

Annotation-Based Transcoding for Nonvisual Web Access

Chieko Asakawa

IBM Japan, Ltd.,
Tokyo Research Laboratory
1623-14, Shimotsuruma, Yamato-shi,
Kanagawa-ken 242-8502, Japan
+81 46 215 4633
chie@jp.ibm.com

Hironobu Takagi

IBM Japan, Ltd.,
Tokyo Research Laboratory
1623-14, Shimotsuruma, Yamato-shi,
Kanagawa-ken 242-8502, Japan
+81 46 215 4557
takagih@jp.ibm.com

ABSTRACT

These days, Web authors try to describe as much information as possible in one page using various types of visual effects. This information is visually fragmented into groupings. Blind users read the Web contents in tag order, but visually fragmented groupings are not accessible using tag order reading. In addition, the Web contents are designed to be visually appealing using a lot of images. This style makes nonvisual Web access harder. Therefore we decided to develop an annotation-based transcoding system to convert already-existing Web pages to be accessible, which works between a Web server and a user. It consists of two components, one for structural annotations and one for commentary annotations. Structural annotations are used to recognize visually fragmented groupings as well as to show the importance and basic role of each group. Commentary annotations are used to give users a useful description of each grouping. In this paper, we will describe our transcoding method for nonvisual Web access based on the annotations.

Keywords

Nonvisual Web access, transcoding system, structural annotation, commentary annotation, blind

INTRODUCTION

Currently existing Web pages are mainly designed for desktop computers by taking into account the screen appearance. Web authors often try to present as much information as possible on the screen. This information is visually fragmented and organized into groups, so that sighted users who use desktop computers can easily recognize the page layout and the arrangement of the

contents at a glance. Each group is set off by using visual effects such as the different background colors, layout tables, spacing and so on. Each visually fragmented grouping has a role such as header or footer of a page, index, advertisement, main content and so on.

These days, Web contents are used not only for desktop computers, but also for small-screen devices and as voice output. For these purposes already-existing Web contents should be rendered to be suitable for each device, but mix of important and less important information and visual effects make this task much harder. Although it is crucial for such a rendering system to analyze each visually fragmented group and its role, it is almost impossible to automatically detect them only by parsing HTML documents.

To solve this kind of problem for small-screen devices, transcoding systems [6] have been developed, however, we found that the requirements are different for small-screen devices and voice output devices. The main issue for small-screen devices is to split a page using each visually fragmented group and to display each grouping on a small-screen taking into account the limited capabilities of small devices.

Our objective is to create an accessible page for the blind who can only access the Web through voice output. Blind users usually access the Web using screen readers [7][11][18] or talking Web browsers [5][9]. These applications allow them to read the contents in the linear order of the HTML tags which are written in an HTML document. It is impossible for the blind users to distinguish visually fragmented groupings only from the sequence of tags read to them. Also, it is difficult to recognize the role of each group, since images are often used to show the role. So these days, Web content is becoming increasingly inaccessible to blind users.

We therefore propose using annotation-based transcoding to improve nonvisual Web access for voice output. Our proposed annotations consist of two types, structural annotations and commentary annotations. Structural annotations are used to recognize visually fragmented

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ASSETS'00, November 13-15, 2000, Arlington, Virginia.
Copyright 2000 ACM 1-58113-314-8/00/0011...\$5.00.



Figure 1: A sample of a complicated page for the nonvisual Web access (www.cnn.com)

groupings as well as to show the importance and basic role of each group. Commentary annotations are used to give users a useful description of each group. In this paper, we will first discuss the necessity of annotations for transcoding and report related work. We will then describe the annotation-based transcoding for nonvisual Web access as well as provide more information about the two components of our annotation system.

After describing an authoring tool for these annotations, we will give some examples of transcoded pages. Finally, we will offer some conclusions and plans for the future.

NECESSITY OF ANNOTATIONS

Our research goal is to offer the Web as a new information resource for the blind. As the first step in this direction, we have developed a talking Web browser called Home Page Reader (HPR) [1][2]. HPR allows blind users to navigate through Web pages using a numeric keypad. The user interface of HPR is very simple, so even computer novices

can easily start accessing the Web after two or three hours' training. It can read the Web content directly following tag order as well as other screen readers, so it is very important for nonvisual Web access that the contents should be laid out according to the tags in a logical and meaningful way. Also, the contents are requested to be compliant with the HTML 4.0 standard and the guidelines [16] of Web Accessibility Initiative (WAI) [13] of the World Wide Web Consortium. These days, however, Web authors rarely follow the HTML 4.0 standard or the WAI guidelines, since they want to control the detailed visual appearance of their Web content.

We have been studying methods for transcoding already existing Web pages to be suitable for voice output. We have been applying two approaches, one for automatic structural transcoding and one for annotation-based transcoding. Both methods have pros and cons. Automatic transcoding does not require any manually produced annotations, while annotation-based transcoding needs them. Annotation-based transcoding can create an accessible page for voice output similar to the visual appearance with annotations, while automatic structural transcoding cannot deal with visual effects nor add further description of the original content. In this section, we will focus on two critical issues for nonvisual Web access which automatic transcoding cannot handle, and then we will explain the necessity of annotations to resolve the issues.

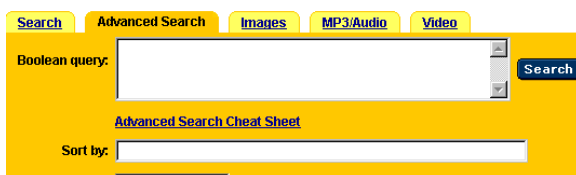


Figure 2: An example of a complicated form description (www.altavista.com)

Use of Visually Fragmented Groupings

Web authors try to present as much information as possible in one page. For this purpose, they fragment a page into groupings, using various types of visual effects, such as background colors, layout tables, spacing, horizontal lines and so on. Each visually fragmented grouping has a role as header or footer of a page, index, advertisement, main content, shopping list and so on. These groupings and their roles are easily recognized at the first glance, since Web authors take account of the visual layout on the screen. This authoring method makes it hard for the blind to navigate through Web pages.

Figure 1 is a part of a top page of "cnn.com" which uses a lot of layout tables. There are several issues in this page for nonvisual Web access. We will report two of them. One is that the main content (C) in the Figure 1 appears under the date (A) on the screen. It is, however, described by using a layout table in the HTML document, so the tag order is (A), (B) and (C). When blind users read it, they hear the link list at (B) after the date (A). The main contents (C) are read after this. It means that blind users cannot recognize the grouping which consists of (A) and (C). If the date and the main contents are written in one cell of the table (within one <td> (Table Data) tag), they can be read and recognized as one group.

The other issue is that there is no delimiter between each group allowing them to be distinguished nonvisually. The logo at location (D), for example, is read right after the video description box at location (E) without any notification. Blind users cannot recognize if they are related to each other or not. They need to listen to the entire contents very carefully to understand the page layout and its contents. If the changes between groups were indicated by voice or sound, the nonvisual Web access would be much easier. To resolve this issue, structural annotations would be useful. Sighted annotators can specify each visually

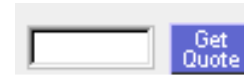


Figure 3: An example of a non-descriptive form

fragmented grouping very easily. When the system can recognize each grouping, it is possible to reorder each group as well as add notifications of the boundaries of each group.

The Difficulty of Understanding the Role of Each Group

It is also very hard for the blind to recognize the role of each visually fragmented grouping. Web authors often use images to indicate a role for the group, since there is not much space on the screen. Group (F) in Figure 1 is one example. In this group, "click here" is the only text, so it is impossible for the blind to understand the role of the group. When an annotation which indicates the role of the group as "advertisement" is provided, they can understand. That this group is less important. Figure 2 is another example. The type of the form in the Figure 2 is changed by clicking one of the links above the form. It is hard for the blind users to notice the role of the link list without any notification of a grouping as well as a description of the form. When an annotation which describes the role of this form as "the type of the form is changed by clicking one of the links above the form" is provided, they can understand that the form is used for several purposes. Annotations for form elements would be very useful. In the case of Figure 3, for example, the role of this form is presented by an image which says "Get Quote". So sighted users can easily recognize that this form is for searching a quote, however, blind users only get the information that there are an input box and a submit button in the form, so there is no way for them to recognize the role of this form.

In this way, annotations can be used effectively to transcode already-existing Web pages to be suitable for voice output for the blind, and to give users useful descriptions for Web

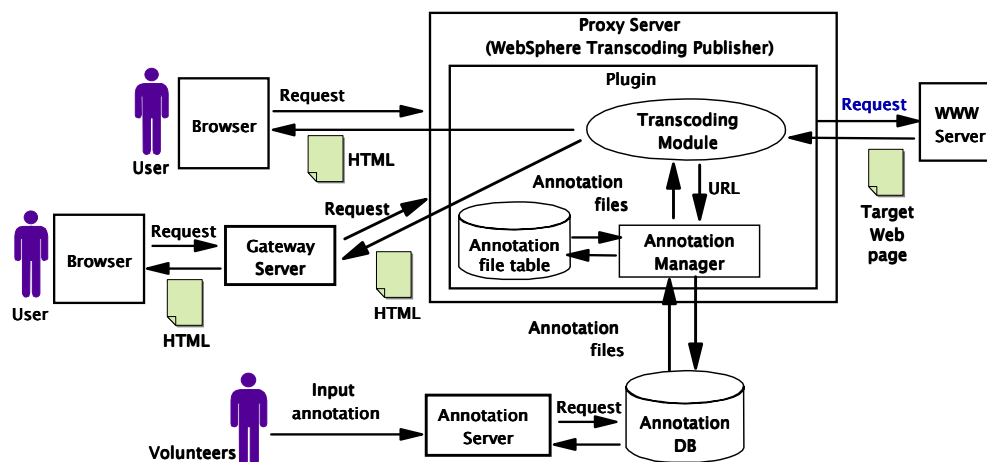


Figure 4: The system architecture

page navigation. Without transcoding Web pages or letting Web authors know the importance of the Web accessibility, the Web will be increasingly inaccessible to blind users in the near future.

RELATED WORK

In this section, we will summarize related work and compare it to our transcoding system.

(1) Annotation-based Web Content Transcoding [6]

This system is designed to transcode HTML documents to be suitable for small-screen devices. It uses external annotations to fragment a page for various kinds of small-screen devices. Therefore, it needs to take account of physical and performance constraints of the devices, such as screen size, memory size and connection bandwidth. Both systems recognize visually fragmented groupings by the external annotations. However, they differ in that their main objective is to fragment a page for small-screen devices and our objective is to reorder visually fragmented groupings according to the importance of each grouping and give useful descriptions for voice output. We referred to their vocabularies for structural annotations, but our commentary annotations are defined by our unique requirements for nonvisual Web access.

(2) Web Access Gateway [14]

This site is provided for visually impaired users. It transcodes already existing Web pages automatically on the fly. They try to remove images and create a text only page, serialize all text information by removing tags, change font size and color settings, enlarge images and so on. With

regard to serializing the text, HPR and JAWS have already implemented it on the client-side. This site does it on the server-side. We think that many functions are focused on users with poor vision, and it is quite different from our transcoding system which is mainly focused on nonvisual Web access based on annotations.

(3) Commentary Annotations

The study of commentary annotations has been becoming popular, with examples such as ComMenter [10], CoNote, Group Annotation Transducer [11] and Semantic Transcoding [8]. The primary objective of these systems is to share comments on Web content. Our objective is to describe a set of elements of HTML tags for voice output, so the use of our commentary annotations is different from these systems.

(4) Others

ALT-server [3] was introduced in the Evaluation and Repair Interest Group [15] of WAI. This proposal is similar to our system only regarding descriptive alternative text for images. The use is limited to this purpose while our system has capabilities for structural and commentary annotations. Commentary annotations are used not only for an image, but also for any element using HTML tags. They propose to extend a browser on the client-side to use a corresponding external file. Our proposal does not require any modification on the client-side, and the whole system works on the server-side.

Table 1: Vocabularies of our annotations

Tag	Attribute	Description
<annotation> ...</annotation>		start and end of an annotation file
	url=' '	the address of corresponding HTML documents. It can be written as a regular expression
	author=' '	e-mail address of an annotator
<alternatives> ...</alternatives>		description of substitute text of an image
	path=' '	description of xpath of the image
<group> ...</group>		a set of elements
	title=' '	a title for this group
<importance/>	members=' '	description of xpath of each element in the set of the group
		description of the importance of a group or an element
	path=' '	description of xpath of the element
	value=' '	a real number showing the importance from -1 to 1 (default = 0)
<role/>		a role of a group or an element
	path=' '	description of xpath of the element
		selection of items for a role
		"proper content" = main content of a page
		"updated index" = frequently updated indexes
		"general index" = general indexes (not updated often)
	type=' '	"norole" = default value "header" = header of a page "footer" = footer of a page "advertisement" = banner or advertisement "delete" = group can be deleted "layout table" = a table tag is used for layout purposes (only for a table elements)
<description> ...</description>		free description of a group or an element. Examples of elements: <form>, <input>, <table>, , <area> of image map
	path=' '	description of xpath of the element

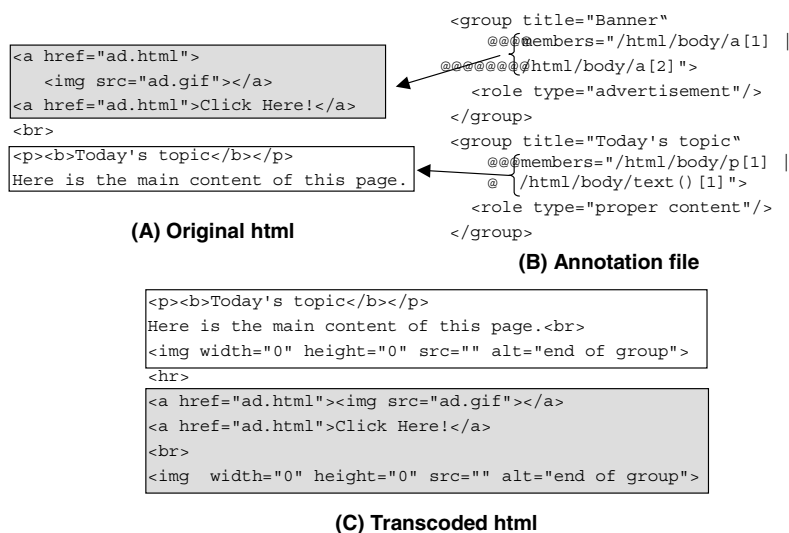


Figure 5: An example of structural annotation-based transcoding

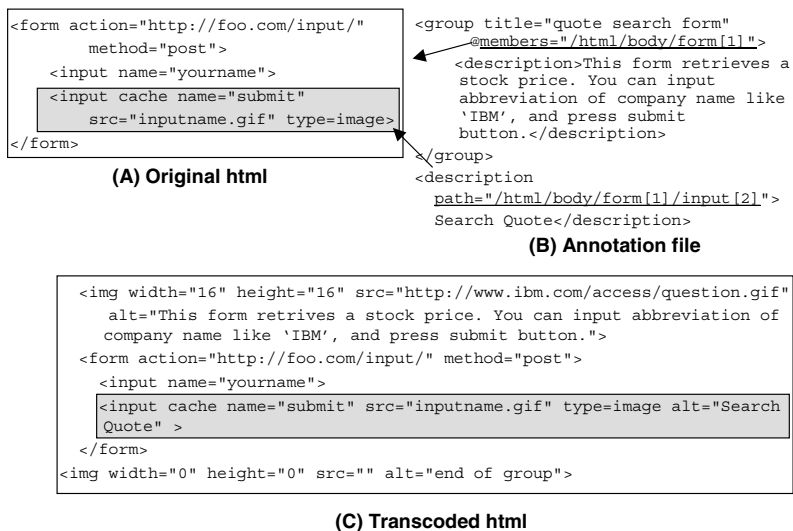


Figure 6: An example of commentary annotation-based transcoding

OVERVIEW OF OUR ANNOTATION-BASED TRANSCODING

In this section, we will first describe the architecture of our system. Then annotations which consist of two components, such as one for structural annotations and one for commentary annotations will be described as well as our transcoding method. Finally, we will discuss the required characteristics of an authoring tool for our annotations.

Architecture

Figure 4 shows the architecture of our system. The system consists of three components—a proxy server, an annotation database and an annotation server. The proxy server is the main component that transcodes a target HTML document. Users access our system by setting a proxy server on their browser, or they can use it through a gateway server which does not require any setting on the client-side. However, the gateway server sometimes causes problems in the redirection process. We use the IBM WebSphere Transcoding Publisher (WTP) [17] as a proxy server and our transcoding system is implemented as a plug-in to WTP. It consists of a transcoding module, an annotation manager and an annotation file table. When the transcoding module receives a target HTML document, it sends the URL of the target HTML document to the annotation manager. The annotation manager retrieves an id list of proper annotation files and sends it to the annotation database. Finally, the transcoding module will receive the annotation files which are sent by the annotation database according to the id list, and transcodes the target HTML document using the annotation files.

The annotation server receives annotation files which are created by volunteers. It registers them to

the annotation database. Each annotation file is basically specified for one URL. However, one page is sometimes similar to others, and in such cases, the annotation files can be shared among those pages. This allows the annotation server to accept wild card notations for URLs from annotators.

When an annotation file is registered to the annotation database, the annotation server checks if the annotation file can be used with any other pages in the site by calculating the similarity of the DOM trees[5]. To be more specific, it

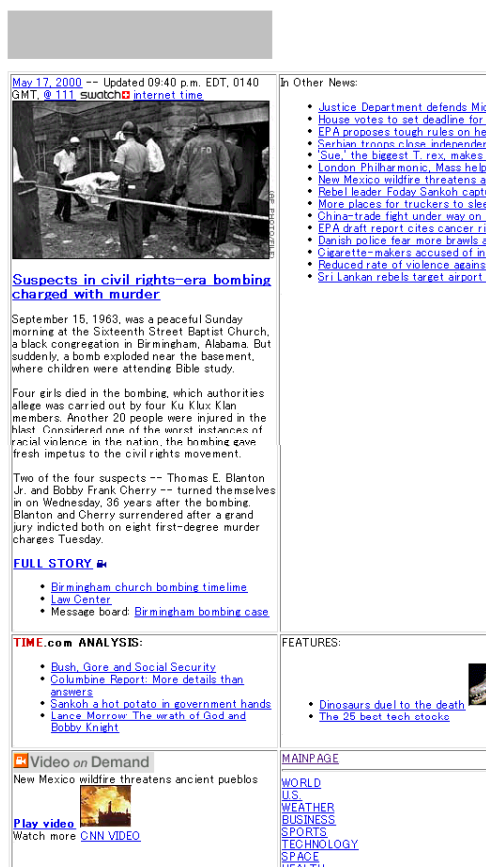


Figure 7: A snapshot of a transcoded page (www.cnn.com)

compares the specified branch of the annotated DOM tree with each branch of DOM trees of other pages in the site. When it detects the similar branch, it will ask users if the annotation can be applied as the annotation of the branch. In this way, the system supports annotators that they will not describe complicated regular expressions of URLs and duplicated annotations.

Structural Annotation

As mentioned above, Web contents are often visually fragmented into groupings to describe as much information as possible in one page using various types of visual effects. It is hard for the system to recognize each visually fragmented grouping automatically. Table 1 shows the vocabulary used in our annotations. The transcoding module uses structural annotation to recognize visually fragmented groupings. In the structural annotations, a `<group>` tag indicates each group. The `members` attribute in a `<group>` tag specifies a set of elements of a group. The `members` attribute is written using XPath[20]. Each group has a role, so the `role` attribute is used for the basic role of a group. An `importance` attribute is used to show the

importance of a group. The transcoding module reorders each grouping by referring to its role and importance. When a group of today's articles on a news site is annotated as "proper content" in the `role` attribute, for example, it would be placed at the top of a page. When a group of banners is annotated as "advertisement" in the `role` attribute, it would be placed at the bottom of a page.

Figure 5 shows a simplified example of an original HTML document, its annotation file and the transcoded HTML document. The transcoding module inserts a delimiter for each group using alternative text for the image which does not appear on the screen.

As mentioned above, Web authors use a lot of visual effects to show a role for each grouping, and this makes nonvisual Web access harder. Commentary annotations are used to give users a useful description of each group. Also, they can be used to describe HTML tags, such as `<table>` tag, `<form>` tag, `` (image) tag, `<area>` tag and so on. Figure 6 shows a simplified example of an original HTML document, its annotation file and the transcoded HTML document. The transcoding module inserts the comment in the `<description>` tag into the original document as the alternative text for an image.

Another useful example of commentary annotations is to provide a comment for each `<area>` tag in an image map. Image maps are used in many Web pages, but there is usually no alternative text in each `<area>` tag, so it is very useful to add such annotations. In addition, since such image maps do not change often, the annotation files can often be used for a long while.

An Authoring Tool for Annotations

It is extremely time-consuming to describe annotations manually using a simple editor. It is crucial to have a WYSIWYG authoring tool for annotations which is easily operated by any computer user.

Our authoring tool is based on a browser's DOM interface and an XML parser[19]. The XPath is determined using the browser's DOM structure. When annotators specify a visually fragmented grouping on the screen, they can select all the objects in a group using a mouse. Then the authoring tool internally finds an XPath notation to correspond the selected text information. It tries to find the smallest higher-level node that includes all of the objects during this process. For example, assume that there are two paragraphs in one cell. When a user selects only the first paragraph, the XPath should be ".../td[1]/p[1]". When a user selects both paragraphs, the XPath should be ".../td[1]" rather than ".../td[1]/p[1] | .../td[1]/p[2]". In this way, when the contents in the cell are changed, the annotation for the cell may remain useful. Finally it defines the XPath in the `members` attribute in the `<group>` tag of an annotation file. When they annotate a role for a group, a dialog with a selection

```

<annotation url="http://www.cnn.com/" time="*"
  author="takagi@jp.ibm.com">
  <group title="Today's top article"
    members="/html/body/table[1]/tbody[1]/tr[4]/td[1]/table[1]/tbody[1]/tr[2]/td[3] |
/html/body/table[1]/tbody[1]/tr[4]/td[1]/table[1]/tbody[1]/tr[3]/td[3]/table[1]/tbody[1]/tr[1]/td[1]">
    <role type="proper content"/>
    <importance value="1"/>
  </group>
  <group title="Category index"
    path="/html/body/table[1]/tbody[1]/tr[4]/td[1]/table[1]/tbody[1]/tr[2]/td[1] |
/html/body/table[1]/tbody[1]/tr[4]/td[1]/table[1]/tbody[1]/tr[3]/td[3]/table[1]/tbody[1]/tr[1]/td[1]">
    <role type="general index"/>
  </group>

```

Figure 8: A part of the annotation file of www.cnn.com

```

<annotation url="http://www.altavista.com/cgi-bin/query?pg=aq&what=web/" time="*"
  author="takagi@jp.ibm.com">
  <group title="advanced search form"
    members="/html/body/table[3] | /html/body/table[4] |
/html/body/table[5]">
    <description>
    A form for searching detailed conditions. Input
    conditions into the first input box, named "boolean
    query". You can use "AND" and "OR" notations.
    </description>
  </group>

```

Figure 9: The annotation of the form in www.altavista.com

menu of roles will appear. They can also specify the importance of a group at the same time. When they annotate an element with a description, they first select the target element and then an input box will appear.

EXAMPLES OF TRANSCODED PAGES

In this section, we describe our transcoding policy based on the annotations. Figure 1 is a snapshot of the original CNN top page and Figure 7 is a snapshot of a transcoded CNN top page. Figure 8 is a part of the annotation file of "www.cnn.com". There are 13 visually fragmented groupings in the snapshot. Each grouping is annotated with its role and importance. The roles should be defined by annotators, but the importance is sometimes automatically defined by its role. When the role is "advertisement", for example, its importance is defined as "-0.8". It means that this group is placed at the bottom of a page. When an annotator defined some group as "-1", it will be placed at the extreme bottom of the page. The "proper content" is defined as "0.8", so it is placed at the top of the page. When there is no annotation for the role and importance, the system automatically assigns the default settings, as "no role" for the role and "0" for the importance. The system inserts an image map at the top of a page to link directly to each group in the page. In the case of Figure 7, there are 13 links in the image map. And each <area> tag has

alternative text which indicates a role of each group. The system also inserts an "end of group" comment at the end of each group as an alternative text for an image to present a voice delimiter between each group. This image is sized as "0" in an tag, so sighted users will not see it.

We described the necessity of commentary annotations using Figure 3. Figure 9 is an annotation file to solve this problem. Figure 10 is an example of the voice output by HPR, and Figure 11 is a snapshot of the transcoded page. Visually, the image for the alternative text is presented as a question mark to allow sighted users to check the comment. In this way, blind users will effectively understand Web content by accessing transcoded pages produced by our annotation-based transcoding system.

CONCLUSION

In this paper, we first discussed the necessity of annotation-based transcoding and summarized some related work. Then we described the overview of our transcoding system. Our system transcodes already-existing Web pages to be suitable for voice output using external annotations. Our proposed annotations consist of two components, one for structural annotations and one for commentary annotations.

Structural annotations are used for reordering each visually fragmented grouping according to its importance. Commentary annotations are used for giving useful description of the contents. We then discussed a WYSIWYG authoring tool for our annotations which can be easily used by any computer user. Finally, we showed examples of transcoded pages produced by our transcoding method based on the annotations.

Now we have to think of ways to put our system into practice to improve the Web accessibility for the blind. It would be the best if we could ask original content authors to annotate their pages by themselves. However, it will take time to let them know the necessity of the annotations. It is becoming well known that the inclusion of alternative text for images and image maps is important, but in many pages we still find images and image maps without any alternative text. So we hope to ask volunteers to help describe the annotations for our system. The notation of our annotations is designed to be as simple as possible to encourage many volunteers' participation. Also, we try to use one annotation file for as many pages as possible by checking the similarity of neighboring pages. We can also extend the lifetime of the annotations by defining appropriate XPath in the

members attribute so they are not affected by every update of a few sentences.

Next, we hope to create transcoded pages which look almost the same as the original pages using Cascading Style Sheets. This approach will also improve copyright protection by not creating completely new pages during the transcoding process. Then, we plan to support the capability of users to use profiles to apply their personal requirements flexibly. One example would be allowing the user to define whether or not descriptions are needed. When users are computer novices, they would need to have as many descriptions as possible by taking full advantage of all annotations. However, when they become more advanced, they would probably prefer to have less verbose descriptions.

For annotators, we have not yet provided annotation-time transcoding on the client-side, so we would like to enhance our system to allow this functionality. It is also important for the system to recognize when a page layout which affects an existing annotation file is changed. If it is possible, we would like to automatically warn annotators to modify annotation files. In this way, annotation files can be used longer with fewer modifications.

We hope that our annotation-based transcoding will improve nonvisual Web access for the blind, as well as spreading awareness of the importance of Web accessibility.

REFERENCES

1. Asakawa C. User Interface of a Home Page Reader, in *Proceedings of The Third International ACM Conference on Assistive Technologies ASSETS '98* (April 1998), 149-156.
2. Asakawa C. User Interface of a Non-visual Web Access System, in *Transactions on Information Processing Society of Japan* Vol. 40 No.2 (Feb. 1999), 453-459.
3. Dardailler D. The ALT-server ("An eye for an alt"), <http://www.w3.org/WAI/altserv.htm>
4. Home Page Reader: IBM Accessibility Center, <http://www-3.ibm.com/able/>
5. Document Object Model (DOM), World Wide Web Consortium, <http://www.w3.org/DOM/>
6. Hori M, Kondoh G, Ono S, Hirose S, and Singhal S. Annotation-Based Web Content Transcoding, in *Proceedings of The Ninth International World Wide Web Conference (WWW9)* (May 2000).
7. JAWS, Henter-Joyce, Inc., <http://www.hj.com/>
8. Nagao K. Semantic Transcoding: Making the World Wide Web More Understandable and Usable with External Annotations, in *Proceedings of International Conference on Advanced in Infrastructure for Electronic Business, Science, and Education on the Internet* (2000).
9. pwWebSpeak: The Productivity Works, http://www.prodworks.com/catalog/catalog_pwwebspeak.html

```
[A form for searching detailed conditions.
  Input conditions into the first input
  box, named "boolean query". You can use
  "AND" and "OR" notations.]
Search | Advanced Search | Images |
MP3/Audio | Video

(Start of form 1)
Boolean query: [Textarea:   ]
[Search: Submit Button]
```

Figure 10: An example of the voice output by HPR for Figure 11

Figure 11: A snapshot of the transcoded form in Figure 2

10. Roscheisen M., Mogensen C., and Winograd T. Shared Web annotations as a platform for third-party value-added information providers: Architecture, protocols, and usage examples, in *Technical Report CSDTR/DLTR. Computer Science Department, Stanford University* (1995).
11. Schickler A. M., Mazer S. M., and Brooks C. Pan-browser support for annotations and other meta-information on the World Wide Web, in *Proceedings of Computer Networks and ISDN Systems* Vol. 28 (1996).
12. Thatcher, J. Screen Reader/2: Access to OS/2 and the graphical user interface, in *Proceedings of The First Annual ACM Conference on Assistive Technologies ASSETS '94* (Nov. 1994), 39-47.
13. Web Access Initiative (WAI): World Wide Web Consortium (W3C), <http://www.w3.org/WAI/>
14. Web Access Gateway: the Association of C and C++ Users, <http://www.accu.org/cgi-bin/access/access>
15. Web Access Initiative (WAI) Evaluation and Repair Tools Interest/Working Group: World Wide Web Consortium (W3C), <http://www.w3.org/WAI/ER/IG/>
16. Web Content Accessibility Guidelines 1.0: World Wide Web Consortium (W3C), <http://www.w3.org/TR/WAI-WEBCONTENT/>
17. WebSphere Transcoding Publisher, IBM Corporation, <http://www-4.ibm.com/software/webserver/transcoding/>
18. Window-Eyes: GW Micro, Inc., <http://www.gwmicro.com/>
19. Xerces: Apache XML Project, The Apache Software Foundation, <http://xml.apache.org/>
20. XML Path Language (XPath): World Wide Web Consortium (W3C), <http://www.w3.org/TR/xpath>