IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB12807

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Adaptive Web Browsing Using Web Mining Technologies for Internet-Enabled Mobile Handheld Devices

Wen-Chen Hu, Dept of Computer Science, University of North Dakota, PO Box 9015, Grand Forks, ND 58202, P 701-777-2413, F 701-777-3330, wenchen@cs.und.edu

Jyh-haw Yeh, Dept of Computer Science, Boise State University, 1910 University Dr., Boise, ID 83725-2075, P 208-426-3034, F 208-426-2266, jhyeh@cs.boisestate.edu

Hung-Ju Chu & Sheng-Chien Lee, Dept of Computer & Info. Science & Engineering, University of Florida, PO Box 116120, Gainesville, FL 32611, F 352-392-1220, {hchu, rlee168}@cise.ufl.edu

ABSTRACT

Using Internet-enabled mobile handheld devices to accessing the World Wide Web is a promising addition to the Web and traditional e-commerce. Mobile handheld devices provide convenience and portable access to the huge information on the Internet for mobile users from anywhere and at anytime. On the other hand, the traditional Web pages are mainly designed for viewing on desktop or notebook computers. They usually do not suit the devices well because the pages, especially the large text files, can not be properly, speedily displayed on the microbrowsers due to the limit of screen size, narrow network bandwidth, small memory capacity, and low computing power. This research investigates a method for large text summarization for mobile handheld devices. This method applies a Web usage mining technology—Web navigation pattern discovery and applications—to adaptive mobile Web browsing. An experimental example is given. Other related issues such as mobile commerce systems will also be discussed in this research.

INTRODUCTION

With the introduction of the World Wide Web, electronic commerce has revolutionized traditional commerce and boosted sales and exchanges of merchandise and information. Recently, the emergence of wireless and mobile networks has made possible the extension of electronic commerce to a new application and research area: mobile commerce, which is defined as the exchange or buying and selling of commodities, services, or information on the Internet through the use of mobile handheld devices. In just a few years, mobile commerce has emerged from nowhere to become the hottest new trend in business transactions. Despite a weak economy, the future of mobile commerce is bright according to the latest predictions:

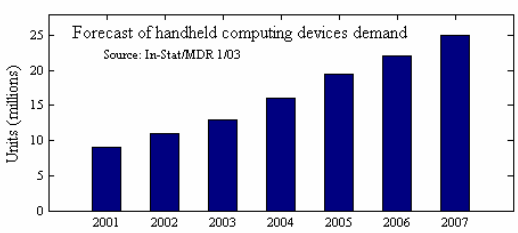
- Figure 1 shows the growth in demand for handheld computing devices (not including smart cellular phones) through 2007, as predicted by the research firm In-Stat/MDR (PalmInfocenter.com, 2003).
- It is estimated that 50 million wireless phone users in the United States will use their handheld devices to authorize payment for premium content and physical goods at some point during the year 2006 (Reuters, 2001).
- Realizing the advantages to be gained from mobile commerce, companies have begun to offer mobile commerce options for their customers in addition to the electronic commerce they already provide (The Yankee Group, 2002).

Regardless of the bright future of mobile commerce, its prosperity and popularity will be brought to a higher level only if information can be properly, speedily displayed. The traditional Web pages are mainly designed for the use of desktop or notebook computers. They usually do not suit the devices well because the pages, especially the large text files, can not be properly, speedily displayed on the microbrowsers due to the limit of screen size, narrow network bandwidth, small memory capacity, and low computing power. Therefore, loading and visualizing large text documents on handheld devices become an ardous task. This research investigates a method for large text summarization for mobile handheld devices. This method applies Web mining technologies—Web navigation pattern discovery and applications—to adaptive mobile Web browsing. This approach improves the readability and download speed of the mobile Web pages. Other related issues such as mobile commerce systems will also be discussed in this research.

MOBILE COMMERCE SYSTEMS

This section first introduces a mobile commerce system and then illustrates how it is used to perform a mobile transaction.

Figure 1. Forecast of demand for mobile handheld computing devices



System Structure

A mobile commerce system is inherently interdisciplinary and could be implemented in various ways. Figure 2 shows the structure of a mobile commerce system and a typical example of such a system (Hu, Lee, & Yeh, 2004). The system structure includes six components: (i) mobile commerce applications, (ii) mobile handheld devices, (iii) mobile middleware, (iv) wireless networks, (v) wired networks, and (vi) host computers.

1. Mobile commerce applications: Electronic commerce applications are numerous, including auctions, banking, marketplaces and exchanges, news, recruiting, and retailing, to name but a few. Mobile commerce applications not only cover the electronic commerce applications, but also include new applications, which can be performed at any time and from anywhere by using mobile computing technology, for example, mobile inventory tracking.
2. Mobile handheld devices: An Internet-enabled mobile handheld device is a small general-purpose, programmable, battery-powered computer that is capable of handling the front end of mobile commerce applications and can be operated comfortably while being held in one hand. A mobile handheld device includes six major components: (i) a mobile operating system, (ii) a mobile central processor unit, (iii) a microbrowser, (iv) input/output devices, (v) a memory, and (vi) batteries (Hu, Yeh, Chu, & Lee, 2005).
3. Mobile middleware: The major task of mobile middleware is to seamlessly and transparently map Internet contents to mobile stations that support a wide variety of operating systems, markup languages, microbrowsers, and protocols. WAP and i-mode are the two major kinds of mobile middleware. According to an article in (Eurotechnology, n.d.), 60 percent of the world's wireless Internet users use i-mode (NTT-DoCoMo, n.d.), 39 percent use WAP (Open Mobile Alliance, n.d.), and 1 percent use Palm middleware. Table 1 compares i-mode and WAP.
4. Wireless and wired networks: Wireless communication capability supports mobility for end users in mobile commerce systems. Wireless LAN, MAN, and WAN are the major components used to provide radio communication channels so that mobile service is possible.
5. Host computers: A user request such as database access or updating is actually processed at a host computer, which contains three major kinds of software: (i) Web servers, (ii) database servers, and (iii) application programs and support software.

An Example of Mobile Commerce Transaction Processing

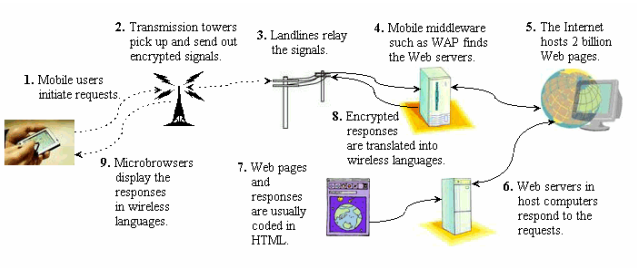
To explain how the mobile commerce components work together, Figure 2 shows a flowchart of how a user request is processed by the components in a mobile commerce system, along with brief descriptions of how each component processes the request.

1. Mobile commerce applications: A content provider implements an application by providing two sets of programs: client-side programs, such as user interfaces on microbrowsers, and server-side programs, such as database access and updating.

Table 1. Comparisons of the two major types of mobile middleware

	WAP	i-mode
Developer	WAP Forum	NTT DoCoMo
Function	A protocol	A complete mobile Internet service
Host Language	WML	CHTML
Major Technology	WAP Gateway	TCP/IP modifications
Key Features	Widely adopted and flexible	Highest number of users and easy to use

Figure 2. A flowchart of a user request processed in a mobile commerce system



2. Mobile handheld devices: Handheld devices present user interfaces to the mobile end users, who specify their requests on the interfaces.
3. Mobile middleware: The major purpose of mobile middleware is to seamlessly and transparently map Internet contents to mobile stations that support a wide variety of operating systems, markup languages, microbrowsers, and protocols.
4. Wireless networks: User requests are delivered to either the closest wireless access point (in a wireless local area network environment) or a base station (in a cellular network environment).
5. Wired networks: This component is optional for a mobile commerce system. However, most computers (servers) usually reside on wired networks such as the Internet, so user requests are routed to these servers using transport and/or security mechanisms provided by wired networks.
6. Host computers: Host computers process and store all the information needed for mobile commerce applications, and most application programs can be found here. They include three major components: Web servers, database servers, and application programs and support software.

ADAPTIVE WEB BROWSING USING WEB MINING TECHNOLOGIES

HTML Web pages usually do not suit Internet-enabled mobile handheld devices well because the pages can not be properly, speedily displayed on the microbrowsers of the devices due to

- small screen size,
- narrow network bandwidth,
- low memory capacity, and
- limited computing power and resources.

To relieve this problem, this research applies Web usage mining technologies to Web page summarization for handheld devices.

Web Usage Mining

World Wide Web data mining includes content mining, hyperlink structure mining, and usage mining. All three approaches attempt to extract knowledge from the Web, produce some useful results from the knowledge extracted, and apply the results to certain real-world problems. The first two apply the data mining techniques to Web page contents and hyperlink structures, respectively. The third approach, Web usage mining, the method used by this research, is the application of data mining techniques to the usage logs of large Web data repositories in order to produce results that can be applied to many practical subjects, such as improving Web sites/pages, making additional topic or product recommendations, user/customer behavior studies, etc.

A variety of implementations and realizations is employed by Web usage mining systems. Figure 3 shows a generalized structure of Web usage

Figure 3. A Web usage mining system structure

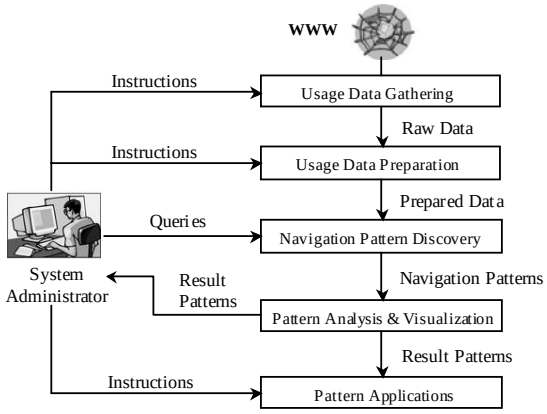
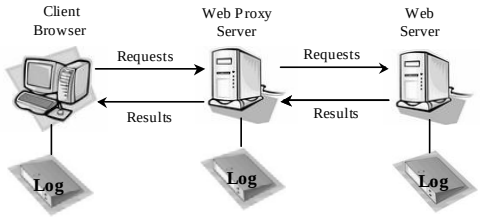


Figure 4. Three Web log file locations



mining system (Hu, Yang, Lee, & Yeh, 2005). A Web usage mining system performs the following five major tasks:

1. Usage data gathering: Web logs, which record user activities on Web sites, provide the most comprehensive, detailed Web usage data. A log file can be located in three different places: i) Web servers, ii) Web proxy servers, and iii) client browsers, as shown in Figure 4.
2. Usage data preparation: Log data are normally too raw to be used by mining algorithms. This task restores the users' activities that are recorded in the Web server logs in a reliable and consistent way. This phase should at a minimum achieve the following four major tasks: (i) removing undesirable entries, (ii) distinguishing among users, (iii) building sessions, and (iv) restoring the contents of a session.
3. Navigation pattern discovery: This part of a usage mining system looks for interesting usage patterns contained in the log data. Most algorithms use the method of sequential pattern generation, while the remaining methods tend to be rather ad hoc.
4. Pattern analysis and visualization: Navigation patterns show the facts of Web usage, but these require further interpretation and analysis before they can be applied to obtain useful results.
5. Pattern applications: The navigation patterns discovered can be applied to the following major areas, among others: (i) improving the page/site design, (ii) making additional product or topic recommendations, and (iii) Web personalization. Learning user/customer behavior and Web caching, less important applications for navigation patterns, are also worth studying.

An Example

This example shows how to apply Web usage mining technologies to adaptive Web browsing for handheld devices. Figure 5 shows an HTML page—the World Health Organization of the United Nations—which is

Figure 5. The HTML page of the World Health Organization

too large to be properly displayed on a microbrowser of a mobile handheld device. An HTML page can provide the following information:

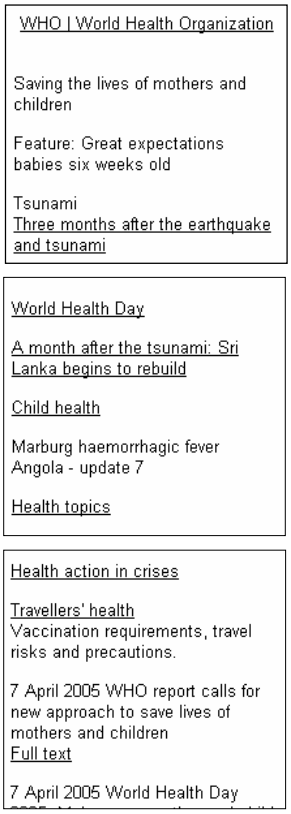
- Content: Web page content provides the most accurate and full-text information. However, it is also the least-used information for a search engine since content extraction is far less practical than other methods.
- Descriptions: Web page descriptions can either be constructed from the metatags or submitted by Webmasters or reviewers. A metatag is an HTML tag that provides information (e.g., author, expiration date, a list of keywords, etc.) about a Web page. Page descriptions are either from the metatags or submitted by Web masters or reviewers.
- Hyperlinked text: Hyperlink text is normally a title or brief summary of the target page.
- Hyperlinks: Hyperlinks contain high-quality semantic clues to a page's topic (Chakrabarti et al., 1999). A hyperlink to a Web page represents an implicit endorsement of the page being pointed to. However, exploiting this link information is challenging because it is highly noisy.
- Keywords: Keywords can be extracted from full-text documents or metatags. Before obtaining keywords from a full-text document, some filtering operations are applied to the document. Typical operations would include the removal of common words using a list of stopwords, the transformation of uppercase letters to lowercase letters, word ranking, etc. (Korfhage, 1997).
- Page structure: An HTML page is usually organized as a tree, in which top-level nodes are normally more important than lower-lever nodes.
- Page title: The title tag, which is only valid in a head section, defines the title of an HTML document. A title is usually chosen that makes sense with no context.
- Text with different fonts: Emphasized text is usually given a different font to highlight its importance.
- The first sentence: The first sentence of a Web document is likely to give crucial information related to the document, as it is usually an introduction or an abstract.

From the above HTML page information and the Web usage mining technologies, the following two major steps are used to achieve adaptive Web browsing for handheld devices:

- The first sentence of each paragraph is kept. The rest of the paragraph is ignored.
- The hyperlinks are sorted based on popularity by the Web usage mining.

Figure 6 shows three microbrowser screen dumps (one page), which are from the results of applying the proposed methods to the page at Figure 5. Note that the proposed methods are still in the very early stage of development. This is probably the best example the methods could get so far. The major reason of this—ineffectively summarizing a Web page for handheld devices—is the high complexity of Web pages, many of which are not written by humans. Instead, they are generated from

Figure 6. Screen dumps of the results after applying the proposed methods to the page in Figure 5



software such as Dreamweaver or include a variety of scripts such as Javascript and ASP (Active Server Pages).

CONCLUSIONS

It is widely acknowledged that mobile commerce is a field of enormous potential. However, it is also commonly admitted that the development in this field is constrained. There are still considerable barriers waiting to be overcome. One of the barriers is adaptive Web viewing for handheld devices. Unable to effectively viewing Web pages, mobile commerce can not be brought to a higher level. A mobile commerce system is inherently interdisciplinary and could be implemented in various ways. The system structure includes six components: i) mobile commerce applications, ii) mobile handheld devices, iii) mobile middleware, iv) wireless networks, v) wired networks, and vi) host computers.

This research applies Web usage mining technologies to adaptive Web viewing for handheld devices. Web usage mining is the application of data mining techniques to the usage logs of large Web data repositories in order to produce results that can be applied to many practical subjects, such as improving Web sites/pages, making additional topic or product recommendations, user/customer behavior studies, etc. A Web usage mining system must be able to perform five major functions: i) data gathering, ii) data preparation, iii) navigation pattern discovery, iv) pattern analysis and visualization, and v) pattern applications. This approach improves the readability and download speed of mobile Web pages.

REFERENCES

Chakrabarti, S., Dom, B.E., Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., & Kleinberg, J. (1999). Mining the Web's link structure, *IEEE Computer*, 32(8), 60-67.

Eurotechnology. (n.d.). *Frequently Asked Questions about NTT-DoCoMo's i-mode*. Retrieved October 9, 2005, from http://www.eurotechnology.com /imode/faq.html

Hu, W.-C., Lee, C.-w., & Yeh, J.-h. (2004). Mobile Commerce Systems, Shi Nansi, editor, *Mobile Commerce Applications*, pages 1-23, Idea Group Publishing.

W.-C. Hu, W.-C., Yang, H.-J., Lee, C.-w., & Yeh, J.-h. (2005). A technical perspective of World Wide Web usage mining, John Wang, editor, *Encyclopedia of Data Warehousing and Mining*, pages 1242-1248, Idea Group Publishing.

Hu, W.-C., Yeh, J.-h., Chu, H.-J., & Lee, C.-w. (2005). Internet-enabled mobile handheld devices for mobile commerce. *Contemporary Management Research*, 1(1), 13-34.

Korfhage, R.R. (1997). *Information Storage and Retrieval*, John Wiley & Sons, pages 105-144.

NTT-DoCoMo. (n.d.). *i-mode*. Retrieved October 2, 2005, from http://www.nttdocomo. com/ .

Open Mobile Alliance, (n.d.). *WAP (Wireless Application Protocol)*. Retrieved September 14, 2005, from http://www.openmobilealliance.org/tech/affiliates/wap/wapindex.html

PalmInfocenter.com. (2003). *PDA Market still Poised for Growth*. Retrieved September 21, 2005, from http://www.PalmInfocenter.com/view_Story.asp? ID=5050 .

Reuters. (2001). *The Yankee Group publishes U.S. Mobile Commerce Forecast*. Retrieved October 05, 2005, from http://about.reuters.com/newsreleases/art_31-10-2001_id765.asp .

The Yankee Group. (2002). *Over 50% of Large U.S. Enterprises Plan to Implement a Wireless/mobile Solution by 2003*. Retrieved September 28, 2005, from http://www.yankeegroup.com/public/news_releases/news_rlease_detail.jsp?ID=PressReleases/news_09102002_wmec.htm .