Multi-Layer Dialog Generation for Non-Visual Web Access

Yevgen Borodin
Department of Computer Science
Stony Brook University
Stony Brook, NY 11794, USA
borodin@cs.sunysb.edu

ABSTRACT

People with visual disabilities use screen-readers to browse the Web. The existing screen-readers have limited ways of presenting Web page content. I propose to turn non-visual Web browsing into a multi-layer mixed-initiative dialog-based interaction between users and computers. The suggested layers of dialog navigation are: basic screen-reading, DFS or BFS, and domain-specific. The support of adaptive dialogs is also planned. This research is aimed at improving and accelerating non-visual Web browsing for blind users.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*natural language, Voice I/O*; H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*navigation*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Language generation*

General Terms

Design, Standardization, Human Factors

Keywords

Web Navigation, Voice Browsing, non-Visual, Screen Reader, Multi-Layer, Domain-Specific, Mixed-Initiative, Dialog Generation, User Interface, VoiceXML, VXML.

1. INTRODUCTION

The Web has become an indispensable aspect of our lives. In our daily activities we turn more and more to it for information: from checking the weather to searching for air tickets or getting directions. We use it for reference, news, shopping, etc. The Web was designed for graphical modes of interaction, making all these activities simple for sighted users. Users with visual disabilities, however, have to use screen readers, which process Web pages sequentially making the process of Web browsing complicated and time-consuming.

A typical screen reader simply speaks out the content of a page to the user. The only way to interact with a screen reader is by means of shortcut keys, which allow skipping paragraphs, searching, pausing, changing the rate of speech, etc. Some screen readers also provide their users with extended features, such as summarization [11], lists of current and visited links [3], etc. All these features greatly improve Web-browsing. However, they provide little control over the way the information is presented to the user. Also, the interaction between the user and the screen reader is one-sided and requires improvement.

In this paper I propose to study dialog generation for non-visual Web access to provide blind people with better ways of exploring Web content. This research will improve non-visual Web browsing by offering multi-layer dialog-based interaction with screen readers. The approach will build on features of the best screen-readers, extending the levels of interaction and information presentation.

ASSETS'06, October 22–25, 2006, Portland, Oregon, USA.
ACM 1-59593-290-9/06/0010.

2. RELATED WORK

To provide blind users with multi-layer Web access, the system will have to implement non-visual Web browsing; use a collection of methods for Web-page segmentation and analysis; and implement dialog generation and processing.

Non-Visual Web Access has been addressed by a number of research projects [9, 1]. Among the well-known screen readers there are JAWS [3] and IBM's Home Page Reader [8]. BrookesTalk [11] summarizes Web pages. Braille-Surf [2] converts HTML into Braille. These systems give their users reasonable facilities to browse the Web. However, most screen readers present information sequentially and provide little control over the form of presentation. In contrast, I propose to give visually disabled users access to Web-page content at multiple levels, keeping all of the standard features of the best screen readers.

Web Content Analysis uses segmentation of Web pages into blocks of data, which are then analyzed, grouped, labeled, etc. Most of the techniques used for segmentation are either domain specific [8] or rely on sets of manually specified rules [10]. Some of these approaches are not suitable for dynamically changing Web sites. I build on the previous work on structural and semantic analysis described in [7, 5]. The geometric partitioning method I am using is fully automated and scalable over domains and does not depend on manually specified rules or domain knowledge.

Dialog Generation is a large area of research in NLP and substantial research has been done on various aspects of dialog generation, including VoiceXML dialogs. Mixed initiative dialogs and generation of cooperative responses using VoiceXML are described in [4]. An example of use of VoiceXML dialogs in Web browsing is described in [6]. However, dialog generation for non-visual Web Browsing is not a well-studied subject. My work is based on and extends the research described in [7].

3. APPROACH

Non-visual Web navigation can be viewed as a mixed-initiative dialog between a human and a Web browser, in which the person asks questions and the browser gives answers or asks for clarification. The browser can make suggestions and prompt the user for input, e.g. if a Web form has to be filled out. User input can be in text or speech. The new approach will allow to access Web pages at three different layers of dialogs. The system will generate all three layers and will allow the user to switch between these layers and browse the Web at the level, which is most convenient for a given task.

Basic Screen-Reader functionality is the first layer of dialog generation. At this level the interaction between the user and the screen-reader is limited to an event-driven dialog, in which the page is simply read out to the user. The user can press standard screen-reader shortcut keys that control the flow of the dialog, e.g. skip, pause, repeat, change voice settings, etc. All these shortcuts are implemented within the dialog interpreter, while the dialog itself can be a simple narration, in which the screen-reader has the initiative and the user only selects the dialog direction.

DFS/BFS navigation is the second level of content presentation, requiring more sophistication in generating dialogs. The user should be able to freely navigate on the Web page using breadth-first, depth-first, and mixed approaches. The user will be able to choose which part of the page to listen to, at which level of detail, as well as get brief summaries of the page parts. The implementation of this level will require a mix of simple partial-document summarization techniques to be able to abstract, classify, and label different types of data.

Domain-Specific level of dialog generation will require specialized templates, e.g. news, shopping, etc. The dialog generator will analyze the Web page content and try to fill the corresponding template with information. The resulting dialog will give the user a more structured view of the page. The implementation of this layer may also require the use of classifiers to determine to which domain a page belongs.

Dialogs will be enriched with *earcons*, special sounds that could help distinguish between plain text, links, visited links, taxonomies, etc. Dialogs will be made adaptable to the user's choice of vocabulary, verbosity, navigation style, etc. This will require the use of sophisticated speech grammars to be able to interpret a variety of user commands.

4. PROJECT STATUS

Infrastructure. This work is a part of the HearSay [7] project[1]. The HearSay system has basic non-visual browsing facilities and can perform simple structural and semantic analysis of Web page content. The system provides a frame tree representation of Web pages, which can be further analyzed, grouped, and partitioned. The frame tree content is used to generate VoiceXML dialogs that will then be processed by a voiceXML interpreter. To be platform independent, the system is being developed entirely in JAVA.

VoiceXML Interpreter. Since no suitable open-source VXML interpreter was available, I am implementing a custom VXML interpreter that will comply with the VoiceXML specification and allow both voice and keyboard input. However, the interpreter is designed to go beyond the specifications and provide more control over the dialogs, speech properties, and event-handling.

Dialog Generator. The HearSay system already implements the first layer of dialog interaction, supporting basic screen-reading and extended speech controls. The implementation of the second layer of dialog generation will start as soon as reasonable partial-document summaries can be obtained. Dialog templates have to be developed for the third layer of domain-specific dialogs. And, finally, the dialog generation system and the VoiceXML interpreter have to be extended to support adaptive dialogs.

Summarization. The second layer of dialog generation will require the use of summarization. Most of the research on Web page and document summarization concentrated on full- and multi-document summarization. I am currently investigating summarization techniques that will be able to provide intelligent labels, summaries, and abstracts of parts of a Web page. The choice of techniques will depend on the type of information in any given part of the Web page.

Evaluation. The research goals of this project have been formulated through collaboration with Helen Keller School for the Blind (HKSB) at Hempstead, NY. The design of the project is guided by individuals with visual disabilities who are teachers at HKSB. The ideas have been obtained and clarified in meetings with instructors and students of the school. A series of progressive evaluations will be performed by the students of the school as the system takes shape.

5. CONCLUSION

In this paper I proposed to research dialog generation techniques as part of the ongoing HearSay project [7] for non-visual Web browsing. The goal of my research is to provide better ways of presenting Web page content by means of audio. A multi-layer dialog-based interaction has the potential to make the Web more friendly and accessible for people with visual disabilities. The system can be also adapted to provide Web access by phone.

[1]Supported by NSF grants CCR-0311512 and IIS-0534419.

6. REFERENCES

[1] C. Earl and J. Leventhal. A survey of windows screen reader users: Recent improvements in accessibility. In *Journal of Visual Impairment and Blindness*, 1999.

[2] D. Hadjadj and D. Burger. Braillesurf: An html browser for visually handicapped people. In *Proceedings of Tech. and Persons with Disabilities Conf.*, 1999.

[3] http://www.freedomscientific.com.

[4] K. Komatani, F. Adachi, S. Ueno, T. Kawahara, and H. G. Okuno. Flexible spoken dialogue system based on user models and dynamic generation of voicexml scripts.

[5] S. Mukherjee, I. Ramakrishnan, and A. Singh. Bootstrapping semantic annotation for content-rich html documents. In *Intl. Conf. on Data Engineering (ICDE)*, 2005.

[6] http://www.internetspeech.com.

[7] I. Ramakrishnan, A. Stent, and G. Yang. Hearsay: Enabling audio browsing on hypertext content. In *Intl. World Wide Web Conf. (WWW)*, 2004.

[8] H. Takagi, C. Asakawa, K. Fukuda, and J. Maeda. Site-wide annotation: Reconstructing existing pages to be accessible. In *ACM Intl. Conf. on Assistive Technologies (ASSETS)*, 2002.

[9] G. Weber. Programming for usability in non-visual user interfaces. In *ACM Intl. Conf. on Assistive Technologies (ASSETS)*, 1998.

[10] S. Yu, D. Cai, J.-R. Wen, and W.-Y. Ma. Improving pseudo-relevance feedback in web information retrieval using web page segnmentation. In *Intl. World Wide Web Conf. (WWW)*, 2003.

[11] M. Zajicek, C. Powell, and C. Reeves. Web search and orientation with brookestalk. In *Proceedings of Tech. and Persons with Disabilities Conf.*, 1999.