Can we improve Web Accessibility for users with below average cognitive capacity?

Jenny Jerrams-Smith   David Heathcote***

*University of Portsmouth, School of Computing, Burnaby Terrace
1-8 Burnaby Road, Portsmouth PO1 3AE, UK

**Bournemouth University, Psychology Research Group, DEC, Talbot Campus, Poole, BH125BB, UK

dheathco@bournemouth.ac.uk
Jenny.Jerrams-Smith@port.ac.uk

Abstract

Previous empirical work has suggested that performance in using the Web as an information resource deteriorates when the cognitive load imposed by the content of Web pages exceeds the cognitive capacity of the user [3], [7]. This problem is particularly pronounced in users with below average cognitive capacity. The work reported in this paper describes the design, rationale and development of an adaptive interface for the use of special needs and elderly users with reduced cognitive capacity. ABEL (AccessiBility and Effective Learning) conducts an assessment of the user's cognitive capacity and then selects and directs the user to Web pages which impose an appropriate level of cognitive load. ABEL incorporates an innovative means of measuring individual variation in user working memory capacity which is suitable for special needs and elderly users. ABEL is likely to improve Web accessibility for educational and recreational purposes particularly for groups for whom Web use has previously been found to be problematic.

Keywords : working memory, accessibility, special needs, adaptive interface, user model, Google

1. Introduction

The Web is a very extensive information resource which has the potential to be of considerable value to special needs users. However, in practice special needs users are often presented with sites that they cannot make full use of because the Web site developer has not included the appropriate functionality within the Web page. For example, blind users who have the text of a site read to them by a screen reader have no way of knowing the content of an image unless it is described within the HTML in which the Web page is written (using the alt attribute of an image tag). Although there are now legal requirements for companies to make their sites accessible to all users, there remain many sites that are not accessible to special needs users.

With this in mind, this paper considers methods of alleviating the problems for special needs

Web users who are cognitively impaired. In addition to those with innate cognitive impairments or simply below average capacity, there are also many people throughout the world for whom old age is accompanied by a diminution of cognitive function [1], [7]. For these people, the textual complexity of some web pages may overload their diminished cognitive capacity. This research project therefore aims to help users with limited cognitive ability to use the Web successfully resulting in higher levels of motivation and frequency of Web use.

The research has three principal objectives: First, to find a method to efficiently and unobtrusively identify individuals who require and can benefit from adaptive Web assistance. Second, to find a method to rate Web pages for their suitability for cognitively impaired users. Third, to put the first two together by designing, building and testing an adaptive interface that rates Web pages and finds and presents the most suitable Web pages to cognitively impaired users.

The aim of the ABEL adaptive system is therefore that it both identifies the users who have reduced cognitive capacity and finds the Web pages that are appropriate to them. It is extension, advancement and innovative application of the CAIN system [4]. CAIN was intended for use by university students who were learning academic domains such as cryptography and poetry. It relied upon human experts to rate Web pages according to their level of difficulty and level of detail. This enabled CAIN to present Web pages to the user in increasing level of difficulty and increasingly more detailed pages.

ABEL is said to be an adaptive system because it provides different results for different individuals. It does this by means of a User Modeller that identifies the individuals who have reduced cognitive capacity. In addition, it has a Google-Return-Sites Modeller that ranks Web sites returned by Google according to their cognitive complexity.

A qualitative difference between CAIN and the new adaptive interface (i.e. ABEL: AccessiBility and Effective Learning) is that ABEL is automated and can rate Web pages (returned by Google) for their suitability for users with cognitive impairments. It does this by calculating a measure for each page based on indicators of difficulty, such as the average number of words per sentence. However, the effectiveness of an adaptive interface is dependent upon the acquisition of accurate user information. Without such information the interface would be unable to offer any adaptive advice. The measurement of Working Memory Span (WMS) [2] provides an indication of individual differences in cognitive capacity. However, [3] demonstrated that some elderly users are unable to complete standard tests of Working Memory span like that of [6]. In order to address this problem White et al. sought to develop alternative measures of working memory capacity more suitable for cognitively impaired users.

2. Developing the User Modeller

Individual differences in working memory capacity have been used as an indicator of general cognitive ability. However, the standard methods for measurement of working memory span (WMS) were too demanding for some users. Therefore a new method of assessing approximate WMS was devised which records a user's performance on a range of relevant variables while they are engaged in simple computer tasks. In order to maintain high levels of motivation these tasks are presented in a game-like format. A simple game-like working memory task was devised that requires participants to type a word while simultaneously remembering the name of a picture they are later asked to recognise (recognition is generally more accurate than recall).

This software for finding users with cognitive impairments was tested on twenty three participants in order to evaluate its suitability for incorporation into the ABEL adaptive interface. Twelve of the twenty-three participants were over 65 years of age. The 16 participants who made at least one error completed 392 tasks and made 84 errors. Five distinct types of error were identified and this led to the conclusion that simple game-like tests can be used to differentiate between individuals in terms of cognitive capacity.

An experiment was conducted [7] whose goal was to investigate whether a game-like task would be able to identify individuals with very poor working memory capacity without using mental arithmetic tests (such as those used in Turner & Engle's working memory span test). If the task is capable of discriminating between users on the basis of working memory capacity, it may prove to be a useful component of the ABEL adaptive interface. It was anticipated that the majority of participants would have no difficulty with the task but that low working memory span users would find the task particularly demanding.

Working memory can be operationally defined as the temporary storage and concurrent processing of information [1]. Thus, unlike the measurement of passive short-term memory, working memory span tests must have a storage *and* a simultaneous processing component. For example, [6] measured working memory span by presenting participants with a series of nouns (e.g. snow, house, cat etc.). After the presentation of each word a simple arithmetic equation was presented and the participant was required to verify the solution as true or false (e.g. $5 + 7 - 3 = 9$, T/F?). The participant's task was to recall as many of the words as possible in the serial order in which they had been presented. Since participants had to store the words in memory while processing the arithmetic problems, this task clearly meets the requirement that a working memory task must have temporary storage and concurrent processing elements. However, one problem with Turner & Engle's test is that some participants may have particular difficulties with the arithmetic component of the task. Thus performance in the task may be partly influenced by individual differences in arithmetic ability rather than working memory capacity per se. For this reason a working memory task was devised which was largely independent of arithmetic and verbal ability. The storage component was the retention of object pictures in memory and the concurrent processing load was imposed by the typing task. The processing load was incrementally increased as a function of the difficulty of the typing task.

Twenty-three participants took part in the experiment. Twelve were over 65 years of age (classified as the elderly group). Nine attended a day centre for the elderly and three attended a computer course for seniors. Eight of the under 65 group (classified as the young group) were disabled and attended a day centre for the handicapped at least one day a week. Their disabilities were mainly caused by strokes, cerebral palsy or degenerative illnesses.

A number of 3-letter, 4-letter, 5-letter and 6-letter concrete nouns with no repeating letters were selected (the words-to-be-typed). Sixteen simple black and white line drawings were selected from a computer-drawing package. The name of one of these drawings was randomly allocated to one of the three-letter words, becoming the picture-to-be-chosen. The same procedure was used for a four-letter, five-letter, and six-letter word. Together these four words and pictures formed one trial. The procedure was repeated for four more trials. Two different computer screens were designed. On the first screen a sentence (for example "Type the word *car* and then point at a picture of a *chair*.") appeared using one of the words-to-be-typed (in bold) and its allocated picture. This was presented both visually and verbally as it was read aloud by the computer, using the synthesised female voice 'Victoria'. The sentence remained on screen for 5 seconds. The participant was then expected to type or point at the on-screen keyboard the word-to-be-typed (in the above example the word "car"). If the participant was unable to remember the word they could continue by pressing the blue "Carry on" button. This was the only item allocated a colour and was so designed to help non-readers. Once the last letter of the word had been typed (or the "Carry on" button pressed) the screen changed to one containing the sixteen line drawings, in four rows of four. The participant was required to select the picture corresponding to the picture-to-be-chosen word that had been given on the previous screen (in the above example the chair). The procedure was repeated for a four-letter word, followed by a five-letter word, and finally a six-letter word. This was followed by a blank screen and the participant was offered the chance to stop or continue to the next trial (5 trials in total). During the development of the software the position of the pictures on each picture screen was randomly allocated. The words typed, and the pictures chosen were recorded with the participant's name, the date and the time. The pressing of the "Carry on" button was also recorded. The software automatically recorded if the answer was right, wrong or forgotten, and the time taken. The first set of four items was taken as a training exercise, and help given if required. These results were discarded.

Seven participants (two in the elderly group, five in the young group) were correct 100% of the time. In the majority of experiments this "ceiling effect" corresponds to a 'scale attenuation effect' i.e. a weakening of the scale of measurement. However, in this particular experiment it was not only predicted in advance, but a part of the design. The "ceiling" was deliberately designed to be low so that special needs individuals that might need extra help when performing simple tasks in a standard interface could be identified, and their data studied in greater depth. Ultimately it was hoped that a test could be designed for use in a user modelling system that differentiated between individuals.

It was found that incorrect items could be classified into five categories: - forgotten word-to-be-typed; forgotten picture; mistyped word (letter missing or repeated letter); misspelled word (other incorrect words); and wrong picture.

The total number of errors in the different categories can be seen in Table 1. The 16 participants who made at least one error completed 392 tasks (out of a total of 616 tasks), and made 84 errors.

The average time each participant took to complete one task (including the trial tasks) varied from 10 seconds to 45.3 seconds. On average the elderly group took twice as long to perform each task (26.4 seconds) as the young group (13.1 seconds). They also completed fewer tasks (mean 22 items) than the young group who completed all the tasks.

Table 1: The total number of errors and their breakdown into 5 classifications.

Total number of errors	Wrong picture	Spell error	Typing error	Forget picture	Forget word
84	25	13	16	20	10
100%	30%	15%	19%	24%	12%

An analysis of the number of the picture-to-be-remembered items was performed. The mean of the percentage of the pictures lost (wrong pictures + forgotten pictures) was found to be 21.9%. Participants who had a higher percentage error in the number of pictures lost than the mean were classified as having a low working memory. The rest were classified as having a high working memory. Eight participants were classified as having low working memory, two from the young group, and six from the elderly group. In a previous set of experiments 10 of the participants had taken a [6] type working memory span test (for information on these experiments see [3] and [7]). The results from these tests classified these participants as either having a low or high working memory span. The mean number of errors in words-to-be-typed with 6 letters was 3.75 errors, which is 250% higher than errors in the words with 3 letters. Similarly the mean of the pictures lost associated with the words-to-be-typed of 6 letters was 3.75, which is 150% higher than the loss of pictures associated with words of 3, 4, or 5 letters. This gives some tentative support to the hypothesis that for some individuals failure to chunk the words-to-be-typed causes an increase in cognitive load and a resulting increase in picture loss as the 'magical figure' of 7 items-to-be-remembered [5] is approached.

Of the 10 participants who also took the Turner & Engle type working memory span test six had the same classification on both tests. Spelling difficulties caused one young participant (time taken 13 seconds) to be classified as low on this test. However, numeracy proved less of a problem for him and he was classified as high on the Turner & Engle test. The three classified as high on this test but low on the Turner & Engle test (time taken 13.8, 14.7 and 22.7 seconds) experienced no spelling difficulties and were able to chunk the letters together resulting in a reduced cognitive load.

As expected the majority of the participants experienced little difficulty in performing the tasks. Fourteen of the twenty-three participants made three or less errors, and completed all the tasks. However, four of the elderly group (time taken 28.9, 34.1, 35.2 and 42.6 seconds) and one of the young group (time taken 21.4 seconds) demonstrated extremely poor working memory performance (25% or more pictures lost). Other participants had problems typing the words and it was found that a set of heuristics could predict participants who have problems spelling simple words (time taken 13, 21.4, 35.2 and 42.6 seconds) and would benefit from a word prediction or voice recognition program.

Simple game-like tests can therefore be used to identify users who as a result of reduced cognitive capacity could benefit from viewing only those Web pages selected by ABEL on the basis of their relatively lower cognitive load. We are currently incorporating these findings into the ABEL interface.

3. Developing the Google-Return-Sites Modeller

The ABEL adaptive interface is a front-end to the Google search engine and contains two major components : first, a User model, with its associated User modeller; and second, a Google-Returned-Sites modeller and its associated Google-Return-Sites model. The User modeller provides the user with a test based on that described above and stores the result of that test (i.e. whether or not the user has cognitive impairment) in the User model. The Google-Return-Sites modeller takes the first ten results returned by Google and constructs a Google-Return-Sites model that ranks the ten Google-Return Sites from most suitable for a cognitively impaired user to least suitable.

The ranking of the ten Google-Return sites by the Google-Return-Sites modeller is informed by research discussed by [2]. This empirical work shows that for a wide range of readers of differing abilities there is a correlation between reading comprehension and Working Memory Span (WMS). In addition, a number of studies have demonstrated that there is a difference in disambiguation ability between readers having high and low WMS. High WMS readers are able to resolve an ambiguity in text even if there are several sentences between it and the disambiguating information. However, low WMS readers cannot do this. For instance, they studied children who were read stories with an inconsistent ending (eg that a mother praised her son for not giving sweets to his sister). If the disambiguating information (that his sister is on a diet) is given several sentences later, the high WMS children can afterwards give the appropriate explanation of the inconsistency whereas the low WMS children cannot.

We therefore decided to focus the ranking of the ten Google-Return pages on the resolution of ambiguity. We theorise that the greater the use of pronouns, the less comprehensible a (Google-Return) page will be, because the reader is forced to hold onto previous information. Likewise, we theorise that shorter sentences will be easier to comprehend than longer ones.

For each of the returned sites, the Google-Return-Sites modeller works out the average number of pronouns per sentence and the average number of words per sentence. Therefore the ranking of the Google-Return-Sites model is based on the rating that each site receives, according to its difficulty of comprehension, in particular focussing on disambiguation. Working memory appears to play an important role in the syntactic analysis of sentences for meaning with the result that individual differences in user working memory span may interact with the syntactic complexity of text pages. Therefore one further goal is to rate sites on the basis of their overall syntactic complexity which in turn will be based on a number of linguistic variables.

We are continuing to develop the ABEL adaptive interface, providing a front-end to the Google search-engine. ABEL uses its User modeller and User model to classify each user as having either high WMS or low WMS. ABEL uses its Google-Return-Sites modeller and Google-Return-Sites model to rank sites according to their comprehensibility. It is hoped that combining the two strands of this research will ultimately assist individuals with reduced cognitive capacity be enabling them to find the Web more accessible and more useful.

References

[1] Baddeley, A. D. (1986). *Working Memory*. Oxford University Press.

[2] Baddeley, A.D. (1990). *Human Memory: Theory and Practice*. LEA.

[3] Jerrams-Smith, J., Heathcote, D., & White, L. (1999). Working memory span as a usability factor in a virtual community of elderly people. Proceedings UKAIS Annual Conference. 7-9 April 1999, York University, UK.

[4] Lamas, D., Jerrams-Smith, J. and Heathcote, D. (2000). Using directed World Wide Web navigation guidance : an empirical investigation. Proceedings Ed-Media Annual International Conference.

[5] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review*, **63**, 81-97.

[6] Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, **28**, 127-154.

[7] White L. (2002). Working Towards the Provision of ‘Everyone’ Interfaces for Elderly and Disabled Users of ‘Everyday’ Computers. Unpublished PhD Thesis: University of Portsmouth.