Chapter X

Adaptive Mobile Web Browsing Using Web Mining Technologies

Wen-Chen Hu
University of North Dakota, USA

Yanjun Zuo
University of North Dakota, USA

Lei Chen
Sam Houston State University, USA

Chyuan-Huei Thomas Yang
Hsuan Chuang University, Taiwan

ABSTRACT

Using mobile handheld devices such as smart cellular phones and personal digital assistants (PDAs) to browse the mobile Internet is a trend of Web browsing. However, the small screens of handheld devices and slow mobile data transmission make the mobile Web browsing awkward. 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. A Web usage mining system must be able to perform five major functions: (i) usage 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.
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, that is, 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. The future of mobile commerce is bright according to the following predictions:

- Figure 1 shows the growth in demand for smart mobile devices including handhelds, wireless handhelds, and smart cellular phones through 2006, as estimated by the research firm Canalys (2004a, 2004b, 2004c, 2005a, 2005b, 2005c, 2005d, 2006).
- Cumulative sales of smartphones will reach 1 billion units by the first quarter of 2011 according to IDC, a market research company (Symbian Limited, 2006).
- According to various reports, the estimated worldwide shipments of the following three equipments in 2006 are:
  - PDAs and smartphones: 84 million (Gartner, Inc., 2006),
  - Cellular phones: 986 million (cellular-news.com, 2006), and
  - PCs: 250 million (Silicon Valley Daily, 2006).

The worldwide shipments of PDAs and smartphones in 2006 had a 57% increase from the same period last year, according to Gartner, Inc. Smartphone shipments bolstered the market, growing 75.5% to reach 34.7 million units, more than four times the size of the PDA market. PDA shipments increased by 5.7%, totaling 7.4 million units. Though the unit sales are less than one tenth of the worldwide mobile phone sales in 2006, they are not too far away from the worldwide PC sales in 2006.

Regardless of the bright future of mobile commerce, its prosperity and popularity will be assured only if information can be properly, speedily displayed. Traditional Web pages are designed for desktop or notebook computers and loading and visualizing large text documents on handheld devices are often arduous tasks. They usually do not suit the devices well because the pages, especially the large text files, cannot 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 summariz-
Adaptive Mobile Web Browsing Using Web Mining Technologies

ing large text for mobile handheld devices. This method applies a Web usage mining technology, that is, Web navigation pattern discovery and applications, to adaptive mobile Web browsing. An experimental example is given. Other related issues such as mobile handheld devices will also be discussed in this research.

BACKGROUND

Mobile users interact with mobile commerce applications by using small wireless Internet-enabled devices, which come with several aliases such as handhelds, palms, PDAs, pocket PCs, and smartphones. To avoid any ambiguity, a general term, mobile handheld devices, is used in this chapter. A mobile handheld device is small enough to be held in one hand and is a general-purpose, programmable, battery-powered computer, but it is different from a desktop PC or notebook due to the following three major special features:

- **Limited network bandwidth**: This limitation prevents the display of most multimedia on a microbrowser. Though the WiFi and third generation (3G) networks go some way toward addressing this problem, the wireless bandwidth is always far below the bandwidth of wired networks.
- **Small screen/body size**: This feature restricts most handheld devices to using a stylus for input.
- **Mobility**: High mobility of handheld devices is an obvious feature that separates handheld devices from PCs. This feature also creates many new applications such as mobile recommendations that normally cannot be done by PCs.

Short battery life and limited memory, processing power, and functionality are additional features, but these problems are gradually being solved as the technologies improve and new methods are constantly being introduced. Figure 2 shows a typical system structure of handheld devices, which includes six major components: (i) a mobile operating system, (ii) a mobile central processing unit, (iii) a microbrowser, (iv) input/output devices, (v) memory, and (vi) batteries (Hu, Yeh, Chu, & Lee, 2005b). Brief descriptions of the components follow:

1. **Mobile operating systems**: They are the core software of handheld devices. Mobile operating systems are different from those

![Figure 2. A system structure of mobile handheld devices (Hu et al., 2005b)](image-url)
in desktop computers as they include the following additional features: (i) power management to prolong the battery life, (ii) real-time capability for time-critical operations such as voice communication, and (iii) wireless infrastructure for wireless communication.

2. Mobile central processing units: Mobile CPUs are the core hardware of mobile handheld devices and the performance and functionality of the devices are heavily dependent on the capabilities of the processors.

3. Microbrowsers: Microbrowsers are Internet browsers specifically designed for use in mobile handheld devices. They differ from desktop browsers in several ways, specifically the languages they use, security, footprint, and smaller windows. The last feature, smaller windows, enables them to provide simplified interfaces, although it also eliminates much of the desktop browser’s multimedia functionality, such as streaming audio and video support.

4. Input/output devices: There is only one major output device, the screen, but there are several popular input devices; in particular, keyboards and touch screens/writing areas that require the use of a stylus.

5. Memory: Three types of memory are usually employed by handheld devices: (i) random access memory, (ii) read-only memory, and (iii) flash memory.

6. Batteries: Rechargeable Lithium Ion batteries are the most common batteries used by handheld devices.

**MAIN THRUST OF THE CHAPTER**

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

- Small screen size.
- Narrow network bandwidth.
- Low memory capacity.
- 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 (Cooley, 2000). 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, and so forth.

![Figure 3. A Web usage mining system structure](image-url)
A variety of implementations and realizations is employed by Web usage mining systems. Figure 3 shows a generalized structure of a Web usage mining system (Hu, Yang, Lee, & Yeh, 2005a). 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 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 Webmasters 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, Dom, Kumar, Raghavan, Rajagopalan, Tomkins et al., 1999). A hyperlink to a Web page represents an implicit endorsement of the
Adaptive Mobile Web Browsing Using Web Mining Technologies

Figure 5. The about page of the World Health Organization

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, and so forth (Korfhage, 1997).

Page structure: An HTML page is usually organized as a tree, in which top-level nodes are normally more important than lower-level nodes.

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

1. The first sentence of each paragraph is kept because the sentence usually contains important information or an introduction about the paragraph.
2. The rest of the paragraph is ignored. By doing this, the page size can be reduced significantly.
3. The hyperlinks are sorted based on popularity from the Web usage mining. The proposed system installs a counter on each Web page. Each page visit increases the counter by one. A popular page therefore has a higher number in its counter.

Figure 6 shows three microbrowser screen dumps (one page), which are from the results.

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 three major steps are used to achieve adaptive Web browsing for handheld devices:

- The first sentence of each paragraph is kept because the sentence usually contains important information or an introduction about the paragraph.
- The rest of the paragraph is ignored. By doing this, the page size can be reduced significantly.
- The hyperlinks are sorted based on popularity from the Web usage mining. The proposed system installs a counter on each Web page. Each page visit increases the counter by one. A popular page therefore has a higher number in its counter.

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 software such as Adobe Dreamweaver or include a variety of scripts such as JavaScript and active server pages (ASP).

**FUTURE TRENDS**

Many methods are proposed for rendering Web pages for microbrowsers. This research uses Web usage mining technologies to summarize Web pages and then displays the results on microbrowsers. Other promising methods include:

- **A block importance model:** This method extracts and presents more condensed Web search results to mobile users by using a block importance model (Xie, Miao, Song, Wen, & Ma, 2005), which assigns importance values to different segments of a Web page.
- **Link analysis:** Yin and Lee (2004) use a ranking algorithm similar to Google’s PageRank algorithm to rank the content objects within a Web page. The PageRank model suggests the reputation of a page on a topic is proportional to the sum of the reputation weights of pages pointing to it on the same topic. That is, links emanating from pages with high reputations are weighted more heavily. This allows the extraction of only important parts of Web pages for delivery to mobile devices. This method provides savings in the wireless traffic and downloading time while providing a satisfactory reading experience on the mobile device.
- **Page and form summarization:** This method (Buyukkokten, Kaljuvee, Garcia-Molina, Paepeke, & Winograd 2002) breaks each Web page into text units that each can be hidden, partially displayed, made fully visible, or summarized. A variety of methods are introduced that summarize the text units. In addition, HTML forms are also summarized by displaying just the text labels that prompt the use for input.
- **Page reformatted:** This method changes page format, but not its content. Two commercial products use this method:
  - **ACCESS:** ACCESS’ NetFront browser includes smart-fit rendering technology (n.d.), which intelligently adapts standard Web pages to fit the screen width of any mobile device enabling an intuitive and rapid vertical scrolling process, without degrading the quality or usability of the pages being browsed. Concretely, the following process is performed:
    - Images larger than the screen width are scaled down to fit the screen width.
    - Tables larger than the screen width are split and laid out vertically, as shown in Figure 7.
  - **Opera:** Opera’s small-screen rendering technology (n.d.) reformats a Web page to fit it

![Figure 7. A Web page table split by ACCESS’ NetFront Browser](image)
inside the screen width and eliminates the need for horizontal scrolling. All the content and functionality is still available, it is only the layout of the page that is changed. Figure 8 shows an example of Opera’s method.

- **Page structure**: One example of this method analyzes the structure of an existing Web page and splits it into small and logically related units that fit into the screen of a mobile device. Chen, Ma, and Zhang (2005) organize a Web page into a two level hierarchy with a thumbnail representation at the top level for providing a global view and index to a set of subpages at the bottom level for detail information.

- **Visitor’s context**: This method displays the Web content based on the visitor’s context such as location, information interests, and device capabilities. For example, nearest gas stations or affordable motels are displayed based on a mobile user’s location and preferences. Pashtan, Kollipara, and Pearce (2003) propose a method adapting Web content to a user’s dynamic context for Motorola Lab’s Web-enabled museum system.

At this moment, it is almost impossible to predict which methods will prevail in the future because device or microbrowser manufacturers tend to use their own technologies.

**CONCLUSION**

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 awkward browsing of large text Web pages by using handheld devices, which include six components: (i) mobile operating systems, (ii) mobile central processing units, (iii) microbrowsers, (iv) input/output devices, (v) memory, and (vi) batteries. Unable to effectively view Web pages, mobile commerce cannot be brought to a higher level. 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, and so forth. 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**


