An Evaluation of DIADEM Assisted Online Form Completion

<table>
<thead>
<tr>
<th><strong>Journal</strong></th>
<th><em>15th Americas Conference on Information Systems</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>AMCIS-1005-2009</td>
</tr>
<tr>
<td><strong>Submission Type</strong></td>
<td>Paper</td>
</tr>
<tr>
<td><strong>Mini-Track</strong></td>
<td>E-Inclusion for Older Citizens &lt; Emerging Issues in IS Research</td>
</tr>
</tbody>
</table>
An Evaluation of DIADEM Assisted Online Form Completion

Author 1

Author 2

Author 3

Author 4

ABSTRACT: The DIADEM project aims to develop a web-based application in the form of an Expert System (ES) to assist cognitively impaired older-adult users in the task of interacting with and completing online transactions. Having recently developed the first experimental version of the application, this study reports on the preliminary findings of user trials carried out in three European countries to evaluate this early version of the application. Of the 94 users that took part in the trials, 77 were identified as users that were likely to present with some degree of mild cognitive impairment, and thus were included in the analysis stage. The key findings of the study indicate that users of DIADEM assisted form filling seemed to report comparatively high levels of satisfaction, particularly when considered against what is considered a typical level of satisfaction for this user group. Furthermore, as a result of a statistical analysis, the application appears to provide significantly increased levels of assistance for users presenting with higher levels of cognitive impairments, and therefore achieves its goal of catering for this particular target user group.

Keywords
Accessibility, Assistive technology, Older-adults, Online-forms, Web-based transactions.

1. INTRODUCTION

The European population is ageing, and as a result, the proportion of older adults is on the increase. In 2007, the European old-age-dependency-ratio was 25.2%, by the year 2055 it is projected to rise to 52.5% (Eurostat 2009). Ageing is often accompanied by declines in the senses mobility and cognitive ability (Smith et al. 1996; Mioshi et al. 2006), as a result it is increasingly common for older-adults to be living within their homes with restricted access to social services, and carrying out everyday tasks such as bill payment, banking and grocery shopping (Grundy 2003). Governments across Europe now see online services as providing new opportunities for older-adults to maintain levels of access to a range of services despite the effects of ageing (Money et al. 2008).

Until recently, paper-based forms have been the primary means of accessing government services, such as welfare, housing and tax returns. Users are now being expected to complete web-based equivalents of these forms (online-forms) to access these services. A large proportion of online-forms however, are not designed specifically for the needs of the older-adult and thus present significant usability challenges (Czaja and Lee 2008). Research to date has typically focused on developing general web-content design guidelines, and in some cases, has considered the needs of the older-adult to be fully incorporated by the needs of disabled users. For example, the World Wide Web Consortium Web Content Accessibility Guidelines (WCAG) (W3C 1999) provides 14 guidelines for the design of general web-content for disabled users. The United States general services administration section 508 (1998) also provides standards for supporting inclusive access for disabled users. The National Institute of Ageing (NIA), recognises older-adults as a user group in their own right, and presents twenty-five guidelines for general web-content for users aged 60+ (Morrell et al. 2002). Online-forms however differ from general web-content due to their predominantly question and answer based structure, and thus the usability challenges that arise as a result of interacting with such forms differ from those posed when browsing web-content in general (Lines et al. 2007). However, there seems to be very little research presenting online-form specific design guidelines, and an even smaller amount specific to the older-adult user (Sayago and Blat 2007). In terms of the older-adult user group, only three small scale studies were found in research literature that present design guidelines for online-form content. Lines et al. (2004) and Lines et al. (2007), carried out small scale studies, in the UK. The most recent of these (Lines et al. 2007) validated results from a previous study (Lines 2004) and as a result presented 13 online-form design guidelines. Sayago and Blat (2007) also carried out a small scale study, involving one online-form and seven Spanish older-adult users. The
findings revealed that users prefer checkboxes and radio-buttons as opposed to list-boxes. They also required optional fields to be grouped into separate sections. Furthermore, research relating to the general web-content guidelines that exist, found that the majority of e-Government web-content still fails to conform to the most basic design guidelines (Choudrie and Ghinea 2005).

Older-adult users are likely to benefit from a client-side application that adapts and personalises online-form content so that it better conforms to web-content and online-form design guidelines. Such an application would provide older-adult users with a standardised and familiar interface, which would enable them to access online-form content more efficiently and effectively. The Delivering Inclusive Access to Disabled and Elderly Members of the community (DIADEM) project aims to achieve this by developing a web-based application that adapts and personalises existing online-forms based on the individual user’s preferences, and interactions with on-screen content. Having recently completed the development of the first DIADEM prototype application, we have carried out user trials across Europe to evaluate the usability of DIADEM enabled online-forms, and to establish whether the ES rules used by DIADEM are sensitive to cognitively impaired older adults’ needs. The trial design and associated results are presented in this study.

The remainder of this paper is structured as follows. Section 2 provides a conceptual description of the proposed DIADEM application. Section 3 presents details of the user trials carried out to evaluate the application. Section 4 presents some preliminary key findings of the user trials. Section 5 concludes the paper.

2. OVERVIEW OF DIADEM ARCHITECTURE

As far as the user is concerned, the DIADEM application operates as a “client-side” extension to their web browser. The application has an Expert System (ES) element that monitors user interactions, and personalises the user interface to alleviate usability challenges faced by the user. DIADEM operates alongside existing web-architectures. Thus, the DIADEM web-services interface serves as an intermediary between existing application software, the DIADEM server, and ultimately the user or client site. The client component is data driven by XML control files from the server, so that it can interface to a range of services. From the client side, the DIADEM plug-in makes it possible to carry out the final enablement and transformation of existing online-forms into the standardised DIADEM online-form user interface. Figure 1 provides a conceptual overview of the DIADEM application architecture and how online-forms rendered in a standard browser are transformed into a DIADEM enabled form.

Figure 1: DIADEM application architecture

The DIADEM user interface and its functionality is developed directly from existing older adult online-form requirements specifications (blind ref) as well as online form design guidelines derived from DIADEM user trials (blind ref). It provides users with a consistent and familiar interface which conforms to online-form design guidelines developed specifically for the older-adult target user group. The ES element of DIADEM monitors interactions whilst the user accesses online-form content, which may include: user inactivity; typing speed; key stroke accuracy; mouse click accuracy; frequency of backward navigation through the dialogue; frequency of
submission errors; question completion rates. The system analyses these interactions, and attempts to diagnose when the user appears to be in a problem state. Certain sequences of user interaction activity trigger rules in the DIADEM ES rule set, which result in the system providing appropriate and personalized assistance to the user.

3. USER TRIALS

User trials were carried out across Europe to evaluate the recently developed first experimental version of the DIADEM application. The primary goal of the trials was to effectively identify cognitively impaired older adults, and establish whether the ES rules used by the DIADEM application were sufficiently sensitive to cognitively impaired users interactions. In this section, a description of the trials design is presented.

3.1. Engaging users

Initially, local councils were approached for advice on sourcing potential participants for the trials. As a result, contact details for a number of voluntary organisations and contacts within local hospitals were obtained. In total 94 participants were recruited to take part in the trials across 3 different countries. The main source of participants was from voluntary organisations who seemed enthusiastic about the project and very willing to provide access to large numbers of older adult volunteers who regularly offer their time to take part in research. Although some participants were sourced from local hospitals, patient confidentiality, and constraints on staff time made participant recruitment more difficult via these sources. Clear communication with candidate older adult participants is important, to avoid potential confusion, and to ensure that participants are fully informed and aware of the requirements of the task (Dickinson et al. 2007). Initially a trial information sheet and covering letter was sent, outlining the goals of the study and the tasks they would be asked to perform. In the first instance, this information was often provided on our behalf, via the voluntary agency. Participants then contacted the voluntary agency if they wished to take part in the trials, and gave consent for their contact details to be passed on so that further arrangements could be made by telephone. Any questions and clarification was then provided to participants verbally. All participants were offered free transport to and from the venue. A confirmation letter, including clear directions to the venue, and the agreed time and date of the trial session was sent to all participants. A week before the trials commenced, participants were telephoned, to remind them that they were not obligated to take part in the study, and to confirm attendance and travel arrangements.

3.2. Selecting appropriate online-forms

During the trials, users were asked to complete a DIADEM enabled online-form. It was important that the online-forms used in the trials were sufficiently 'complex' to ensure that a broad range of user scenarios were elicited as users interacted with the online-forms, and that the DIADEM application demonstrated that it was capable of successfully transforming such forms. However, at present there is little research into evaluating online-form complexity for this user group. Consequently, prior to the trials, all partner countries carried out the Bespoke Online-forms Selection (BOFS) procedure which we specifically designed to identify appropriately complex online-forms for the older adult target user group.

BOFS is a solution for assessing a range of generic online-form criteria including: the total number of pages in the online-form; the amount of scrolling required within a form; whether the online-form gives an idea of the user’s location within the context of the whole form and so forth. According to Miller & Jarrett (Miller and Jarrett 2001), there are five HTML input mechanisms, these being: drop-down boxes, radio buttons, check boxes, hyperlinks, and type-in boxes. Since online-forms are primarily concerned with users inputting information, BOFS also assesses online-form complexity according to these five types of input mechanism, and thus can be applied to any online-form that has been designed in standard HTML 4.0 format, regardless of the genre of the online-form or indeed the specific subject content of the online-form. The generic nature of BOFS means that it is a versatile tool that can be used in a wide range of contexts, which is particularly useful when assessing online-forms in a variety of languages, or indeed across a variety of subject content. A detailed description of the BOFS method and case examples of how it was implemented in the DIADEM trials is provided in (blind ref).

As a result of carrying out the BOFS assessment, four appropriately complex online forms were identified and used in the trials. These are as follows:

- Italy (City of Turin): Self-certification about the civil status of people form.
- UK (City of Sheffield): Citizens to change address form.
- Norway (BlueGarden): Travel expenses form.
- Norway (More): Safety alarm application form.
3.3. Procedure

Initially, informed consent was sought from each user attending the trial session. Each was then asked to complete the ACE-R exam (Mioshi et al. 2006), which typically took less than 10 minutes to complete. In addition to the background information collected as part of the ACE-R exam, such as number of years spent in full time education, and date of birth, users were also asked to self-report whether they had noticed a decline in cognitive function. This information would be used later, to identify user data that was appropriate to be included at the analysis stage.

For the main trials task, users were presented with a DIADEM enabled online form, and were asked to complete it. Users completed the online form individually in laboratory conditions, the only other person present in the room was the researcher conducting the trial session. No time constraints were imposed on users, who were instructed to take as long as they needed to complete the form. A log-file was generated, recording each user’s interactions with the online-forms for each trial session. The data recorded in the log-file served as a valuable source for gaining further insights into the older-adult’s interaction experiences. One of the key features of the DIADEM application is to provide assistance to the user by monitoring user interaction measures, and triggering rules from the ES rule set, which result in triggering the system to provide the user with pro-active assistance.

The log-file data contained a record of the number of times ES rules that were triggered throughout each session. Therefore, this made it possible to carry out an evaluation of the frequency of ES rules triggered for each individual user, compared with their respective levels of cognitive decline, which may be measured based on corresponding ACE-R scores. The outcome of this evaluation would reveal whether the DIADEM application appeared to achieve its goal in providing increased levels of assistance to users with comparatively high levels of cognitive decline. Other information recorded in the log-files included; session start time; session finish time; number of keystrokes; number mouse clicks; number of errors on first submission attempt; number of times the help function was accessed help. Although it was not anticipated that this data would be of immediate use for purposes of evaluating DIADEM version one, such data was still collected, as it likely to provide valuable benchmark data against which future versions of the DIADEM application could be evaluated against.

On completion of the DIADEM enable online-form, users completed a short satisfaction questionnaire. There are a number of tried and tested questionnaires, specifically designed to measure satisfaction of software usage. Some examples include, the Questionnaire for User Interaction Satisfaction (QUIS) (Schneiderman 1997), Website Analysis and MeasureMent Inventory (WAMMI) (Claridge and Kirakowski 2007), and Software Usability Measurement Inventory (SUMI) (Kirakowski and Corbett 1993). After reviewing a range of satisfaction questionnaires, the Software Usability Measurement Inventory (SUMI) questionnaire was chosen as most appropriate for these trials. It consists of 50 short and clearly worded statements, and can be administered relatively quickly (typically in less than 10 minutes). One significant benefit of SUMI is that the results can be compared with the benchmark SUMISCO database, which is a collection of over 3000 SUMI questionnaire results collected for a wide range of users and software applications.

4. KEY FINDINGS

After all trials had been completed, a number of data analysis tasks were carried out in order to understand the implications of the data collected. This section outlines the analysis and reporting activity undertaken, and includes a summary of the key findings.

4.1. Users data included in analysis

In order to ensure that the findings of this study was likely to represent users with some degree of mild cognitive impairment, user data was only included in the analysis stage if they match a minimum of one of three inclusion criteria:

1) User aged 70+. All users aged 70+ were automatically included at the analysis stage. The rate of cognitive decline is known to accelerate with age (Salthouse 1982). For example, a meta-analysis shows that by the age of 40, an individual’s cognitive speed is expected to drop by 20%, this is expected to drop further, to 60% by the age of 80 (Salthouse 1982). Therefore, it is considered likely that older-adults aged 70+ would present with some level of cognitive decline.

2) The user has had 12 years or less full-time education, and self-reported a decline in memory or cognitive functioning. A high level of education is believed to have a protective affect on cognitive decline (Albert et al. 1995). Therefore those who reported to have spent 12 years or less in full time education, were more likely to present with cognitive decline. Furthermore, users were asked whether they have noticed or have been aware of any decline in memory or cognitive functioning. Self-reporting of changes in cognitive functioning or memory is an accepted means of identifying
individuals who are suffering from some level of cognitive decline or cognitive disability (Reisberg et al. 1982).

3) The user scores lower than average on at least one sub-scale of Addenbrooke’s cognitive examination (ACE-R). Although the ACE-R is a relatively new screening tool, it is now widely accepted that the ACE-R is significantly more sensitive than more commonly used cognitive examinations such as the MMSE (Larner 2006). Like the MMSE, the ACE-R is still considered to be a brief, sensitive, and inexpensive screening tool that can typically be administered in less than 10 minutes. It is also shown to be more effective in picking early cognitive dysfunction/mild cognitive declines (Larner 2006) compared with the MMSE.

Of the 94 users that took part in the trials, 77 were included at the analysis stage of which 39 were female, 38 male, with an overall average age of 67.8. In the UK a total of 33 users were eligible, 21 of which were female and 12 were male. In Italy, a total of 31 users were eligible, 11 female and 20 male. In Norway, 13 users were eligible 7 female and 6 male. The pan-European sample comprised of 39 females, 38 males, with an average age of 67.8. Overall, 31 end-users were aged 70+, 31 end-users self-reported to have experience a decline in memory or cognitive functioning and were in full-time education for 12 years or less. A total of 58 of the 77 end-users were eligible for analysis by scoring below the ACE-R inclusion thresholds for one or more of the five ACE-R subscales. Figure 2 presents a count of the number of users eligible for each of the three inclusion criteria and their respective combinations for the user trials.

![Figure 2: User inclusion criteria](image)

4.2. Reported SUMI scores

By default, the SUMI questionnaire is purchased as a product that includes a data analysis service. The vendor required SUMI scores to be sent for analysis, presented in a predefined standardised template. An analysis of the results was returned by the vendor, which included a comparison of the DIADEM SUMI scores against the SUMISCO database, which is made up of over 3000 SUMI questionnaire responses, for a range collected for a range of software, and may be used as a benchmark against which scores can be compared.

SUMI reports results on five sub-scales (Efficiency, Affect, Helpfulness, Control, Learnability) and one overall scale (Global), which is the sum of the five sub-scales. Efficiency represents the extent to which users feel that the software helps them with their work. Affect relates to the user's emotional reaction to the application. Helpfulness is the extent to which the application is self-explanatory. Control measures how in control of the application the user feels. Learnability is a measure of how quickly the user feels they were able to master the use of the application. Table 1 presents the overall SUMI scores achieved by the DIADEM application for each of the five sub-scales and the overall Global scale.
Table 1: SUMI scores compared with SUMISCO database

<table>
<thead>
<tr>
<th>Scale</th>
<th>UF</th>
<th>Ucl</th>
<th>Median</th>
<th>Lcl</th>
<th>LF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>80</td>
<td>50</td>
<td>47</td>
<td>44</td>
<td>14</td>
</tr>
<tr>
<td>Affect</td>
<td>84</td>
<td>59</td>
<td>56</td>
<td>52</td>
<td>21</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>82</td>
<td>55</td>
<td>52</td>
<td>49</td>
<td>15</td>
</tr>
<tr>
<td>Control</td>
<td>77</td>
<td>51</td>
<td>49</td>
<td>46</td>
<td>20</td>
</tr>
<tr>
<td>Learnability</td>
<td>91</td>
<td>55</td>
<td>51</td>
<td>47</td>
<td>4</td>
</tr>
<tr>
<td>Global</td>
<td>84</td>
<td>55</td>
<td>52</td>
<td>49</td>
<td>17</td>
</tr>
</tbody>
</table>

SUMI scores have been transformed so that they may be compared against the SUMISCO database average, which fixed at 50 for each of the scales. The Median represents the middle score when the scores are arranged in numerical order. The Ucl and Lcl are the Upper and Lower Confidence Limits respectively. They represent the limits within which the theoretical true score lies 95% of the time for this sample of users. The UF and LF are the Upper and Lower Fences. They represent values beyond which it may be plausibly suspected that a user is not responding with the rest of the group.

Therefore the above results may be interpreted as follows. If a scale score does not encompass 50 within its Ucl/Lcl range, it may be considered 95% certain that it is above (or below) the database average. Sub-scales achieving Ucl/Lcl ranges above 50 indicate that the DIADEM outperformed the SUMISCO database, and that the application was comparatively well received by users. Conversely, sub-scales achieving Ucl/Lcl ranges below 50 indicate that DIADEM scored below average compared with the SUMISCO database. Sub-scales achieving Ucl/Lcl incorporating 50 indicate that DIADEM achieved similar scores to the SUMISCO database.

As can be seen, Efficiency, Helpfulness, Control, Learnability and Global all incorporated 50 in the Ucl/Lcl ranges. Therefore it may be inferred that DIADEM enabled forms achieved levels of user satisfaction, similar to the benchmark SUMISCO database. The Ucl/Lcl range for Affect was 59/52, which was above the SUMISCO database average, indicating DIADEM enabled online-forms on average were more likeable than the benchmark SUMISCO database scores. The Affect sub-scale achieved above average the SUMISCO benchmark. These results are encouraging, when considering that these scores were achieved from an older-adult user group, that were likely to present with some level of cognitive decline. Indeed, a study by Nielsen (Nielsen and Norman 2002), found that web-content is twice as difficult to use for older-adults aged 65 and over, compared with younger users. Hence, if DIADEM was not effectively catering for older-adult’s needs, SUMI satisfaction scores may well have been expected to be considerably lower than the benchmark SUMISCO database averages.

4.3. ES rule trigger frequency

The prototype DIADEM application trailed four high-level rules that allowed the DIADEM application to provide assistance to users as they interacted with the online-forms. In order for a high-level rule (or rule group) to be triggered, the system required a number of low-level rules to be triggered. Essentially, the four rules tested in these trials represented an aggregated system response to a number of lower-level rules being triggered. To follow is a brief description of each high-level rule, and how the logic used to trigger the responses during the end-user trial sessions was defined:

- **Rule groups 1 and 2** are designed to detect and respond to user activity where they appeared to have ‘lost focus’ on any of the form’s input fields.

- **Rule groups 3 and 4** are designed to identify when the user becomes inactive and provide a response that helps them regain concentration and continue. The guidance offered depends upon the point the user has reached in the transaction process and whether there are incomplete or incorrect items on the screen.

To gain an idea of the frequency of rules fired during end-user trails, the log-files collected for each of the 77 users included in the analysis phase were analysed. Figure 3 presents the results of this analysis, showing each of the high level rules fired, and the number of times each was triggered by the 77 end-users that took part in the UK, Italy and Norway trials.
Rule 3 was triggered 89 times during the trial sessions, the most frequent out of all rules. Rule 1 was next highest with 74 triggers, followed by Rule 2 with 51 triggers, and Rule 4 with 37 triggers. The results show that all rules that were anticipated to trigger during the user trials did so, and that no rules that were implemented failed to be triggered as a result of normal end-user interaction activity.

4.4. ES high level rule triggers versus ACE-R scores

One important aspect of the trials was to evaluate the rule triggering mechanisms employed in the DIADEM application. In particular, whether the DIADEM ES rule triggers are sensitive to users who presented with comparatively high levels of cognitive decline. One approach to exploring this further was to compare the frequency of ES rules triggered in each session with the ACE-R scores achieved by each respective user (which was used to represent level of cognitive decline). It was proposed that if the ES rules were being triggered effectively, then the number of ES rules triggered within a session would increase as the users ACE-R score decreased (the lower the ACE-R score, the higher the level of cognitive decline). The log-files recorded for the UK participants were analysed and the number of ES rules triggered were counted for each user. A Pearson's r correlation test was carried out, comparing the number of ES rules triggered with ACE-R scores achieved by each of the 33 UK users. Since we hypothesised that frequency of ES rule triggers would increase as the ACE-R scores decreased, a one-tailed test was used.

Table 2 shows the results of the test.

<table>
<thead>
<tr>
<th>ACE-R score</th>
<th>Number of ES rule triggers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>(-.298(*))</td>
</tr>
<tr>
<td>Sig. (1-tailed)</td>
<td>0.046</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
</tr>
</tbody>
</table>

As can be seen, there was a significant negative correlation between ACE-R scores and the number of ES rule triggers during a session \(r = -0.298, n = 33, P < 0.05\). Therefore the DIADEM application was successful in triggering significantly more ES rules for users that achieved comparatively low ACE-R scores (presenting with higher levels of cognitive decline). These results are promising, as they indicate that the DIADEM application appears to be sensitive to the user level of cognitive decline, and provides increased levels of assistance accordingly.
5. CONCLUDING REMARKS

This paper has presented details of user trials carried out across Europe to evaluate an experimental version of the DIADEM system to assist elderly users. The outcomes of the trials have demonstrated promising results. In particular, from a satisfaction perspective, older-adult users reported levels of satisfaction on par with the general population for all sub-scales of the SUMI satisfaction questionnaire, with the exception of the Ucl/Lcl range, for which the DIADEM application outperformed the general population scores. This is encouraging when considering that older-adult users typically find web-content twice as difficult to use compared with younger users (Nielsen and Norman 2002). An analysis of the frequency of ES rule triggers, also revealed that all of the high-level ES rules implemented appeared to trigger regularly during the trial sessions. UK log-file data revealed that the DIADEM application provides more frequent assistance to users that present with higher levels of cognitive decline. These are also promising results, which provide valuable support to the overall direction of the project. In particular, the findings indicate that the measures used to monitor user interactions, and the mechanisms used to trigger ES responses appear to be particularly sensitive to users that present with high levels of cognitive decline. Future work will involve further analysis of the data collected during these trials, and developing the DIADEM application further. In particular, the ES rule base will be developed further so that increased assistance may be provided to the target user group. This will be accompanied by one final round of user trials with a more comprehensive experimental version.

ACKNOWLEDGEMENTS

(Removed to blind paper).

REFERENCES


