HearSay: Enabling Audio Browsing on Hypertext Content

I. V. Ramakrishnan
Department of Computer Science
Stony Brook University
Stony Brook, NY 11794-4400
ram@cs.stonybrook.edu

Amanda Stent
Department of Computer Science
Stony Brook University
Stony Brook, NY 11794-4400
stent@cs.stonybrook.edu

Guizhen Yang
Department of Computer Science and Engineering
University at Buffalo
Buffalo, NY 14260-2000
gzyang@cse.buffalo.edu

ABSTRACT

In this paper we present HearSay, a system for browsing hypertext Web documents via audio. The HearSay system is based on our novel approach to automatically creating audio browsable content from hypertext Web documents. It combines two key technologies: (1) automatic partitioning of Web documents through tightly coupled structural and semantic analysis, which transforms raw HTML documents into semantic structures so as to facilitate audio browsing; and (2) VoiceXML, an already standardized technology which we adopt to represent voice dialogs automatically created from the XML output of partitioning. This paper describes the software components of HearSay and presents an initial system evaluation.

Categories and Subject Descriptors

D.2.11 [**Software Engineering**]: Software Architectures; H.4.3 [**Information Systems Applications**]: Communications Applications—*information browsers*

General Terms

Algorithms, Design, Experimentation

Keywords

World Wide Web, HTML, VoiceXML, structural analysis, semantic analysis, audio browser, user interface

1. INTRODUCTION

As the World Wide Web has matured, an enormous amount of information, ranging from news to entertainment to scientific discoveries, has been made readily accessible to the general public. Recently Google announced that its search engine has indexed 4.3 billion Web pages. Furthermore, it is estimated that the Web is still constantly growing at a rapid pace, doubling in size every 8 months. However, most Web documents are designed primarily for graphical display and interaction. This means that individuals with visual disabilities and/or highly mobile individuals have reduced access to the many resources the Web provides.

Creating audio browsable Web content has become the focus of intensive research efforts by industrial enterprises (*e.g.*, IBM, AT&T, Microsoft) and standardization organizations (*e.g.*, W3C). New markup languages, such as VoiceXML [4] and SALT [3], and new voice browser systems, such as IBM's WebSphere Voice Server and Lucent Technologies' MiLife VoiceXML Gateway, are emerging to facilitate the creation, publishing, and exchange of audio browsable Web content.

Copyright is held by the author/owner(s).
WWW2004, May 17–22, 2004, New York, New York, USA.
ACM 1-58113-844-X/04/0005.

However, adapting to voice browser technology still remains a significant burden for many Web content providers. In order to deliver their content to different browsers with different visual and audio modalities, Web content providers have to design and encode their Web content in different forms using various standard markup languages such as HTML, VoiceXML, and WML [5] — a tremendous development and maintenance overhead. Such diversity in content formats has become an obstacle to the wide adoption of voice browser technology by Web content providers.

Although voice browsers that can read HTML documents exist (*e.g.*, [2]), they do not adequately convey to users the logical structure and semantics of content in Web documents, nor do they provide users with easy ways to select which parts of a document to listen to. Consequently, users who are visually disabled and/or mobile waste a considerable amount of time and attention listening to irrelevant information.

Our goal is to develop intelligent systems that can enable people to interact directly, quickly, and easily with Web content using non-visual modalities, primarily audio. In this paper we present the HearSay system, a prototype audio Web browser we have developed. HearSay is based on a novel technology, *partitioning*, for structural and semantic analysis of HTML documents. HearSay is completely speech driven. It supports interactive *exploratory browsing*, presenting information about an HTML document's logical structure and content and permitting users to select which parts of the document to listen to and when to navigate to a new page. The key ideas underlying this system are:

1. Novel structural analysis techniques that leverage the structural annotations of HTML documents to partition Web content into semantic units, which are organized and then presented to users as a concept hierarchy (or *partition tree*). We have also developed efficient heuristics to identify semantic labels for nodes in the partition tree.

2. Novel semantic analysis techniques that are tightly coupled with structural analysis. We use lexical associations derived from WordNet, combined with ontologies, to more accurately separate semantic units and to assign informative labels to partitions. In addition, we have proposed effective heuristics to propagate semantic associations based on structural cues in HTML documents, further improving the performance of our techniques.

3. Template-based automatic VoiceXML dialog creation using partition trees. This permits on-the-fly generation of dialogs for efficient, user-directed interactive audio browsing. An