# The "Author Once, Present Anywhere" (AOPA) Software Platform

Helen M. MENG*, P. C. Ching,** Tien-Ying FUNG*, Yuk-Chi LI*, Man-Cheuk HO*,
Chi-Kin KEUNG*, Wai-Kit LO*, Tin-Hang LO*, Kon-Fan LOW* and Kai-Chung SIU*

**Human-Computer Communications Laboratory*,
*Department of Systems Engineering and Engineering Management*,
*** Digital Signal Processing Laboratory*
*Department of Electronic Engineering*,
*The Chinese University of Hong Kong*,
*Hong Kong SAR, China*

{hmmeng, tyfung, ycli, mcho, ckkeung, wklo, thlo, kflow, kcsiu @se.cuhk.edu.hk, pcching@ee.cuhk.edu.hk}

## 1. Introduction

AOPA is a universally accessible software platform that supports Chinese Web content development for displayless voice browsers, mobile mini-browsers and regular Web browsers in E-business services provision. Universal accessibility refers to accessibility through displayless voice browsers for telephones or for the elderly / visually impaired; mobile mini-browsers for Internet-ready phones and PDAs, and regular Web browsers for desktop PCs. AOPA enables Web content/service providers, ISPs, and ASPs to author and maintain a single content repository, whose content automatically adopts usability-optimized presentation styles to reach the client devices of diverse form factors. Web visitors using mobile handhelds or telephones will outnumber those using desktop PCs within three years. Universal accessibility enables information dissemination to a much wider audience. This is critical and beneficial to B2B/B2C E-commerce, M-commerce and voice-enabled E-commerce.

AOPA eliminates major redundancies and inefficiencies in a common practice to achieve multiple accessibility – the same content is re-authored for every alternative form factor in client devices. AOPA leverages emerging standards from the W3C to decouple content specification (in XML) from presentation specification (in XSL). Platform components include: (1) An HTML-to-XML transcoder to process existing Web content. (2) Reference XSL stylesheets encoding usability-optimized presentation styles for specified content sources and client devices. (3) The first reference implementation of Chinese VXML to enable displayless, Cantonese voice browsing, by integrating with Chinese text processing, Cantonese speech recognition and synthesis technologies developed in The Chinese University of Hong Kong. The overview of AOPA platform is illustrated in Figure 1.

This paper reports on our progress in developing the HTML-to-XML transcoder, authoring reference stylesheets, developing core speech technologies, and the use of various markup language and standards in the AOPA platform.

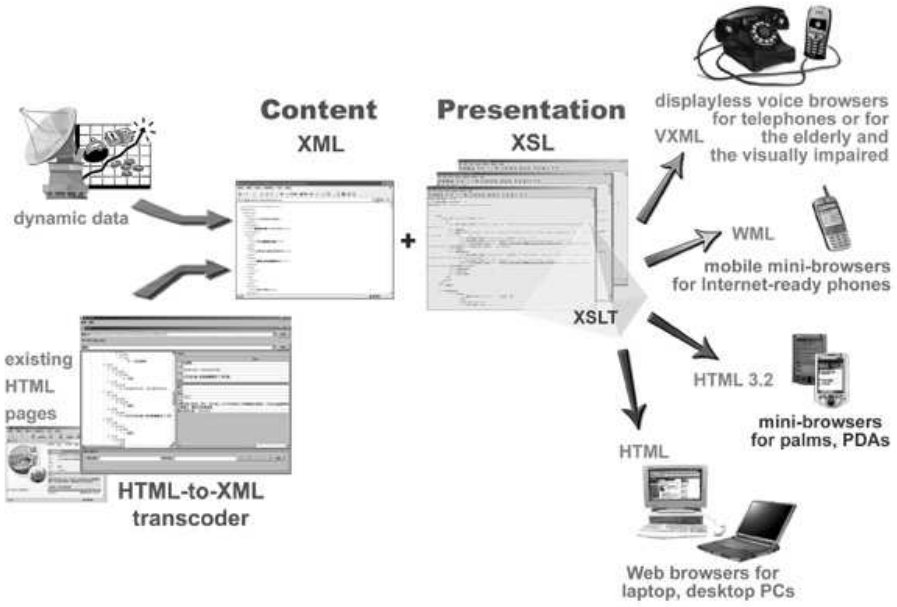


*Figure 1. An overview of the AOPA platform. The single, unified XML repository can have its content displayed with usability-optimized presentation styles on client devices of diverse form factors.*

## 2. CU[1] HTML-to-XML Transcoder

The CU Transcoder is a software toolkit for HTML-to-XML transcoding. Web content is conventionally described in terms of HTML that ties content with presentation. This implies that the presentation style of a content repository cannot be easily customized for different usage contexts without duplication of the content itself. AOPA achieves universal accessibility by leveraging W3C's XML and XSLT technologies to decouple content and presentation style. HTML-to-XML transcoding is a sequential process that includes: (1) robust HTML parsing; (2) locating selected content and (3) structuring the selected content. The steps are elaborated in the following.

**Step 1: Robust HTML Parsing**

Users can issue HTTP requests to obtain their desired HTML page(s). These are analyzed by the robust HTML parser. Analysis begins with a "sanitization" step that automatically checks through the original HTML page and corrects for improper elements such as missing tags or spurious HTML tags. Hence the HTML parser is robust to minor errors and inconsistencies in the original HTML page. Analysis continues by the use of W3C's Document Object Model [1] to parse the HTML page into a *tree* structure (see Figure 2). The tree representation eases subsequent processing such as content location and extraction.

[1] *CU* abbreviates Chinese University of Hong Kong.

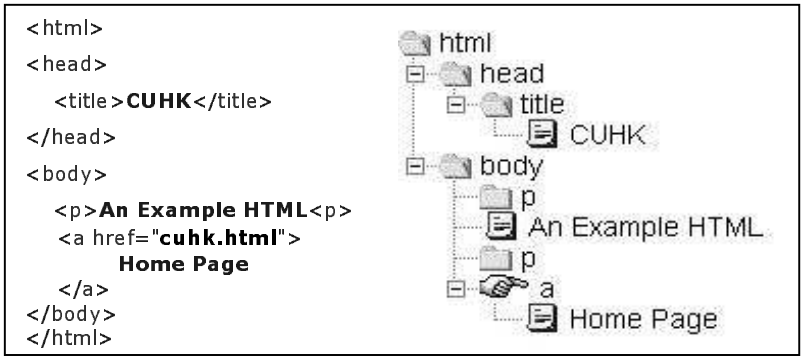


*Figure 2. An example HTML page (inset left) and its parse tree structure (inset right).*

**Step 2: Locating Selected Content**

Users can locate desired content based on the HTML tree representation by simple drag-and-drop actions on the CU Transcoder GUI (see Figure 3) which is designed with usability considerations. The location of the selected content is captured in the form of a path (as also known as the HTPath), which is designed with reference to W3C's XPath [2]. While XPath is used to locate selected content in an XML document, HTPath is for selected content in an HTML document. HTPath has the advantage of *reusability* – it can be applied to identically-structured pages, such as the hundreds of weather information pages for the world cities.

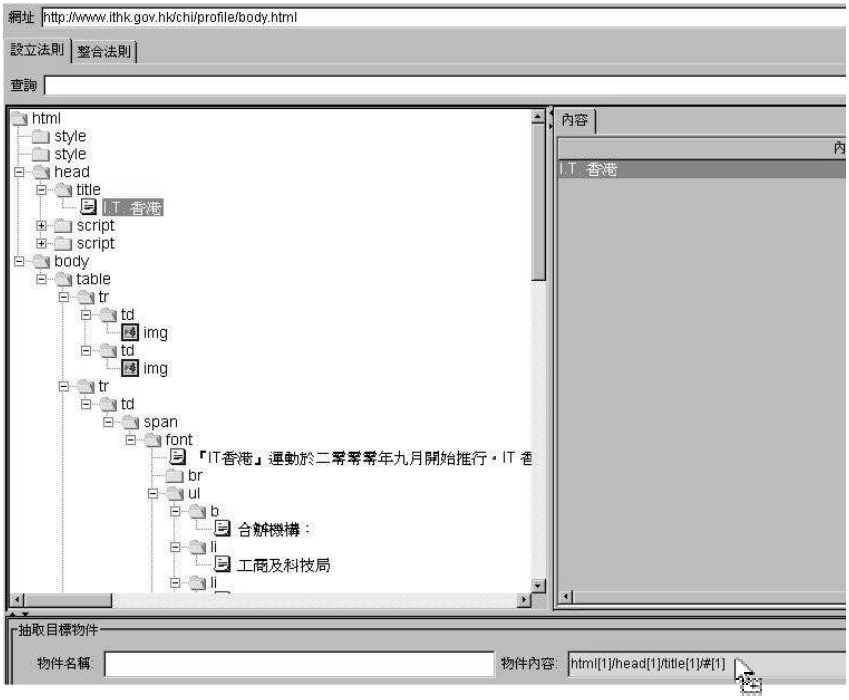


*Figure 3. The Graphical User Interface of the CU Transcoder. The selected content (I.T. 香港) is reflected by the HTPath (html[1]/head[1]/title[1]/#[1]).*

**Step 3: Structuring Selected Content**

We have defined and developed the Content Management Markup Language (CMML) with reference to W3C's XML Schema 1.0 [3]. Users can cast the selected content into *tagged* CMML structures. Table 1 compares the tag semantics between the XML schema and CMML. Similar to the HTPath locators mentioned above, user-defined CMML scripts can also be re-used for identically-structured HTML source pages in generating the output XML.

| **Tags** | **XML Schema** | **Semantics in CMML** |
|---|---|---|
| <element> | This tag declares a new XML element | Same as XML Schema |
| <attribute> | This tag declares an attribute for an XML element. | Same as XML Schema |
| <complexType> | This tag declares a complex type element. A complex type element is an XML element that contains other elements and/or attributes. | Same as XML Schema |
| <sequence> | This tag requires elements in the group to appear in the specified sequence within the containing element. | Similar to XML Schema. The declared sequence of child elements in CMML over-rides the sequential order of the corresponding child elements in the source HTML. |
| <choice> | This tag allows only one element in the group to appear within the containing element. Typically, the attribute *maxOccurs* are set to 'unbounded' to allow any number of child elements to appear without any specific order. | Similar to XML Schema. The sequence of the elements is not specified. The sequential order of the corresponding elements in the source HTML is used as default. |

*Table 1: Comparing the semantics of tags between the XML Schema and CMML.*

Users can implement each step via the CU Transcoder's GUI, which is designed for *ease-of-use*, e.g. drag-and-drop actions on the HTML tree for content location and selection as well as defining and editing the CMML to re-structure the resulting XML content. A second advantage of the CU Transcoder is *re-usability* – HTPath locators can be re-used in content location from similarly structured HTML pages and CMML scripts can be re-used for similarly structured content to produce output XML. In addition to the two advantages mentioned previously, the CU Transcoder implementation can also generate XML output that stores multilingual and multimedia content. Multilinguality is represented by Unicode and multimedia data (e.g. image, audio, binary data from WORD documents and PDFs, etc.) are encoded as base64 strings. The output XML forms the single, unified content repository. The content can be customized for various usage contexts using usability-optimized presentation stylesheets. This is described in the following section.

## 3. References Presentation Stylesheets

Content in the single, unified XML repository are automatically tailored for usability-optimized presentation in client devices with different form factors, e.g. PDAs, Internet-ready phones and regular Web browsers. Usability-optimized presentation styles are specified in terms of XSL stylesheets. We use the ITHK official website (http://www.ithk.gov.hk) to demonstrate the capabilities of the AOPA technologies. All content from the ITHK website has been extracted and transcoded from HTML into XML by means of the CU Transcoder. We have developed and released three reference stylesheets that correspond to three different form factors in client devices:

(1) Personal Digital Assistants (PDAs):

- Fujitsu Pocket LOOX (OS: PocketPC2002, screen size: 320x240 pixels, 5.3x7.6cm)
- O2 XDA (OS: PocketPC2002, screen size 320x240 pixels, 5.3x7.6cm)

(2) WAP phones:

- Nokia 3650 (OS: Symbian, screen size: 176x208 pixels, 3.4x4.1cm)

(3) Desktop PCs (OS: Microsoft Windows, screen size: 1024x768 pixels, 30.5x23.3cm)

As new devices with new form factors emerge, new stylesheets can be released for the AOPA platform to achieve universal usability.

We have defined a set of general strategies that enhance information visualization on client devices with various form factors. These strategies are applied to the design of our reference XSL stylesheets and are listed as follows:

(1) ***Pagination:*** Automatic pagination divides a lengthy page into multiple shorter pages. This is important for visual display on devices with small screens, e.g. the Pocket PCs and WAP phones.

(2) ***Avoid horizontal scrolling:*** Multi-column HTML tables are re-formatted to single- or double-column tables in order to fit on small screens as well as to minimize the amount of horizontal scrolling needed. Users only need to scroll vertically. This enhances the ease of navigation.

(3) ***Abbreviate dates:*** Long Chinese dates on the ITHK source HTML pages are shorten automatically to Arabic dates to tailor for small screen displays (see Figure 4).

(4) ***Selective image display:*** The presentation stylesheet specifies a set of (less important) images in the original webpage to be removed for display on small screens. In their place the stylesheets specifies the display of a textual position holder derived from the *alt tag* of the image. As such the stylesheet aims to generate simple and clear information display for small devices. Fewer images imply shortened page transmission times over wireless networks.

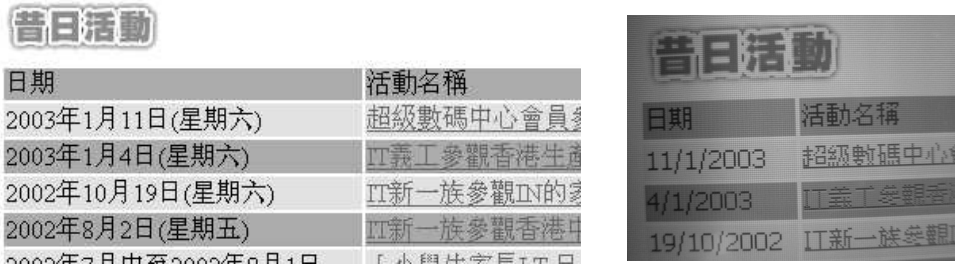


*Figure 4. In the first column of the table, dates are displayed in its full Chinese format for desktop PC (left). For Pocket PC, dates are shortened to Arabic dates to minimize space required (right).*

(5) ***Image processing:***

- ***Automatic image re-scaling*:** Pictures are re-scaled dynamically (with locked aspect ratio) to fit different form factors. To implement this strategy, all the pictures of the demo website are stored in the form as *base64 strings* in an image server. Upon user request, the server adjusts the image with parameters (e.g. *width*, *quality* and *format*) specified by the client and render an image accordingly. Thus picture re-scaling is achieved by specifying different values for the "*width*" parameter in the XSL presentation stylesheet for different devices. For example, the ITHK logo can be re-scaled to half of its original size (see Figure 5 for the original image size on the left and the resized image on the right) when the associated image parameter "*width" is set to 0.5*. This strategy helps users to see the whole picture easily without scrolling back and forth.


*Figure 5. The ITHK logo in its original scale (left) and re-scaled to half of its original size (right)*

- ***Automatic conversion of picture format*:** The AOPA image server supports two image formats: *jpeg* and *wbmp*. *jpeg* is used for desktop PCs and Pocket PCs, *wbmp* format is used for WAP phones. Pictures are automatically converted to a format compatible with the client device, as specified in the XSL presentation stylesheets. Figure 6 shows an example of the original ITHK logo in *jpeg* format and the same logo converted to *wbmp* format.


*Figure 6. ITHK logo in jpeg format (left) and the same logo converted to wbmp format (right)*

- ***Picture dithering*:** Picture dithering is useful for displays that do not support gray-scale. Dithering uses different patterns of black and white pixels to create shades of gray. The AOPA image server renders dithered images as a default display for WAP phones.
- ***Automatic resolution downgrading:*** Should a client device be constrained due to limited communication bandwidth, the quality of an image may be downgraded automatically for faster streaming. The quality of a picture is adjusted by the image server with the *jpeg quantization algorithm*. It down-samples an image by removing high frequency pixels with a quality coefficient defined by mathematical methods. This strategy shortens download times and reduces storage space for the images.

## 4. Speech Interface Technologies

In order to support universal accessibility, the AOPA platform needs to automatically customize content for display on client devices of diverse form factors. The previous sections present information visualization on small and mobile client devices with limited screen sizes, the AOPA project needs to also consider client devices *without* screen displays, such as the conventional telephone. This is a widespread screenless device for information access, and presents the challenge of speech-based, audio-only interactions. To support speech output from the device, we have developed a Cantonese speech synthesizer, to which refer as CU VOCAL. To support speech input to the device, our team at the Chinese University of Hong Kong has developed a Cantonese speech recognition engine, to which we refer as Recognition Software Building Blocks (CU RSBB).

### 4.1 CU VOCAL

CU VOCAL is a Cantonese text-to-speech synthesis engine which accepts free-form Chinese text (in BIG5 encoding) as input and generates Cantonese speech as output. CU VOCAL generates highly intelligible and natural synthetic speech using a syllable-based concatenative approach [4][5]. CU VOCAL is available as a standalone dynamic-linked library (DLL) which can be

invoked by other programs easily. We have also developed a version of CU VOCAL with a standardized API to enable system developers to easily integrate CU VOCAL into their voice-enabled applications. SAPI (Speech Application Programming Interface) [6] is a standardized API that runs on the Microsoft platform, and a SAPI-compliant version of CU VOCAL has recently been developed. Additionally, we have also developed CU VOCAL into a Web service [7] – this is a new, highly interoperable platform that allows publication of the functionalities of software on the Web. Web services can invoke one another at runtime through Internet.

### 4.2 CU RSBB

CU RSBB is a collection of software building blocks that include a core engine for Cantonese automatic speech recognition as well as tools for rapid prototyping and development of recognizers with domain-specific vocabularies to suit different applications. The latest release of CU RSBB aims to recognize entries from a word list (or vocabulary) defined by the developer. CU RSBB is also available as a DLL, a SAPI-compliant engine and as a Web service.

## 5. Markup Languages and Standards

The AOPA software platform generates different markup languages to cater for a variety of browsers on different client devices. The complete listing is shown in Table 2. In particular, the VoiceXML (VXML) is a W3C standard that supports information access over the conventional telephone via a spoken dialog interaction. The Speech Application Language Tags (SALT) is a relatively new markup language proposed by the SALT forum members [8]. SALT supports not only spoken dialog interactions over the telephone, it also supports multimodal interactions on a desktop PC. The following presents more details about VXML and SALT.

| **Client Devices** | **Content Display Standards** |
|---|---|
| Desktop PC | HTML |
| Pocket PC | HTML3.2 |
| WAP Phone | WML |
| Telephone | VXML, SALT |
| Multimodal on Desktop PC | HTML with SALT |

*Table 2. Complete list of standard markup languages for content display on various client devices*

### 5.1 VoiceXML (VXML)

VXML is a W3C standard markup language [9] for creating spoken dialog interfaces. Applications written in VXML that runs on a telephony server with a VXML “interpreter” (also known as a “voice browser”) can be accessed by a telephone client. We have integrated

CU VOCAL and CU RSBB into an open-source VXML interpreter developed by Scansoft [10]. The integration is based on DLL invocation to obtain a Cantonese-enabled VXML interpreter. Thereafter Cantonese voice-enabled applications can be developed by authoring textual contents with standard VXML tags. Similar to the other markup languages mentioned above, a VXML document can be dynamically generated from XML documents and XSL stylesheets. This extends the accessibility attained by the AOPA platform.

### 5.2 Speech Application Language Tags (SALT)

SALT is another speech interface markup language that extends existing Web markup languages. In addition to telephony access to Web content, it also enables multi-modal accessing webpages [8]. Multi-modal accessing allows multiple modes of interaction in accessing Web contents. For example, users can choose to use voice or keyboard for input, and obtain output information via speech and graphical displays at the same time. A SALT-enabled webpage can be authored with the use of SALT tags and interpreted with a SALT-enabled browser. The SALT tags facilitate easy invocation of the speech synthesis and recognition engines for accessing Web content. We have integrated the SAPI-compliant CU VOCAL and CU RSBB engines with SALT. This enables multi-modal access to Chinese Web contents where users can browse a webpage with the help of Cantonese speech. Such multi-modal accessibility benefits the elderly and the visually impaired people to access information on the Web.

## 6. Summary and Conclusions

AOPA enables Web content/service providers, ISPs, and ASPs to author and maintain a single content repository, which automatically adopts usability-optimized presentation styles to reach the client devices of diverse form factors. We have developed the CU Transcoder, a software toolkit for generating XML documents from existing Web content (in HTML). We have authored a set of reference XSL stylesheets for different devices so that the generated XML documents can be presented with usability-optimized styles to fit different form factors. For information delivery in displayless devices, we have developed two speech components: CU VOCAL (Cantonese text-to-speech synthesizer) and CU RSBB (Cantonese speech recognizer) for the AOPA platform. We focus on two standards, VXML and SALT, to facilitate Cantonese voice browsing and multi-modal accessing. We believe that the AOPA software platform offers versatility in the development of multi-accessible websites that has the potential of greatly enhancing the efficiency of information dissemination.

## 7. References


[1] Document Object Model (DOM) Level 1 Specification, http://www.w3.org/TR/REC-DOM-Level-1/

[2] XML Path Language (XPath) Version 1.0, http://www.w3.org/TR/REC-DOM-Level-1/

[3] XML Schema Part 1: Structures, http://www.w3.org/TR/xmlschema-1/

[4] Fung, T. Y. and Meng, H. Concatenating Syllables for Response Generation in Domain-Specific Applications. Proc. of ICASSP, Istanbul, 2000.

[5] Meng, H. et. al. CU VOCAL: Corpus-based Syllable Concatenation for Chinese Speech Synthesis across Domains and Dialects. Proc. of ICSLP, Denver, 2002.

[6] SAPI5.1 http://www.microsoft.com/speech/

[7] Meng. H. et. Al. CU VOCAL Web Service: A Text-to-speech Synthesis WebService for Voice-enabled Web-mediated Applications. Proc. of WWW, Budapest, 2003.

[8] SALT forum, http://www.saltforum.org/

[9] VoiceXML forum, http://www.voicexml.org/

[10] Scansoft VXML interpreter, http://fife.speech.cs.cmu.edu/openvxi/