# SEMANTIC TRANSCODING:
## Making the handicapped and the aged free from their barriers in obtaining information on the Web

*Shinichi Torihara * **, Katashi Nagao ***

* IBM Research, Tokyo Research Laboratory*
** Keio University, Media and Governance*

## ABSTRACT

Transcoding refers to the process of converting HTML files in order to adapt the physical attributes of each device. We call our transcoding "Semantic Transcoding" because it transcodes at the deep semantic level. The proxy server consists of translation, text summarization, video summarization, text-to-speech, speech recognition, and so on. External annotation is also possible. Semantic transcoding enables the visually challenged to browse Web pages by changing text to speech, and to understand the pictures by reading external annotated comments. The hearing impaired are provided with superimposed dialog generated by the recognition of speech and scene. Furthermore, we will propose a rapid "diagonal" reading method by listening to speech converted from text on the basis of linguistic information such as syntax, new and old information for the blind and visually impaired in the internet speech browser realized by semantic transcoding.

## 1.INTRODUCTION

We will demonstrate the potentials of freeing handicapped people from barriers imposed by age or other limitations to obtain information on the Web by "semantic transcoding"[1].
We propose "the super-structure on the Web" based on adding external annotations [2] to Web documents. This super-structure would enable any user to annotate additional information to any element of any document. We classify annotations into three categories. The first, linguistic annotation, would help the transcoder to understand the semantic structure of documents. The second, commentary annotation, would assist
the transcoder to manipulate non-textual elements such as images and sounds. The third, multimedia annotation, is a combination of the two types. In this paper, we will describe semantic transcoding and exemplify the applications of semantic transcoding in welfare information technology (WIT) for the blind and visually impaired, the hearing impaired and aging people (the weakly sighted).

## 2.SEMANTIC TRANSCODING
## 2.1. Super-Structure on the Web

The conventional Web structure can be considered as a graph on a plain. HTML (HyperText Markup Language) is just used to describe the relationship and structures in networks. HTML authors cannot express their ideas for making anyone understand

only by HTML.

In this paper, we propose a method for turning a plain graph into a three-dimensional structure consisting of multiple layers on the Web. This metalevel structure is based on external annotations on documents of the Web. The following figure represents the concept of our approach.
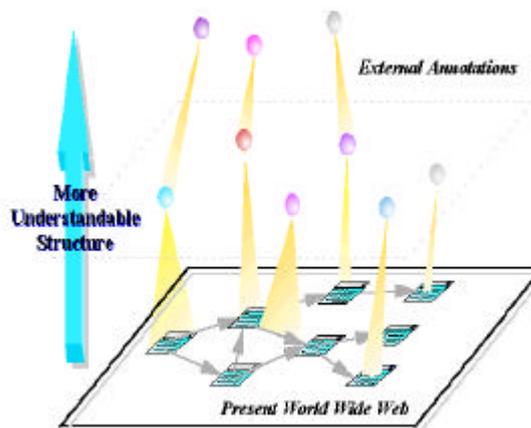


**Figure.1**: Super-structure on the Web

A super-structure on the Web consists of the layers of contents and metacontents. The first layer corresponds to set of metacontents of documents. The second layer corresponds to set of other metacontents of these metacontents. We consider such metacontents as an external annotation in general.

A popular example of external annotation is comments or notes on Web documents "added" by people other than the author. Images without alternative descriptions are not understandable for visually challenged people. If there are comments on these images, they can accept image contents by converting the text of their comments to speech.

## 2.2. Content Adaptation

Content adaptation is a kind of transcoding that considers a user's environment such as devices, network bandwidth and so forth. Such adaptation sometimes involves deep understanding of the original contents. If a transcoder fails to analyze the semantic structure of documents, then the results may cause user

misunderstanding. We assume that external annotations help machines to understand the contents so that transcoding would have higher quality. We call our transcoding based on external annotation "semantic transcoding" because it can transcode at the deep semantic level. The overall configuration of semantic transcoding is shown below in Figure 2.
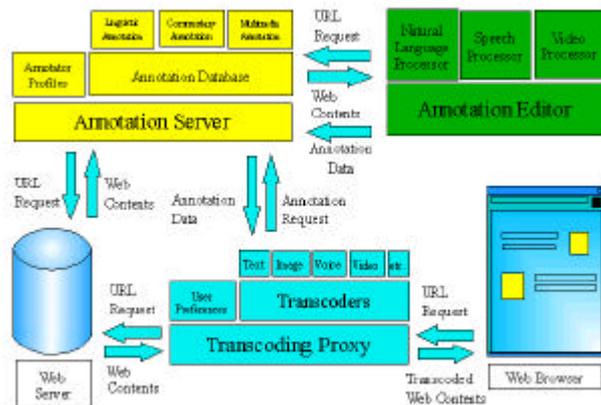


**Figure.2:** Configuration of Semantic Transcoding

There are three main parts in this system: annotation editor, annotation server and transcoding proxy. The remaining parts are a conventional Web server and browser.

## 2.3 External Annotation

Semantic transcoding can associate external annotations with any element of any HTML document. Our annotations are represented as a XML formatted data and divided into three categories (linguistic, commentary and multimedia annotation). Multimedia annotation is a combination of the other two types of annotation.

### 2.3.1. Creation and Management of Annotations

Our annotation environment consists of a client side editor for the creation of annotations and a server for the management of annotations.
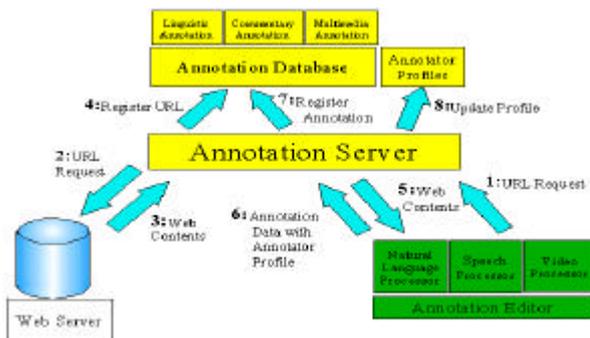


**Figure.3:** Annotation environment

The following are the flows in processing HTML files

1. The user runs the annotation editor and requests a URL as a target of annotation.
2. The annotation server accepts the request and sends to theWeb server.
3. The annotation server receives the Web document.
4. The annotation server calculates the document hash code(digest value) and registers the URL
5. The annotation server returns the Web document.
6. The user annotates the requested document and sends the result to the annotation server with personal data.
7. The annotation server receives the annotation data and relates it with its URL in the database.
8. The annotation server updates the annotation profiles.

### 2.3.2 Linguistic Annotation

Linguistic annotation is to make Web text contents machine-understandable and to perform translation, text summarization, text-to-speech, etc. with much higher quality than currently available text contents. The annotation is generated by automatic morphological analysis, interactive sentence parsing, and word sense disambiguation by selecting the most appropriate paraphrase. We have been developing machine-guided annotation interfaces that conceal the complexity of annotation itself.

### 1.3.3 . Commentary Annotation

Commentary annotation is primarily used to annotate non-textual elements like images and sounds as additional textual information. This type of annotation is also used for text transcoding that combines such comments with the original texts.

### 2.3.4. Multimedia Annotation

Our annotation technique is also applied to multimedia data such as digital video. We developed the techniques for semi-automatic video annotation using a text describing the content of the video. This type of annotation is used to enhance automatic cut detection, characterization of frames in a cut, and scene recognition according to the similarity between cuts.

### 2.4. Transcoding Environment

Semantic transcoding is a combination of several transcoders based on external annotations, which has the functionality of content adaptation according to user preferences. Transcoders are implemented as an extension of an HTTP proxy. Such HTTP proxy is called a transcoding proxy. We will show transcoding environment in Figure 4.

**Figure.4:** Transcoding Environment

IBM's WBI (Web Intermediaries) is used as a development platform to implement our semantic transcoding system [3]. WBI is a customizable and extendable HTTP proxy. WBI provides APIs for the user access level control and the manipulation of input/output data of the proxy. The transcoding proxy based on WBI has the following functions.

1. Maintenance of personal preferences.
2. Getting and management of annotation data.
3. Activation and integration of transcoders.

### 2.4.1. Text Transcoding

Text transcoding is the transformation of text contents based on linguistic annotation. Text summarization is a good example to demonstrate our techniques.

### 2.4.2. Image Transcoding

Image transcoding is to convert images into these of different size, color and resolution depending on user's device and capability. By combing image and text transcoding, the system can convert contents to fit to the client screen size.

### 2.4.3. Voice Transcoding

Voice transcoding generates a spoken language version of documents. There are two types of voice transcoding. One is related to text-to-speech. The other is speech recognition.

### 2.4.4. Video Transcoding

Video transcoding employs video annotation that consists of linguistically-markuped transcripts like closed captions, time stamps of scene changes, representative images (key frame) of each scene, and additional information such as program names, etc.

## 3. SEMANTIC TRANSCODING AND WELFARE INFORMATION TECHNOLOGY (WIT)

Multimedia, multi-modal user interfaces and so-called "universal design" enable all the people in the world including the handicapped people to use computers easily. Semantic transcoding that consists of a combination of several transcoders and external annotations will contribute to welfare information technology (WIT) because our transcoding technologies are empowered by an integration of network-to-network (Internet ) and human-to-human (human network).
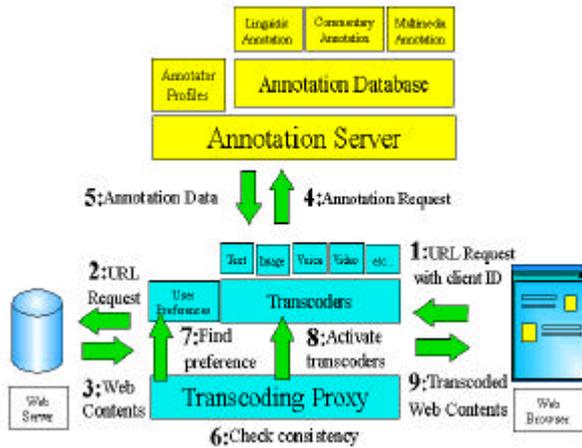
### 3.1. For the Visually Challenged

One of the applications for the blind and visually impaired by semantic transcoding is the Internet speech browser.

The transcoding proxy converts original text elements and annotated text elements to Web documents into voice (MP3 files) by text-to-speech and makes a SMIL (Synchronized Multimedia Integration Language) for a playlist. A client PC can choose, play and understand all of the elements on the Web in the way of stream-playing, not downloading. Configuration of Internet speech browser is shown as below.
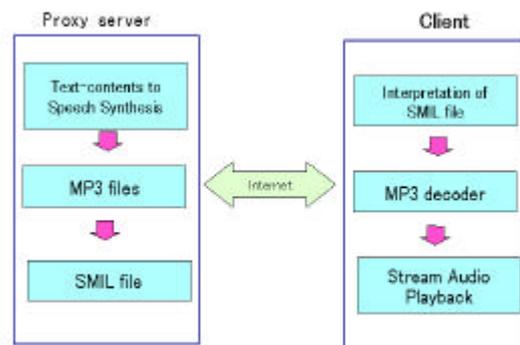


**Figure.5:** Configuration of Internet Speech Browser for the Blind and Visually Impaired

### 3.2. For the Hearing Impaired

Semantic transcoding by speech recognition and scene recognition with external annotations enables the hearing impaired to understand video with sounds (voice) like movies. In addition, the translation of the superimposed dialog into a mother tongue and text-to-speech will be useful to the blind and visually impaired.
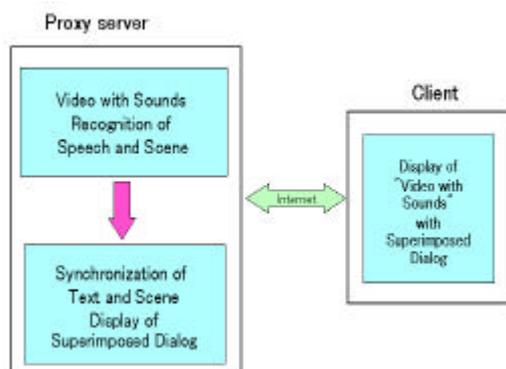
**Figure.6:** Configuration of superimposed dialog for the hearing impaired

## 3.3. For the Weakly Sighted, Including Aging People

There is a tendency that people's sight-power will become weaker as they become older. Each PC user has his/her own parameters in seeing a screen and reading characters, etc. The transcoding proxy helps each people to understand the contents of Web documents depending on his/her preferences in seeing.
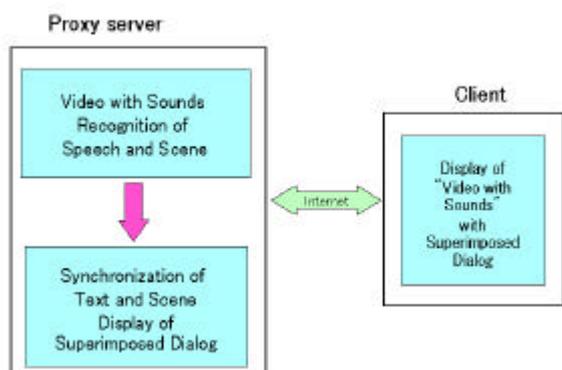


**Figure.7:** Configuration of Browser Customization for the Weakly Sighted

## 4.AN "OBLIQUE" LISTENING METHOD FOR THE VISUALLY IMPAIRED

Normally-sighted people can increase their scanning and reading ability by a method known as "diagonal" reading. The blind and visually impaired, however, must rely on a conversation of written text to speech. The problem with this is that speech sounds generated from text are sequential and linear. If you listen to speech sounds at a faster than normal speed, you will not be able to understand them well. In this study, I propose a rapid "oblique" listening method for the blind and visually impaired by controlling speed based upon the linguistic information (syntax, new and old

information) of any given text[4]. Verbs, nouns and negative adverbs, and new information will be played at relatively slower speed, the remaining parts at a much faster speed. The need for the visually impaired for improved access to widespread interdisciplinary knowledge is obvious. This method will help them share the advantages of "diagonal" reading currently available only to the fully sighted.

## 5. CONCLUSION

We have discussed a total architecture for creating and utilizing external annotations. By using transcoders and annotations, we realized semantic transcoding that automatically customizes Web contents depending on user preferences. Our technologies naturally contribute to those who have handicaps and difficulties in obtaining information on the Web, and to welfare information technology because semantic transcoding is realized by an integration of network-to-network and human-to-human collaboration.

## 6. REFERENCES

1. Nagao, Semantic Transcoding: Making the World Wide Web More Understandable and Usable with External Annotations,SSGRR2000, August 6, 2000.
2. Hori, Annotation of Web Content for Transcoding World Wide Web Consortium, July 10, 1999.
3, Web Intermediaries (http://www.almaden.ibm.com/cs/wbi/index.html)
4, Torihara, An "Oblique" Listening Method for the Visually Impaired (http://www.torihara.com/research/index.html)