THE USE OF INFORMATION RICH WORDS AND ABRIDGED LANGUAGE TO ORIENTATE USERS TO THE WORLD WIDE WEB

M. Zajicek, C. Powell

Introduction

The work described in this paper was funded under the HEFCE initiative for widening access to Higher Education for disabled people, Zajicek & Powell (1997a) The problem of presenting the information content of graphical user interfaces using speech is addressed. Current provision in terms of Braille and speech screen readers tends to be unsatisfactory, Laux et al, (1996), Petrie et al, (1993), and Stephanidis et al, (1995), and makes no attempt to present information on the screen conceptually.

This paper describes and evaluates the use of the helper application WebChat which presents a model of the information content of World Wide Web pages using speech. In particular it uses information rich words and ‘abridged language’ to present information. This is useful for visually impaired people and can be used to present World Wide Web pages over the telephone.

The requirement for up to date information and the accessibility of large volumes of data is a fact of life in the 1990s. Quick orientation to Web pages is essential when scanning for information and a speedy decision must be made as to whether a page is useful or not. Users have been observed scanning standard Web pages for potential information content. For example it was found that they inspect headings and then links on a page. WebChat incorporates a Menubar which offers access to the headings, links and keywords, on the page. For further evaluation of the menubar see Zajicek & Powell (1997b).

This paper describes and evaluates further WebChat facilities designed to enable orientation to the Web page. They are based on the concept of extracting information rich words or sentences from the information contained in the page, and abridging the text so that it can be read out more quickly.

Concepts implemented in existing facilities

We have re-described the structure of a Web page so that it can be ‘read’ out to a visually impaired user over the telephone. This involved forming a conceptual model of the information content of the page and its interrelationships and presenting it in speech form. The Web page has been presented conceptually Zajicek & Powell (1997a) as a collection of textual objects (consisting of heading and accompanying paragraph) and links. WebChat provides access to conceptually different parts of the Web page using function keys and different synthesised voices to indicate conceptually different features of text such as headings, links and paragraphs of text.

A graphical interface is essentially concurrent. Many items of information and options are displayed so
that the user can engage with any one of them at any time. Importantly the user need only engage in those parts that interest her and ignore the rest.

Speech born information is essentially sequential. In WebChat we have enabled a similar concurrency to that in graphical interfaces by conceptually structuring the page to offer options such as lists of links, or lists of headings, lists of keywords found in the text, or to read out paragraphs of text etc. This is achieved by using a menubar which is accessed by speech or function key.

M. Zajicek and C. Powell are at the School of Computing and Mathematical Sciences, Oxford Brookes, University, Headington, Oxford OX3 OBP

Tel: +44 1865 483683 Fax: +44 1865 483666 Email: mzajicek@brookes.ac.uk

The conceptually different parts or objects are ordered hierarchically within heading and sequentially down the page. The HTML code which forms the Web page allows for several headings. This feature is used by WebChat to subset the document and give the user the ability to examine only the elements of text of interest.

A list of keywords was extracted from Web pages using standard information retrieval techniques for library searches as proposed by Luhn in Schultz (1968). These search for significant words in text and are based on the premise that the frequency of word occurrence in an article furnishes a useful measurement of word significance.

The headings links and keywords options have been evaluated Zajicek & Powell (1997b) and it was found that users found the headings list most useful for orientation to the page closely followed by the keywords. Links were less useful as one might expect because they are generally used to describe the contents of the pages that the page under question is pointing to, rather than describing the contents of the page itself.

**Further scanning enabling facilities**

The aim of the further facilities was to accomplish the equivalent of visual scanning of the text. Reading out blocks of text is slow and we wished to establish whether listeners could absorb the meaning of text without listening to all the words. The aim in abridging text is to provide shorter elements to be spoken out but with the content in tact.

Further facilities were:
The option for ‘Abridged text’ where noise words have been stripped out and embedded keywords are presented using different voice types to draw attention to them.

A selection of keywords presented in the abridged sentences in which they are found.

Abridged text is achieved by stripping out all noise words such as ‘and’, ‘the’, and those with a low frequency in the text which might imply that they are less relevant to the text.

An example page with scanning enabling facilities

The page below is the RNIB page on its equal opportunities policy.

Textual content of the page:

Equal Opportunities Policy.

RNIB is working towards becoming an equal opportunities organisation. It is committed to promoting equal opportunities and preventing discrimination. This policy applies both to service delivery and to its own employment practices.

RNIB recognises that as we progress towards this goal the equal opportunities policy will require regular review.

It is committed to oppose direct and indirect discrimination on the grounds of disability, ethnicity or national origin, colour, race, creed, gender, marital status, domestic circumstances, age, HIV status and sexual orientation. RNIB recognises there are groups in society who are disadvantaged through discrimination and will work to redress any inequality in its own service delivery and employment practices.

This policy recognises and incorporates the implications of legislation relating to equal opportunities and recognises and takes into account RNIB's existing policies. Nothing in this policy is intended to prevent RNIB from following best professional practice in recruiting the person most suitable for the job.

RNIB requires its employees, committee members and volunteers to comply with this policy.

RNIB recognises the need to develop procedures to support the implementation of this policy, for instance a training programme and a system of monitoring. These procedures will be incorporated into this as it develops.

This policy was approved by the Executive Council in December 1989
Equal Opportunities Policy. RNIB is working towards becoming an equal opportunities organisation.

Abridged key sentences:

Equal Opportunities Policy. RNIB working towards becoming equal opportunities organisation.
The further scanning enabling facilities were evaluated in terms of the user’s ability to judge how useful a page would be to them with particular information goals in mind and the perceived possibility of the processed text retaining the semantics of the original text.

Six subjects listened to sentences containing keywords, abridged text, and abridged sentences containing keywords for five different Web pages. They were invited to comment on their ability to judge what the page was about. They then listened to the actual text and were invited to comment on the similarity in meaning.

It was found that subjects were quite confident of what a page was about when using scanning enabling facilities. However when questioned further to establish what the page was actually about it was found that they often had misconceptions.

It was also found that word frequency techniques used for extracting keywords produced on average a useful list of keywords. However when applied to the extraction of less useful words for the purposes of text abridgement it was found that information rich words were frequently omitted from the text and the semantics of a sentence were substantially changed. This phenomena can be observed in the example above.

In addition abridged text created by our current word deleting methods was not significantly shorter than the original. The aim of providing shorter elements of text with the semantics in tact was not effectively achieved by this technique.

Further work will involve the development of algorithms which tag words for their semantic content to ensure that information rich words will remain within the abridged text.

References


Schultz, C. K.; 1968; H. P. Luhn: Pioneer of Information Science - Selected Works; Macmillan; London; Pg. 120.


Zajicek M., Powell C., 1997b, ‘Enabling Visually Impaired People to Use the Internet’, IEE Colloquium ‘Computers helping people in the service of mankind’, London

Back to publication list.