FI MU

Faculty of Informatics
Masaryk University

The Architecture of the Blind User Oriented Hypertext AUDIS

by

Ivan Kopeček

FI MU Report Series **FIMU-RS-98-01**

January 1998

The Architecture of the Blind User Oriented Hypertext AUDIS

Ivan Kopeček

Faculty of Informatics
Masaryk University
602 00 Brno
Czech Republic

Botanická 68a, 602 00 Brno Czech Republic
E-mail: kopecek@fi.muni.cz

Abstract. The architecture of the blind user oriented hypertext AUDIS developed at the Faculty of Informatics, Masaryk University Brno, is described in the paper. AUDIS is developed primarily as a support, which would help visually impaired students to study various materials. It can also serve as an Internet browser.

1 INTRODUCTION

Speech synthesizer and "screen access" ("screen reader") software represent basic utilities used by blind users to obtain hypertext information. Although present conventional hypertext and fulltext systems are very sophisticated and user-friendly, they are not very convenient when they are to be used by visually impaired people. The reason is the graphical interface and absence of the features fulfilling special needs of the blind people ([1]).

In what follows we will describe the basic features of the hypertext system AUDIS developed for blind users.

The hypertext system AUDIS is primarily developed to provide comfortable accessibility to textbooks for visually impaired students. It maximally pays attention to the special needs of users and allows for comfortable customization. The system is supported by the utilities for text conversion and rendering. Next versions of AUDIS will allow the system to be used as World Wide Web browser as well.

Even though AUDIS is primarily developed for visually impaired people, the architecture of the system is designed in such a way that the system will be convenient even to a larger group of people with special needs, e.g. for print impaired people.

2 CONTROL OF THE SYSTEM

The user controls the system by means of speech commands combined with keyboard key commands. This access optimizes the effectiveness of the control by joining advantages of both methods. Every command exists in both versions, i.e. in speech command version and in keyboard key version. This is due to possible application of the system for print impaired people.

Speech commands consist of the key words and are recognized by built-in speech recognizer. The system enables customization of the commands by renaming key words or by defining alternatives.

To use the basic speech commands the recognizer need not be trained, it is automatically customized to the user voice. However, some simple training of the recognizer increases reliability of the speech recognizer when larger number of speech commands is used.

Sound is the main output of the system. It can be used in the form of:

- Synthesized voice produced by the syllable based speech synthesizer DEMOSTHENES ([4], [6]). This type of sound output is used for reading hypertext data. The used speech synthesizer applies basic prosodic features based on corpora analysis ([7]) and enables use of various voice types that can be configured by the user. Various voice types are used to distinguish various types of information (e.g. normal text, underlined text, links etc.).
- Sampled voice, which is used for all standard messages to the user. Even sampled voice output can be configured. Various types of sampled voices are used to help the user to distinguish various types of messages.
- Sound generated by the sound synthesizer, MIDI, wave tables or special samples. This type of non-speech sound will be used first for environmental sounds applied to provide feedback to user actions ([2]), secondly it can be used for earcons (non-speech glances used to give the blind user an overview by listening, ([8]).

3 HYPERTEXT DATA STRUCTURE

The following important features influence the data structure:

- Maximal HTML compatibility. This is not only due to Internet but also because of general compatibility of data.
- Maximal possibilities for customization of the system and data.
- The primary goal of the system is to provide comfortable accessibility to students textbook material for visually impaired students - the possible use of AUDIS as a Internet browser has secondary meaning.

There are the following types of data files that are used by AUDIS:

- HTML data; more precisely, the system will accept only a defined subset of HTML (depending on the version of the system). Constructions and expressions outside the defined subset will be ignored.
- System configuration files are used for general customization of the system.
- Data customization files are used for user customization with regard to the data (text information included in the hypertext).

- Task definition data. It is useful to provide the system with the information whether it is used as a local hypertext system supporting textbook material or as an Internet browser.

4 DATA PREPARATION

Hypertext data should be carefully prepared to be used by the visually impaired students in accordance with the system data structure and with the basic features of the system (see Section 6). Although a simple structure based on the defined subset of HTML is acceptable for the system, some other text should be added to the basic text and some additional pieces of information should be included to make the data more comfortable for the visually impaired users. This includes:

- Global modification of the text. The number of the pieces of information that are not convenient for sight impaired (graphics, mathematical expressions, tables etc.) should be reduced and/or explained, if possible, by means of textual form ([3]). Shape of the text should be modified, if possible, respecting the users needs.
- Rendering and interpreting of graphics, tables, mathematical expressions and other non-textual information.
- Defining underlined sentences (sentences containing important information will be phonetically accentuated) and other meta-information in accordance with the data structure conventions.
- Adding summaries in accordance to AUDIS hypertext data structure conventions (see Section 7).
- Reducing the number of links and references, if possible.
- Building the tree structure of the system from basic linear structure; this is supported by the automatic generation of the optimal tree structure (see the next Section 5).

Some other pieces of information will be generated automatically (e.g. earcons, sound glances, environmental sounds).

5 DECISION TREE CONSTRUCTION

Although the information structured by hypertext links is convenient for visually impaired people, too frequent links may decrease lucidity of the text.

To make the structure of the system as simple as possible we use structure of hypertext system consisting of menus (menu pages) and chapters (text pages). Menus contain a title and links only and form the decision structure of the system. Chapters are linearly ordered elements of the text. The links included in the chapters should be used only inside the sequence of the chapters (they should not be used to address menus).

Such a structure is more transparent and lucid for visually impaired users than a general net structure. Formally such structures can be represented by rooted trees (root of the tree is the main menu of the system, inner nodes are menus and leaves correspond to chapters).

When a hypertext structure is created from the basic linear text items we have to define the structure of menus of the decision tree. It is clear that it should be done in such a way that the resulted structure is in some sense optimal from the users point of view. The optimality criterion is motivated by the requirement of the minimal average number of menu items that must be read by the system and evaluated by the user to find the required text page.

The problem of finding optimal rooted trees for such optimality criterion is solved in ([5]). Let us remark that in the quoted paper the criterion is formulated for another problem with different motivation, but it leads to the same formalization as in the present case. Let us briefly recall the main relevant result.

We consider all sets to be finite (card(m) denotes the number of elements of a set m). Rooted trees are denoted by ordered pairs R = (V, A) of vertices (nodes) and arcs (edges). We use also notation V(R) for R and A(R) for A. A root is denoted by r. We denote the set of all successors of a node x by q(x) and by L(R) the set of all leaves of the tree R.

Let R = (V, A) be a rooted tree, $x, y \in V$, and suppose that there exists an (n+1)-tuple $(x_0, x_1, ..., x_n)$ such that $(x_{i-1}, x_i) \in A$, $x_0 = x$, $x_n = y$. The set $\{x_0, x_1, ..., x_n\}$ will be denoted P(x,y) and we use notation $y < x$. If $x_0 = x = r$ then n is called level of y and is denoted lvl(y). Further, lvl(R) will denote the number $\max\{lvl(x); x \in V(R)\}$.

Definition of Optimality. Let R = (V, A) be a rooted tree, $x, y \in V$, card(V) > 1 and $y < x$.

We put $w(x,y) = \sum_{z \in P(x,y)-\{y\}} card(q(z))$ and $W(R) = \sum_{z \in L(R)} w(r, z)$;

The function W represents optimality measure. We shall say that a rooted tree R = (V, A) is W-optimal, if for each rooted tree R' satisfying L(R) = L(R') the condition $W(R) \leq W(R')$ holds;

Definition. A rooted tree R = (V, A) is called L_3 tree if it satisfies:

(1) card(q(x)) < 4 holds for every $x \in V$,

(2) card(q(x)) < 3 implies $q(x) \subseteq L(R)$.

Balanced quasi-ternary tree is an L_3 tree that satisfies:

(1) lvl(x) < lvl(R) - 1 implies card(q(x)) = 3;

(2) lvl(x) = lvl(y) = lvl(R) - 1 implies |card(q(x)) - card(q(y))| < 3;

(3) card(q(x)) $\neq$ 1 for every $x \in V$.

Theorem ([5]**)**. Balanced quasi-ternary trees are W-optimal.

This result makes it possible to propose the decision structure of the hypertext in an automatic and optimal way (in the form of balanced quasi-ternary trees).

6 BASIC FEATURES OF THE SYSTEM

In this section we will present the basic principles on which the system is built from the user point of view.

- The system enables comfortable control by means of combination of speech commands and hot-key commands. Speech commands are supported by speech command thesaurus that allows to express a command in several ways, making the control of the system more intuitive.
- Easy customization and configuration are important features of the system. This is related to the control commands, mode and type of speech synthesis output, graphic output for print impaired people with sight, information data structure, and other properties and options of the system.
- It is important to enable the user to obtain the information quickly and to get always an informational overview. This is supported first by various output speech modes and output speech rates, second by speech summaries, audio glances, earcons and environmental sounds.
- The orientation of the user is supported by the information about the position that is always accessible in speech form as well as in the form of audio glances, earcons and environmental sounds.
- User's speech remarks can be assigned to each information unit (menu, chapter, section, sentence). Each information unit can be underlined or masked. The user and system information attributes (masking, underlining, speech mode and voice types etc.) are always saved and can be always reset.
- The system is provided with tree, linear and fulltext type of searching and with the database of speech bookmarks. Summaries of the related information unit are always associated to a bookmark and can be used for orientation.
- Control of the speech output is fully supported. Each information unit can be skipped or repeated, the speech output can be stopped, speeded up or slowed down.

7 INFORMATION UNITS OF THE SYSTEM

There are four basic information units of the system (here we suppose the data in the form of textbook): menus, chapters, sections and sentences. Menus form tree decision structure (its creation during the stage of data preparation is supported by utilities that yield optimal decision tree for given number of chapters), linearly ordered chapters are leaves.

Chapters are divided into sections. Menus, chapters and sections must be provided with summaries that consist of the text part and audio glances (earcons). The non-text parts of the summaries are generated automatically. The following scheme corresponds to the chapter structure.

Title of the chapter
Summary of the chapter
  Title of the section
  Summary of the section
  Text of the section
  .
  .
  .
  Title of the section
  Summary of the section
  Text of text section

The user may choose whether some types of summaries and audio glances will be masked (then they are not outputted automatically).

Sentences can be non-underlined, underlined, masked and non-masked (in all combinations). The user can overwrite the original attributes (as they were assigned in the time of data preparation) but the original values can be always reset. User-defined attributes can be also saved and reset.

8 CONCLUSIONS

We have presented the basic architecture principles and features of the developed blind user oriented hypertext system AUDIS. The system is developed and tested in close co-operation with the Czech Union of Blind and Visually Impaired. The further work will be oriented to the implementation of basic modules, creation of the data preparation programming support and developing methods for rendering non-text data into speech/sound output.

ACKNOWLEDGEMENTS

The author is grateful to Hana Bubeníčková (Union of Blind and Visually Impaired) and Karel Pala (Faculty of Informatics, Masaryk University) for reading a draft of this paper, for valuable comments, and helpful ideas. The development of AUDIS is supported by the Grant of the Czech Ministry of Education No. VS 97028.

References

1. Bubeníčková, H., Kopeček, I. (1997) Informacni systemy pro nevidome (Information Systems for Blind People, in Czech). Proceedings of *"URBIS '97"*, Brno, April 1997, 34-36.

2. Darvishi, A. (1996) World Wide Web Access for Blind People: Problems, Available Solutions and an Approach for Using Environmental Sounds. Proceedings of *ICCHP'96*, Linz, 1996, pp. 369-373.

3. Kahlish, T (1996) Improving Access to Hypertext Based Study Material for the Blind. Proceedings of *ICCHP'96*, Linz, 1996, pp. 229-236

4. Kopeček, I. (1997) Speech Synthesis of Czech Language in Time Domain and Applications for Visually Impaired. Proceedings of the *2nd SQEL Workshop on Multi-Lingual Information Retrieval Dialogues*, Pilsen, April 1997, pp. 141-145.

5. Kopeček, I (1997) Optimal Trees for Searching in Codebook. Proceedings of *SOFSEM'97, Lecture Notes in Computer Science*, Springer, 1997, pp. 456-464.

6. Kopeček, I. (1997) Syllable Based Speech Synthesis. Proceedings of the *2nd International Workshop Speech and Computer - SPECOM'97*, Cluj-Napoca, 1997, pp. 161-165.

7. Pala, K., Rychlý, P., Smrž, P. (1997) DESAM – Annotated Corpus for Czech. Proceedings of *SOFSEM'97, Lecture Notes in Computer Science*, Springer, 1997, pp. 523-530.

8. Stevens, R.D., Wright, P.C., Edwards, A.D.N., Brewster, S.A. (1996) An Audio Glance at Syntactic Structure Based on Spoken Form. Proceedings of *ICCHP'96*, Linz, 1996, pp. 627-635.

Publications in the FI MU Report Series are in general accessible
via WWW and anonymous FTP:

```
http://www.fi.muni.cz/informatics/report
ftp:   ftp.fi.muni.cz (cd pub/series)
```

Copies may be also obtained by contacting:
Faculty of Informatics
Masaryk University
Botanická 68a
602 00 Brno
Czech Republic