be needed for entering each letter or character (Rebman, Aiken & Cegielski, 2003). Using voice as a medium to operate mobile devices also enables user's hands to engage in some other activities without losing the ability to browse the Internet through voice commands (Feng, Sears & Karat, 2006).

According to a study from Telecom Trends International, the number of mobile commerce users worldwide will grow from 94.9 million in 2003 to 1.67 billion in 2008, and the global revenues generated from mobile commerce are expected to expand from $6.86 billion in 2003 to $554.37 billion in 2008 (de Grimaldo, 2004). A report from ZDNetAsia states that more than half of 3G traffic would be voice and voice is still the platform on which our business is run (Tan, 2005). A study reported by the Kelsey Group claims that expenditures for speech-related services worldwide are expected to reach $41 billion by 2005 (The Kelsey Group, 2001). This report also estimates a 60-65% average annual growth rate for voice services globally by 2005, with the U.S. market expected to be 20-25% of this total. A recent example to the continuation of this trend can be illustrated by an outstanding growth (350 percent increase in quarterly revenue) of speech self-service marketplace reported by Voxify, Inc. (Market Wire, 2006). It is believed that the demand for mobile commerce has created a market for voice-enabled applications accessible by mobile phone.

Traditionally, Interactive Voice Response (IVR) systems are based on proprietary hardware and software technology, with development and deployment tightly integrated on the same hardware platform (Turner, 2004). This has resulted in high development costs. Non-portable proprietary software cannot be deployed on different platforms and it is also inherently difficult to upgrade or modify (Dettmer, 2003). A multi-modal language is needed to support human-computer dialogs via spoken input and audio output. As an optimum solution, VoiceXML (Voice eXtensible Markup Language), a markup language for creating voice-user interfaces, bridges the gap between the Web and the speech world by utilizing speech and telephone touchtone recognition for input and prerecorded audio and text-to-speech synthesis (TTS) for output (Larson, 2003). It is based on the World Wide Web Consortium's (W3C's) eX-