The proactive retrieval system is interrupting my writing

Mari Carmen Puerta Melguizo\textsuperscript{ab}, Anita Deshpande\textsuperscript{a} and Lou Boves\textsuperscript{a}
\textsuperscript{a} Radboud University, Department of Language and Speech, P.O. Box 9103, 6500 HD Nijmegen, THE NETHERLANDS
\textsuperscript{b} Utrecht University, Center for Content and Knowledge Engineering, Department of Information and Computing Sciences, Padualaan 14, 3584CH, Utrecht, THE NETHERLANDS

Writing a scientific paper is an activity in which finding relevant information is essential. Seeking for relevant information is however not an easy task. The Proactive system IntelliGent retrieves and decides which available information is most likely relevant to the written text, and presents it peripherally without user intervention. The decision about what information to present is mainly based on the document currently being written. The information presented can improve the quality of the written text. However, presenting peripheral information can also interrupt the main task of writing. Our research question: is the presentation of information by IntelliGent distracting writers from their main task? We used writing-related tasks with different cognitive load, from high to low: writing, recall, recognition, and copying the title of a text. To explore the trade-off between extra time and quality, we also measured the quality of the written text. We expect to find that interruptions prolong the main task, especially when the task has high cognitive load, but the result is a better document when the information retrieved is relevant for the task in hand.

Keywords: Proactive Retrieval Systems, Interruptions, Cognitive Load.

1 \textbf{INTRODUCTION}

Finding appropriate information while writing a scientific paper is essential but this task is difficult and time consuming. In practice, writers use text editors to write, and keyword-based browsers to gather information. Switching from editor to browser imposes extra-demands on users’ cognitive capacities. A system that can diminish the cognitive load by relieving writers from explicit search and switching between applications would be most welcome.

Proactive Retrieval systems (PRs) retrieve and decide which available information is most likely relevant to the written text, and present it without user intervention. The decision about what information to present is mainly based on the context, in our case, the document that is currently being written. In addition, a global user profile may also play a role. Only a few systems have been designed for this purpose. For example, Watson automatically performs Web searches based on text being written or read in Microsoft Word [2]. Other systems are Letizia [5] and the Remembrance Agent [8]. The existing systems however, are not unequivocally appreciated by their prospective users, among other things because it is not known how and at what moments, information should be presented to be most productive and least disruptive.

The project \textit{À Propos} aims to develop a Proactive Retrieval system that supports users in writing documents, such as scientific articles, reviews, or technical reports. The system should diminish the risk that professionals miss essential information and reduce the time they spend searching for information, as it proactively retrieves relevant, personalized and trustworthy information. With this goal in mind and in collaboration with SEC [9], we are optimizing the PRs IntelliGent\textsuperscript{TM}. The system proactively submits queries and presents the retrieved information while the user is writing a document. The success of the system mainly depends on two factors. First, the retrieved information has to be precise and highly relevant to the written document. Second, the information should be presented in a non-intrusive and timely manner. If the way of presenting the information seriously interferes with the main task, users might stop using the system. Our goal is to explore both issues combining studies in the natural working environment with experimental research directed to explore some factors in more detail. In this paper we present a part of the research we are currently performing.

2 \textbf{THE IntelliGent\textsuperscript{TM} PROACTIVE RETRIEVAL SYSTEM}

While users are writing a text, IntelliGent presents the retrieved information in the visual periphery in the form of a transparent window at the bottom right of the screen (see Figure 1). The window basically contains URLs related to what the user is typing. Clicking in the URLs, the user accesses the corresponding papers from the digital library. The information in the window changes depending upon the text that is being
written and new queries are considered.

Fig. 1. IntelliGent presents peripheral information in the low right corner while writing a document.

To understand how and when users work with IntelliGent in a natural working environment, we conducted a study in which the usability issues of efficiency, effectiveness, and overall satisfaction with IntelliGent were addressed [3]. Users’ general opinion was that the system did not really help them in completing their writing tasks because IntelliGent often retrieved irrelevant information. Furthermore, users perceived that the system interrupted the task of writing and some users simply ended up stopping using IntelliGent. Based on these results we decided to explore in more detail if the presentation of information by IntelliGent was distracting writers from their main task. And even if working with IntelliGent interrupts the writing task, is the quality of the final paper improved when the system presents information that is really relevant?

3 THE EFFECTS OF INTERRUPTIONS IN TASK PERFORMANCE

Presenting peripheral information can provide the user the opportunity to learn more, to do a better job, or to keep track of less important tasks [5]. On the other hand, peripheral informing imposes extra-task demands on user’s cognitive resources and can interrupt the main task. Bailey et al, [1] actually used the concept of computer-initiated interruption to refer to any computer-based task presented to a user by a computer application while the user is performing another computer-based task. Interruptions are often produced because the additional task (i.e. to check if the offered papers are interesting to the writer) has specific demands and there is the need to resume the original task at a later moment. This reorientation task requires the user to remember the status of the primary task (i.e. to write the document). The effects of interruptions on user’s main task performance have being studied frequently [1, 7, 11]. For example, Bailey et al, [1] used six main tasks with different memory load and two interruption tasks. They found a degradation on the time spent on task performance when interruptions where presented. The authors concluded the cause of performance degradation is the additional time needed to resume to the main task.

We performed a preliminary study to explore if the interruptions caused by the presentation of information by IntelliGent required extra time of the users to resume the main task of writing. We also wanted to see if the presentation of related papers, rather than unrelated ones, could possibly have a positive impact on the perceived usefulness of the system by the users In order to check these issues we decided to manipulate the interruption task. In one condition IntelliGent presented related papers to the writer and in another condition unrelated papers were presented.

Bailey et al. also found that tasks with high cognitive load were more affected by interruptions than simpler tasks. Since different writing-related tasks seem to have different cognitive load we decided also to explore this issue. From high to low cognitive load, we included the tasks of writing abstracts, answering questions (recall), recognition of the correct answer and copying the title of an article. We assumed the task of copying a title is the least demanding. Then recognizing the correct answer in a multiple choice question is easier and less demanding than recalling [e.g. 9]. And writing is the most demanding task involving recalling information, planning, translating and reviewing [4].
3.1 Method

Participants. Eighteen subjects participated in the study (10 male). All subjects were students of Information Science (University of Utrecht) and randomly chosen. Students received 5 euros for their participation. The average age was 22.

Design. Two independent variables were manipulated within subjects: main task category and interruption task. The dependent variables were the time on task (TOT), time on interruption (TOI), perceived quality of the performance of the main task (PQT) and perceived usefulness of the information presented by IntelliGent (PU).

Main Tasks. The tasks were selected according to their degree of cognitive load. From high to low:
- Writing an abstract. Participants had first to read a short passage (4-5 sentences) and the task was to write a short passage about the topic they read. For example, whenever they read about video games, participants had to write a short text about this topic so previous knowledge about the topic could also be used.
- Recalling information. A short passage (4-5 sentences) was presented and the task was to read it and then answer three open questions regarding the content.
- Recognizing information. A short passage (4-5 sentences) was presented and the task was to read it and answer three multiple choice questions regarding the content.
- Copying the title of an article. A short passage (4-5 sentences) was presented and the task was to read it and then copy the title of the article.

Interruption Tasks. Three different options were presented:
- No interruption task. In this condition, participants performed the main task without the presence of IntelliGent.
- Related interruption task. When performing the main task, IntelliGent presented articles about the same topic of the main task.
- Not related interruption task. When performing the main task, IntelliGent presented articles about a completely different topic.

The Dependent Variables
- Time on Task (TOT). The amount of time spent performing the primary task (in seconds). This measure did not include the time spent on an interruption task.
- Time on Interruption (TOI). The amount of time spent on an interruption task (in seconds).
- Perceived Quality of the Performance of the Main Task (PQT). Measured with questions based on a 5-point Likert scale (1 meaning low and 5 meaning high).

Procedure. For each task participants had to read first a brief text about one of four possible topics (Alzheimer’s disease, mobile devices, Shakespeare, and games). The texts were abstracts of articles published in different popular science journals. After they had read the abstracts, participants were asked to perform the four main tasks using MS-Word.

Conditions were counterbalanced across subjects and each subject completed the four main tasks, two without interruptions and two with interruption tasks. In the interrupted conditions, we used a mock-up of IntelliGent that was manually controlled by an assistant. Somewhere in the middle of performing the primary task, IntelliGent appeared in the right low corner of the screen with four URLs. Participants were asked to open a specific URL and pay attention to it. Half of the cases participants were presented with articles related to the topic they were performing the main task (e.g. writing an abstract about mobile devices participants were interrupted by IntelliGent and asked to pay attention to an article related to mobile devices). In the not related conditions participants were interrupted with the presentation of articles whose content was not related to the main task (e.g. answering questions about games, the participant was interrupted by IntelliGent and asked to pay attention to an article related to mobile devices).

After each main task was performed, participants were asked about the perceived quality of the main task in a questionnaire.

Camtasia studio software was used to record the time taken to complete each task.

3.2 Results

For each subject, TOT of interrupted and non interrupted tasks, TOI and results of the questions were analyzed. In this paper we present the analysis of the main effects.

TOT: Time on Main Task. We found significant differences in task performance F(3,57)=42.54, p<0.001. Table 1 shows the averages per condition. Comparisons between tasks show that the time
performance of the tasks abstract and recall is significantly higher than the time of the tasks recognition and writing title $F(1,61)=136.12$, $p<0.001$. Tasks with higher cognitive load require more time to be performed.

We did not find any significant differences in TOT between interrupted and not interrupted tasks $F(2,57)=0.46$, $p=0.62$. Our results suggest then that resuming the writing task after an interruption from IntelliGent does not seriously impair task performance.

Table 1. Average and standard deviation (between parentheses) of time performance of the main tasks according to the interruption tasks conditions

<table>
<thead>
<tr>
<th></th>
<th>abstract</th>
<th>Recall</th>
<th>recognition</th>
<th>Writing title</th>
</tr>
</thead>
<tbody>
<tr>
<td>No interruption</td>
<td>108.10(27.79)</td>
<td>118.88(48.78)</td>
<td>22.50(8.19)</td>
<td>24.00(8.10)</td>
</tr>
<tr>
<td>Related Interruption</td>
<td>99.00(38.19)</td>
<td>93.33(61.97)</td>
<td>24.25(14.56)</td>
<td>20.00(5.09)</td>
</tr>
<tr>
<td>Not related interruption</td>
<td>101.25(45.06)</td>
<td>103.50(60.10)</td>
<td>44.40(45.52)</td>
<td>24.00(9.84)</td>
</tr>
</tbody>
</table>

*Interruption task.* We also analyzed if there were differences on time performance of the interruption task (TOI) in the related and not related conditions (averages of 77.00 and 64.50 seconds respectively). The differences were not significant $F(1,30)=1.62$, $p=0.21$.

*Perceived Quality of the Performance of the Main Task.* Analysis of the questions showed significant differences as a function of the Main task $F(3,66)=5.10$, $p<0.005$ and of the interruption task $F(2,66)=5.69$, $p<0.005$.

Table 2. Average and standard deviation (between parentheses) of perceived quality of performance of the main tasks according to the interruption tasks conditions

<table>
<thead>
<tr>
<th></th>
<th>abstract</th>
<th>recall</th>
<th>recognition</th>
<th>Writing title</th>
</tr>
</thead>
<tbody>
<tr>
<td>No interruption</td>
<td>3.00(0.67)</td>
<td>3.40(0.96)</td>
<td>3.6(1.18)</td>
<td>4.50(0.76)</td>
</tr>
<tr>
<td>Related Interruption</td>
<td>4.2(0.96)</td>
<td>3.5(0.58)</td>
<td>4.2(0.83)</td>
<td>4.80(0.44)</td>
</tr>
<tr>
<td>Not related interruption</td>
<td>2.75(0.95)</td>
<td>3.5(1.00)</td>
<td>3.00(1.00)</td>
<td>3.60(0.91)</td>
</tr>
</tbody>
</table>

For the main tasks, pairwise comparisons showed that writing the title was significantly different than the rest (all $p<0.005$). As expected, participants were more satisfied with their performance in this easy task than with the others. No other comparisons were significant.

The differences between the conditions without interruption and the conditions with related interruptions was significant. Differences between related and not related interruptions were also significant (both $p<0.005$). These results indicate that there is not perceived difference between the conditions without interruption and with not related interruptions. In the related condition participants considered the quality of the main task was higher.

4 DISCUSSION AND CONCLUSIONS

Proactive Retrieval systems (PRs) such as IntelliGent retrieve and decide which available information is most likely relevant to the text being written, and present it without user intervention. In this paper we present the results of a preliminary study in which the effects of presenting peripheral information are explored.

The lack of significant differences in the main task performance between interrupted and not interrupted tasks suggests that resuming the writing task after an interruption from IntelliGent does not seriously impair time task performance. Furthermore, results from the questionnaire show that participants judged the quality of their performance better when a related paper was presented than in the conditions in which the paper was unrelated or IntelliGent was not present. We could conclude then that IntelliGent does not require extra time to resume the interrupted writing task and, when relevant information is presented, users are more satisfied with the quality of their performance.

This results seem to conflict with our previous findings [3] in natural working environments. Users felt that the disruptions of the system affected their writing performance and these disruptions were
negatively perceived as very frequently the system offered paper not relevant to the contents of the written text. This discrepancy in results can be explained in two related ways.

First, in our first study, users were university researchers and were asked to use the system while writing their scientific papers. This writing task seems to have more cognitive load than any of the tasks participants were confronted in this experiment and, consequently can be more sensitive to interruptions than the tasks we used in the present study.

Second, this study was designed to get an impression of how IntelliGent could affect several writing tasks. It would be interesting to analyze the interactions between the complexity of the tasks and types of interruptions in detail. However, doing this in a within-subjects design would require that all subjects perform such a large number of tasks that it becomes questionable whether they can complete the experiment without significantly changing their attitude and behaviour during the course of the experiment. In addition, it might be difficult to construct tasks with a sufficiently similar degree of difficulty for all conditions. Alternatively, one might opt for a between-subjects design, accepting the risk that medium size effects may fail to reach significance because of large differences between the cognitive capabilities of the subjects. Perhaps the best solution is to reduce the number of levels in the ‘task type’ factor to two. Since we found that writing tasks with higher cognitive load such as writing abstracts or answering questions, require more time to be performed than simpler tasks, we plan to fully factorial within subjects experiments. In one we will compare abstract and recall and in another we will compare recall and recognition. The first experiment would gauge the effect in two lengthy tasks, whereas the second would give insight into the differences between a long and complex tasks and a short and simple one.

ACKNOWLEDGMENT

We would like to thank the Search Expertise Centrum, Utrecht, The Netherlands for providing us with the ‘IntelliGent’ system and IOP SenterNovem, The Hague, The Netherlands, for funding this project.

REFERENCES