

## Web page adaptation for Universal Access

*Juan F. Lopez, Pedro Szekely*

University of Southern California - Information Sciences Institute  
4676 Admiralty Way, Marina del Rey, California 90292-6695, US  
juan@isi.edu, szekely@isi.edu

### Abstract

Universal Access to the Information Society requires the adaptation of the web content and information to multiple devices and users. This paper presents a generic system that allows users to view any web page on any device. The system automatically gets a web page, transforms it and adapts it to the features of the device and the preferences of the user. The objective of the system is to allow users access the same information anywhere, anytime, and on any device. The approach used is to restructure web pages (typically designed for desktops), so that they can be adequately viewed in small portable devices.

### 1. Introduction

The exploding growth of the Internet and the Information Society is creating new challenges in UI design, because of the universal access requirements for a variety of users and devices. Users expect similar capabilities and accessibility on a small screen device and on a big screen desktop.

It is estimated that by 2003 more than 50% of the Internet access will be through non-PC devices, such as Internet appliances, PDA, web enabled cellular phones and other various handheld devices. All these new portable devices have very different capabilities and features (e.g., display size and resolution, bandwidth, processing power, input/output). Maintaining different versions of a web site for each one of these new devices is labor intensive and impractical.

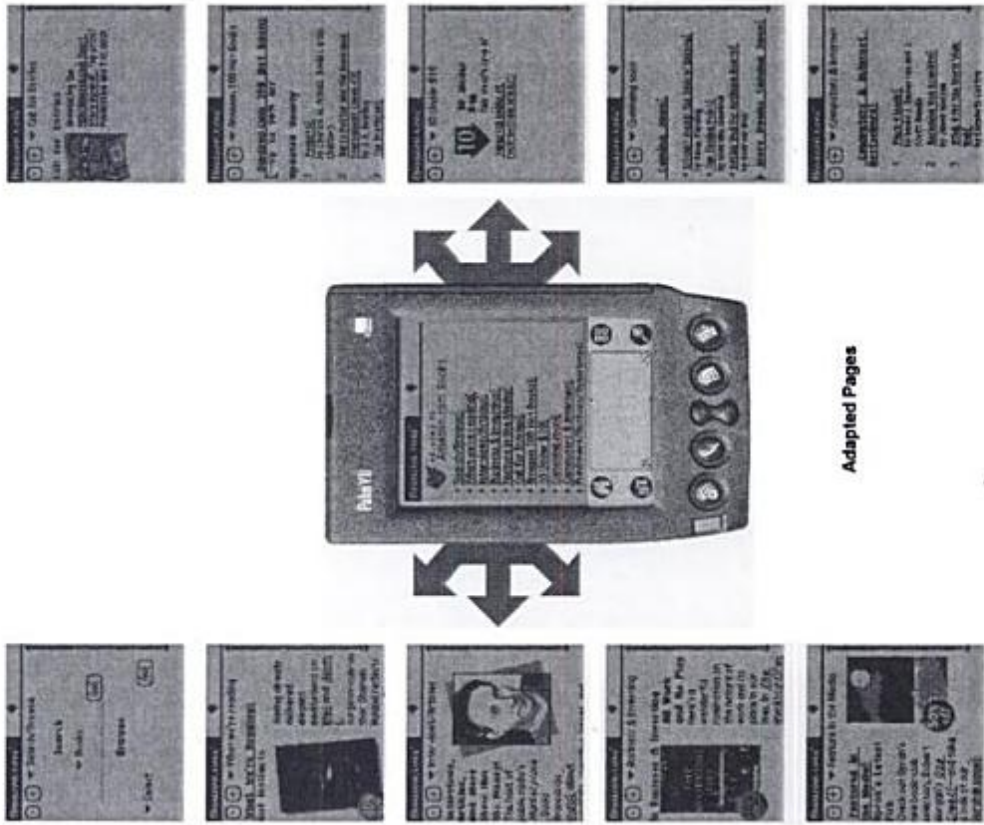
This paper presents a generic system that allows users to view any web page on any device. The system gets a web page, transforms it and adapts it to the features of the device and the preferences of the user. The objective of the system is to allow users access the same information anywhere, anytime, and on any device.

### 2. Example

The following example clarifies what we want to achieve with this system. The left side of the figure shows the input of the system: a page from Amazon books website seen with a regular PC web browser. The right side shows the desired output (several adapted pages) when the user sees the Amazon website with a Palm VII.

Both systems present and make accessible the same information to the user, adapted to the capabilities of the device. Some of the adaptations and transformations illustrated in this example are:

- Each one of the generated adapted pages contains parts of the original HTML code of the page, but adapted to the capabilities of the device (e. g., Palm VII only supports a subset of the HTML 3.2 specifications). Some of the tags or attributes are added, removed or replaced to accommodate the limitations or constraints (e. g., Palm VII does not support JavaScript, nested tables, frames, cookies or Java) [1].
- Graphics and other multimedia elements are adapted: Some graphics are scaled, or the number of colors is reduced (Palm VII supports only 2-bit depth per pixel). Other graphics are replaced by their Alt attribute, a link, or simply removed (e.g., decoration graphics).
- The top of each Palm page contains a navigation menu to enable the user to navigate through the different sections. The system also generates index pages (like the one in the screen of the Palm) with links to the different sections (generated pages). These indexes list the available information to the user. All the HTML code for the navigation and indexes is added by the system (since the navigation did not exist in the original page). Navigation mechanisms allow the user to navigate through the information (the generated adapted pages).



Adapted Pages



Original Page

Figure 1: Example

### 3. The Variety problem

The main problem to implement a system that allows universal access to the information on the web, as the one described above, is what can be called *'the variety problem'*: The system should be generic enough to manage very different classes of information, devices and users:

- Web pages may be very different in content, layout, structure, features, style, languages, standards, etc. There is a very broad variety of pages in the web: *variety of information*.
- New web enabled devices are continually appearing in the market (e.g., Internet appliances, PDA, cellular phones, handhelds) with very different capabilities and features (e.g., display size and resolution, bandwidth, processing power, input/output): *variety of devices*.
- Different kinds of users with different backgrounds, expertise, roles, skills, capabilities, navigation styles, localization, etc. try to access the same information on the web: *variety of users*.

### 4. Approach

The main elements to handle this variety problem are:

- Clean up of the original, desktop-specific HTML [2] and conversion to XHTML [3]: This makes easier to parse and work with the documents, since it is well-formed XML [4]. Transformations can be implemented using XSLT [5]. This also provides a transition path for existing HTML content. The system also accepts content in the new XML/XSL standard.
- Analysis of the document layout and structure to generate an ontology of the document: This includes layout analysis and generalization, main content detection, section classification, and multimedia elements classification (e.g., ornamental vs. information images)
- Document ontology: The document is parsed and analyzed to generate a tree with object classes and relations. Having a well designed and generic enough ontology enables the system to classify each node accurately, and establish the relations with other nodes. A generic ontology makes possible to accommodate very different classes of web pages (information).
- Classification and generation independence: Classification and Generation are independent processes, cleanly separated by the ontology. The system can use different classifiers or generators (automatic or manual) depending on the needs. Publishers can have control on the adaptation by including the ontology into the pages.
- Transformation based: The original document is adapted by a set of transformations. Transformations can be applied sequentially to the original page. Transformations can be reused for different devices or user profiles. New transformations can be easily added to the system.
- Device description and user description: By defining a set of characteristics or capabilities the system can adapt to a great variety of devices and users. Characteristics/capabilities can be added later to expand the system.
- Navigation strategy and navigation mechanisms: Enable the user to navigate through the different pages. Code for navigation may be added to each adapted page to enable the user to navigate through the different pages. The system may also generate index pages with links to the different generated pages.

### 5. System architecture

Our approach is to restructure and transform web pages (typically designed for desktops), so that they can be adequately viewed in small portable devices (automatic adaptation of desktop content for small devices). In a small display all the information cannot be shown at the same time. Normally one desktop page will be displayed as several smaller pages. It is essential to decide what to display and when to display it. Another problem is how the user can navigate between the different elements of information displayed in the page.

The new standards (XML, XSLT, and XHTML) enable us to separate information (content) from presentation (style) and to apply different transformations to the original documents. The system restructures web pages by applying a set of transformations/adaptations to the original document.

The system acts as a proxy, getting the real pages, classifying the contents, transforming the pages into a set of adapted pages, and sending the adapted pages to the client.

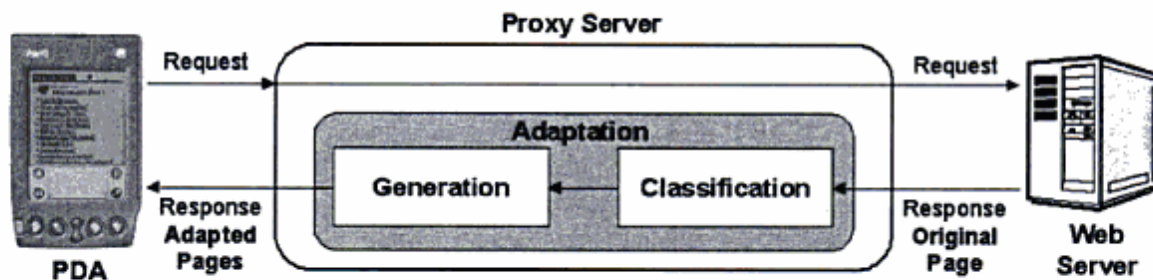


Figure 2: System Architecture

Two main components can be differentiated in the system:

- *Original Web Page Classification*: Classifies the different elements of the original page, divides the page into sections, classifies nodes, etc. Classification may be done manually or by an automatic classifier.
- *Adapted Web Page Generation*: Generates the web pages adapted to the device and user preferences, using the classification and the device and user description. The adapted pages are generated by a set of transformations.

### 5.1 Original Web Page Classification

This component of the system gets the original web page and classifies the different elements of the original page. Classification may be done manually by the publisher (special tags embedded in the HTML) or automatically by a software classifier.

The classification divides the page into different sections. The classification depends on the document structure, layout, discontinuities, fonts, styles, etc. Several types of relations (e.g., spatial, discourse, typographical, contextual, structural, semantic) can be used to segment the document.

Two basic steps are completed in this phase:

- *Original page acquisition*: The input for the system may be HTML, XHTML or XML/XSLT. If the page is HTML then the system cleans up the HTML code and converts it to well formed XHTML.
- *Document parsing and classification to generate the ontology of the page*: The ontology is a tree that represents the different sections, elements or nodes of the document, and the relations between them. In a typical web page different elements can be found: e.g., main content, navigation elements, images. To segment the document the system parses the page and identifies and classifies the different elements in the web page, and the relations between them. The ontology represents an *abstract model* of the web page.

### 5.2 Adapted Web Page Generation

This component of the system generates web pages adapted to the device and user preferences from the classification, the device and user description. To generate the adapted web pages the system does several transformations or adaptations, according to the device capabilities and user description:

- *Navigation Strategy Generation*: Depending on the ontology of the web page, the device features, and the user characteristics, the system constructs a navigation strategy. Each device is characterized by a description of their capabilities: screen resolution, number of color or gray-levels, bandwidth, processing power, standards supported (e.g., HTML, WML, XML, CGI support), multimedia capabilities, input/output, etc. Likewise the user characteristics can be described: expertise level, skills, preferences, etc.
- The navigation strategy represents how the information is presented to the user, which information can be removed, and how the user can navigate between the different elements in the page. The navigation strategy optimizes how the information is transmitted and displayed to the device-user, considering characteristics of the document, devices features and limitations, and user knowledge and preferences.
- *Adapted Web Pages Generation*: Using the navigation strategy the system generates a new page or set of pages with the necessary navigation elements. The new set of pages is a *specific view* of the ontology according to the navigation strategy. One or several sections in the ontology are transformed into one adapted page. The title of each section is used as an anchor for the navigation between different sections/pages

To create the new pages the system generates a set of XSLT transformations that are applied to the original document. These transformations include:

- **HTML Code Adaptation:** Each one of the generated adapted pages contains parts of the original HTML code of the page, but adapted to the capabilities of the device. Some of the tags or attributes are added, removed or replaced to accommodate the device limitations and constraints.
- **Layout Adaptation:** The layout and distribution of the different elements in the page is modified to fit the device screen. Each section can be represented with the full content or just the title of that section as an anchor link to that section. Each section can be displayed with different levels of detail.
- **Images and Multimedia Adaptation:** Some graphics are scaled, or the number of colors is reduced according to the device capabilities and user preferences. Other graphics are replaced by their Alt attribute, a link, a simplified version, or just removed (e.g., decoration graphics).
- **Navigation mechanisms:** Code for navigation is added to each adapted page to enable the user to navigate through the different pages. Also the system may generate index pages with links to the different generated pages.

## 6. Conclusion

In summary, the system classifies the web page according to the ontology of the document, and together with the device and user description it decides a navigation strategy. Using the selected navigation strategy the system generates and applies a series of transformations that converts the original document in a new set of web pages adapted to the device and the user.

The main features and benefits of this system are:

- Adaptive and universal access to the information on the web (browser, device and user independence). Access to the information anywhere, anytime, on any device.
- Acceptance of XML/XSL, XHTML or HTML as input (to have a transition path for existing web content).
- Transformation based: Transformations can be applied sequentially to the original page, and can be reused for different devices. New transformations can be easily added to the system.
- Ontology based: Classification and Generation are independent processes, and cleanly separated by the ontology. The system can use different classifiers or generators (automatic or manual) depending on the needs. Publishers can have control on the adaptation by including the ontology into the web pages.

## References

- [1] Palm, Inc. Web Clipping Developer's Guide. 3 November 1999. [On-line]. Available: <http://www.palmos.com/dev/tech/webclipping/>
- [2] World Wide Web Consortium. HTML 4.01 Specification. [On-line]. Available: W3C Recommendation 24 December 1999. <http://www.w3.org/TR/html4/>
- [3] World Wide Web Consortium. XHTML™ 1.0: The Extensible HyperText Markup Language. A Reformulation of HTML 4 in XML 1.0. W3C Recommendation 26 January 2000. [On-line]. Available: <http://www.w3.org/TR/xhtml1/>
- [4] World Wide Web Consortium. Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation 6 October 2000. [On-line]. Available: <http://www.w3.org/TR/REC-xml>
- [5] World Wide Web Consortium. XSL Transformations (XSLT) Markup Language (XML) 1.0. W3C Recommendation 16 November 1999. [On-line]. Available: <http://www.w3.org/TR/REC-xml>