

# Adapting Web Content for Telephone Users by transcoding XSLT

Mduduzi E. Nxumalo and Daniel Mashao  
Department of Electrical Engineering  
University of Cape, Rondebosch, 7700, South Africa  
mdu@crg.ee.uct.ac.za, daniel.mashao@ebe.uct.ac.za

**Abstract**—the goal of ubiquitous computing is to make information access available anytime, anywhere using any device. However, this comes at the expense of hard work from developers, producing multiple variants of the same information to cater for different user contexts. Different transcoding techniques based on annotating HyperText Markup Language (HTML) have been proposed. This paper proposes transcoding the Extensible Style Sheet Language (XSLT) which transforms back-end Extensible Markup Language (XML) documents into HTML or XHTML for presentation on the Web, to produce the XSLT which transforms the same XML documents to VoiceXML for presentation to telephone users.

**Index Terms**— User Interfaces, Internet, Communication channels.

## I. INTRODUCTION

The Web is arguably the most efficient medium where those with access to computers and the Internet can access information any time. However statistics which were published in September 2006 by [1] estimated that 10.3% of the population of South Africa had access to the Internet. This indicates that most people in South Africa still do not have access to the Internet.

The emergence of devices like Personal Digital Assistants (PDAs), cellular phones and telephones which can access the Internet content promises to make information accessible to anyone, anywhere, anytime, using a device of choice.

However content requires to be written in mark languages suitable for all these devices. XHTML [2] makes content accessible to bigger devices like desktop computers. The lighter version of XHTML called XHTML Mobile makes content accessible to small devices like cell phones and PDAs.

VoiceXML [3] is a markup language used to make speech interfaces accessible to telephone users. While websites are accessed by keying in a Universal Resource Locator (URL), VoiceXML applications are accessed by dialing a telephone number [4], as if you were making a normal telephone call.

Text-to-speech (TTS) tools, a VoiceXML Interpreter and a Telephony Server are needed to host VoiceXML applications. Human recorded audio version of text is better accessed by users compared to synthesized speech because of limitations in the current technology. This makes it more difficult to make content accessible to telephone users.

If the same information written in eleven South African languages is made accessible from three categories of devices, then it should be duplicated eleven times for each device, leading to at least  $11 \times 3 = 33$  duplicates of the same information. This makes it expensive for the government and companies to make information accessible via the Internet and telephones.

The aim of this research is to make it possible for telephone users to access already existing Web content. This paper discusses technical details of producing a VoiceXML based Interactive Voice Response application by transcoding XSLT which transforms back-end XML documents for presentation in the Web.

## II. RELATED WORK

There have been work in creating resources which adapt to different devices. The InfoPyramid [5] framework proposed creating resources of different modalities (e.g. having audio alternative of text) and with different quality to cater for small devices. The XML based framework which defines the adaptability of resources was proposed by [6]. The problem has been to create interfaces which make this information accessible.

Model-based [7] design is a way of designing interfaces for multiple platforms, beginning by planning an interface for each platform. The model for each platform includes information about content to be accessed and about interaction between the user and the system. The problem with this approach is that it requires more work. The alternative to this approach is to design one interface and use transcoding tools to adapt it for other platforms.

“Transcoding is a method for translating one type of code (e.g. HTML) into a different type (e.g. VoiceXML)” [8]. Transcoding has been used to make Web content accessible to mobile devices [9], to blind users [10, 11] and to old people [12].

Annotations [11] are machine understandable descriptions of Web content. They can be embedded within Web content (HTML) or written in a separate file i.e. external annotations. In the context of Semantic Web and transcoding, annotations give hints on how to adapt Web content for alternative access. A more in depth discussion on annotation based transcoding can be found in [13].

There has been work in making the job of creating annotations easier. Kouroupetroglou, et. al. [11] proposed a framework for content engineering where there are people who write annotations, using terms defined in an ontology by other people. Transcoders take both annotations which give hints on how to adapt content and use the vocabulary from ontology to understand these hints (annotations). This creates a community of people working on making the Web accessible to many people, including the blind.

Research has come up with annotation based transcoding of HTML to VoiceXML. Matching of VoiceXML to HTML elements is discussed by [14]. Web Transcoding Publisher (WTP) [15] is a proprietary tool from IBM which converts HTML to VoiceXML. WTP gets hints on how to transcode from instructions defined in the Annotation Language in [16].

However transcoding Web content for voice access is not only the matter of matching HTML to VoiceXML [8]. This is because the Web is designed to be accessed visually, while telephone users (TU) access it non-visually. Web content can be optimized for presentation in the Web by using different colors, formatting e.g. making it bold or using pictures. This optimization can make content into groups e.g. navigation menu in the top, advertisements in the left and right, and main content in the middle [8]. These groups of information can be easily identified by Web Users. However these groups of information cannot be easily conveyed to users who use alternative access [12].

Telephone users rely on what the system reads to them, they cannot go straight to access the information they want as if were surfing the Web. Different navigation techniques which divide the Web page into small segments of information so that it can be better presented to the user have been proposed [8]. Other challenges involve rendering complex HTML tables and forms in a non-visual manner [17] and difficulties in inputting and outputting speech [18].

Takagi et.al. [19] worked on improving voice browsing by making users access important information first and by inserting text which helps the user to “see” different sections of the Web page and different pages of the website. Shao et. al. [8, 20] improved usability of transcoded VoiceXML interfaces by relating Telephone Browsing to Web Browsing, where users are able to use forward and back buttons to move back and forth. Similar work by [11] gives users a choice to browse the non-visual Web, paragraph by paragraph and even line by line. There is also work in ontology based transcoding of Web content for voice access (see [10, 21]).

We identified the following problems with transcoding HTML to VoiceXML:

1. Annotation authors should have knowledge about HTML of the website being transcoded. The problem is that websites often have complex HTML mixed with Java Script and the Cascading style sheet. This makes it difficult for annotation authors who are not the authors of the website.
2. Converting HTML to VoiceXML is repeated even when HTML documents with the same node tree are transcoded. This step is repeated more often if we transcode the same content written in different languages.

This paper discusses the XSLT based transcoding system which distributes the work of creating annotations to the XSLT authors and the content authors.

### III. TRANSCODING XSLT

This section discusses the architecture of the transcoding system which produces VoiceXML from an existing XSLT document which transforms documents to HTML (XSLT of HTML). The transcoding process is divided into two independent steps: Step 1 and Step 2.

Step 1 uses annotations about XSLT of HTML to convert HTML elements to VoiceXML. Its results are the XSLT document with VoiceXML elements instead of HTML (called XSLT of VoiceXML). Step 2 uses the stored XSLT of VoiceXML to transform any given source XML document to VoiceXML. Step 2 is guided by annotations of the given XML document. This transcoding process is depicted in Fig. 1.

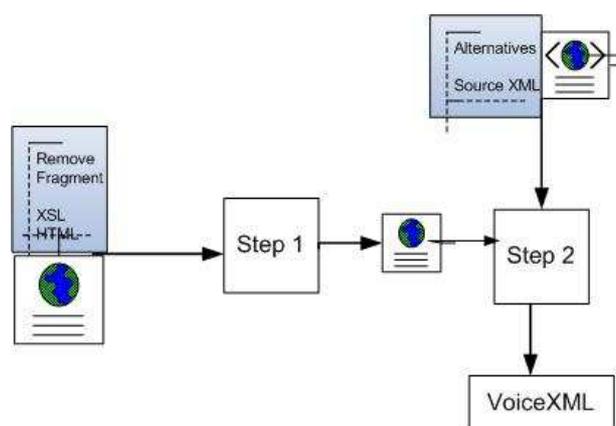


Fig. 1. The architecture of a XSLT based transcoding process is divided into two independent steps: Step 1 and Step 2. Step 1 uses annotations to convert XSLT of HTML to XSLT of VoiceXML. Step 2 uses annotations to transform XML documents to VoiceXML.

#### A. Producing XSLT of VoiceXML

##### 1) Annotations

The input to Step 1 of the transcoding process is XSLT which transforms a given XML document to HTML. This XSLT is written in the context of a user who will browse the website using a visual Web browser like Internet Explorer.

This Step uses annotations which are written by the author of XSLT to change the context to a person who browses the same information using a telephone. These annotations serve the same purpose as in [15, 20] but are written in the context of the XSLT style sheet, not HTML. The differences are discussed in section IV. The results are XSLT which is stored and used to transform any compatible XML documents to VoiceXML. The annotations are used for three main purposes as discussed next.

First, annotations identify sections of the website that are not suitable for voice access; for example pictures or less important content like advertisements. These sections are excluded in the resulting interface. Second, annotations identify alternative resources to be rendered in the results e.g. providing text which can be rendered instead of an image.

Third, voice browsing of the web is done linearly i.e. the website is read from top to bottom and left to right. For better voice browsing experience, the website is divided into fragments that can be accessed individually from the menu. This is discussed in the next section.

## 2) *The Transcoding Process*

Dividing the website into fragments is more difficult in XSLT. Because it is not easy to see how templates call each other. For example [15] uses a policy that HTML heading tags (e.g. h1, h2) indicate the end of the current fragment and the beginning of a new fragment.

This can not be easily achieved because it requires creating and closing VoiceXML forms which are linked to the VoiceXML menu. The first problem is that we need to create form names and link these forms to a Choice element in the menu. This cannot be easily achieved using XSLT as a programming language. The second problem is that output of the XSLT style sheet does not depend on the linear sequence of templates. It depends on how templates call each other. This makes it possible to have fragments which start in one template and end in the other.

The solution is to call the Transcoder as an external application in the resulting XSLT. This application is called in Step 2 and has two main responsibilities. First, it manages the creation of fragments. Second, it makes it possible for the author of the source XML document to contribute with annotations which can change the resulting VoiceXML.

### *B. Producing VoiceXML*

Step 2 of the transcoding process uses XSLT from Step 1 to transform a given XML document to VoiceXML.

### *1) Annotations*

XSLT can be created from knowing the schema of XML documents it transforms. This makes it possible for the author of XSLT to work independent of the author of XML documents. The results of this are that people with different expertise and knowledge can collaborate to make information systems.

The second transcoding process is guided by annotations written by the author of the XML document to be transformed to VoiceXML. These annotations give two different kind of information. The first one is meta-information about information in the nodes of the XML document. This information is about the language (e.g. English or Zulu) in which this content is written. They also provide information about already existing audio alternatives of the XML nodes. For example there can be an audio version (registration\_procedure.wav) of text in a node called minor/registration\_procedure. This audio version may be of higher quality than the one which would be synthesized by a Text-to-speech tool. The Transcoder renders the audio version and ignore the text version. More information about creating information written in different languages and formats is discussed by Smith [5] and by Nivele [6].

The second information given by the annotations identify nodes in the XML document that are not suitable to be accessed by a telephone user. These nodes are excluded in the resulting VoiceXML application.

The next section describes how these annotations are related to the templates in the XSLT style sheet.

### *2) The Transcoding Process*

This transcoding process uses the XSLT Processor and the Transcoder to produce VoiceXML from XSLT from Step 1.

XSLT transforms each node in the source XML tree by applying transformation rules defined in the `xsl:template` whose match attribute corresponds to the path of the node. These rules were initially created with the help of annotations from the author of XSLT in Step 1. Step 1 made it possible to modify these rules by calling the Transcoder in the beginning of each template. The Transcoder was given the list of ancestor nodes of the current node (ancestor-or-self) and the position of the current node. It uses this information to construct the XPath expression of the node in the XML document being processed by the XSLT Processor.

The aim of Step 2 is to make it possible to use annotations from the author of the source XML document to modify template rules. The annotations are given to the Transcoder as an external document. The annotations use XPath expressions to point to specific nodes in the XML document they give information about.

The Transcoder is able to relate annotations to templates since it is called in each template and it knows the XPath expression of the node transformed by the template it is in

and annotations identify the nodes they have information about using XPath expressions.

#### IV. EVALUATION

The advantage of transcoding HTML is that transcoding can take place any where; independent of the existence of the XSLT which produced HTML. XSLT may not be in the client side. This makes the discussed technique to be more suitable for transcoding in the server side where we still have access to XSLT and have information about existing resources in the server.

The advantage of transcoding XSLT is that results from the first transcoding process are the XSLT of VoiceXML which can be stored and maintained separately. This XSLT can be used to transcode the same content written in different languages.

We converted two already existing Web pages of the South African government website. We found it easier to deal with repeated sections in the Website if we annotate XSLT.

Web pages of the same website are usually similar, with certain parts appearing to all pages, e.g. the same menu bar on top and advertisements in the left. Some parts can appear more than once in a page e.g. a navigation menu in the top and in the bottom of the page. Developers can write the repeated section of code once in a separate file and use a scripting language like php to include that section where it is needed.

It may not be easy to identify elements which hold these sections in a complex HTML document because these sections may not be identifiable as a single unit. These sections can be better identified in XSLT because XSLT defines a Processing Instruction (PI) [22] element which writes a processing instruction node to the output document. The ID attribute of Processing Instruction nodes can be used to avoid repeated transcoding of the same section of code if transcoding takes place in the server side.

#### V. CONCLUSION

This paper discussed the architecture of the transcoding system which produces from XSLT of HTML. The transcoding process is divided into two steps. The first step is guided by annotations written in the XSLT author's perspective and produces the XSLT style sheet which transforms XML documents to VoiceXML. The second step uses XSLT from the first step and annotations which are written by the author of the XML document to be transformed. The content author's annotations identify already existing audio alternatives of text.

This architecture makes it easier for different people to collaborate in making multi-lingual web content accessible to telephone users.

Writing annotations which guide the transcoding process is time consuming. Future work will look at incorporating an annotation tool which will help to analyze and explore HTML.

#### ACKNOWLEDGMENT

This research is made possible by financial contributions from The National Research Fund (NRF), The University of Cape Town (UCT) and the UCT Centre of Excellence. Great gratitude from the authors goes to these sources.

#### REFERENCES

- [1] "Internet Usage Statistics For Africa," 2006; <http://www.internetworldstats.com/stats1.htm#afri> ca.
- [2] "The Extensible HyperText Markup Language," 2002; <http://www.w3.org/TR/xhtml1>.
- [3] "Voice Extensible Markup Language (VoiceXML) Version 2.0," 2004; <http://www.w3.org/TR/voicexml20>.
- [4] M. Tsai, "VoiceXML dialog system of the multimodal IP-Telephony-The application for voice ordering service," *Expert Systems with Applications*, vol. 31, pp. 684-696, 2006.
- [5] J. R. Smith, R. Mohan, and C. Li, "Transcoding Internet Content for heterogeneous Client Devices," presented at IEEE International Conference on Circuits and Systems, Monterey, CA, USA, 1998.
- [6] L. Nevile, "Adaptability and accessibility: a new framework," presented at OZCHI 2005, Canberra, Australia, 2005.
- [7] F. Paternò, "Model-based tools for pervasive usability," *Interacting with Computers*, vol. 17, pp. 291-315, 2005.
- [8] Z. Shao, R. Capra, and M. A. Pérez-Quiñones, "Annotations for HTML to VoiceXML Transcoding: Producing Voice WebPages with Usability in Mind.," Computing Research Repository (CoRR), Technical Report cs.HC/0211037 2002.
- [9] H. Kim and K. Lee, "Device-independent web browsing based on CC/PP and annotation," *Journal of Network and Computer Applications*, vol. 18, pp. 283-303, 2006.
- [10] D. R. Lunn, "SADIE: Structural-Semantics for Accessibility and Device Independence," in *School of Computer Science: University of Manchester*, 2005.
- [11] C. Kouroupetroglou, M. Salamasis, and A. Manitsaris, "A semantic-Web based Framework for Developing Applications to Improve Accessibility in the WWW," presented at International cross-disciplinary workshop on Web accessibility (W4A): Building the mobile web: rediscovering accessibility?, Edinburgh, U.K., 2006.
- [12] S. H. Kurniawan, A. King, D. G. Evans, and P. L. Blenkhorn, "Personalising web page presentation for older people," *Interacting with Computers*, vol. 18, pp. 457-477, 2006.
- [13] K. Nagao, Y. Shirai, and K. Squire, "Semantic annotation and transcoding: making Web content more accessible," *IEEE MultiMedia*, vol. 8, pp. 69-81, 2001.

- [14] N. Annamalai, "An Extensible Transcoder For HTML to VoiceXML Conversion," in *Computer Science*: University of Texas at Dallas, 2002.
- [15] M. Lamb and B. Horowitz, "Guidelines for a VoiceXML Solution Using WebSphere Transcoding Publisher," vol. 2007.
- [16] M. Hori, K. Ono, Mari Abe, and T. Koyanagi, "Generating Transformational Annotation for Web Document Adaptation: Tool Support and Empirical Evaluation," *Journal of Web Semantics*, vol. 2, pp. 1-18, 2005.
- [17] E. Pontelli, T. Son, C., K. Kottapally, C. Ngo, R. Reddy, and D. Gillan, "A system for automatic structure discovery and reasoning-based navigation of the web," *Journal of Interacting with Computers*, vol. 16, pp. 451-475, 2004.
- [18] N. Yankelovich, "How do users know what to say?," *ACM Interactions*, vol. 3, pp. 32-43, 1996.
- [19] H. Takagi and C. Asakawa, "Web content transcoding for voice output," presented at 11th International Conference on World Wide Web, Hawaii, USA, 2002.
- [20] Z. Shao, R. Capra, and M. Pérez-Quñones, "Transcoding HTML to VoiceXML Using Annotation," presented at IEEE International Conference on Tools with Artificial Intelligence, Sacramento, California, USA., 2003.
- [21] C. Hsu and S.-J. Kao, "An OWL-based extensible transcoding system for mobile multi-devices," *Journal of Information Science*, vol. 31, pp. 178-195, 2005.
- [22] "XSL Transformations (XSLT) Version 1.0," 1999; <http://www.w3.org/TR/xslt>.

**Mduduzi E. Nxumalo:** is a Masters Student in the department of Electrical Engineering at the University of Cape in South Africa, supervised by Prof. Daniel Mashao.