Contextual Sidebar: Effect of Context Information Quality on the Usability of a Context-aware Application

Sebastian Möller [1], *Hantang Liu* [1,2], *Axel Spriestersbach*[2]

[1] Quality and Usability Lab, TU Berlin, Germany
[2]SAP SE, Karlsruhe, Germany
sebastian.moeller@telekom.de, hantang.liu@sap.com, axel.spriestersbach@sap.com

Abstract

Context-aware applications promise better usability as they make context information which is relevant to the user accessible and can thus improve interaction flow. However, the quality of the presented information, as well as the way they are presented to the user, can be expected to have an impact on whether an improvement can be achieved. In this paper, we present a quantitative assessment on the improvement for a typical productive application in which context information is presented in an adaptive sidebar. Manipulating the context information quality in terms of accuracy and completeness, our experiment shows that both have a significant effect on usability: Whereas information accuracy significantly impacts execution time, completeness affects task result accuracy and mental effort. Overall, the user experience of the context-aware application was judged to be of high pragmatic and medium-to-high hedonic quality.

Index Terms: usability, context information, adaptation, context-aware systems

1. Introduction

Context-aware applications which are able to proactively respond to different usage contexts and environments become more and more visible. In such an application, context information has an impact on the way people interact with other people or with objects. The required context information constantly changes (with the context) and originates from distributed sources. This results in a tendency of context information to be incomplete and, to some degree, uncertain. Therefore, context information is usually imperfect and may not be worthy to use, which we call the problem of the quality of context information.

Depending on how precisely the context information fits to the purpose of why the application is used by a certain user, the interaction between user and system might be improved or worsened. In the end, this might lead to more satisfied users, higher user experience, and ultimately better acceptance of the application compared to a non-context-aware application. However, it is still unclear how the quality of the context information, as well as the degree of coverage of what is necessary for the task to be fulfilled, influence the interaction flow. Thus, in this initial study, we concentrate on two major aspects of the (functional) usability, namely effectiveness and efficiency, and only to a lesser extent on user experience. We will do that for a typical productive application where we see the potential of a positive impact of context awareness, namely an application used for making cost reimbursement claims for business trips inside a large company.

The paper will be structured as follows. First, we will have a closer look into the nature of the context information (Chapter 2). In Chapter 3, we will explain our solution for displaying such context information to the user with our typical application, using a so-called “contextual sidebar”. Chapter 4 presents an outline of our subjective experiment which was carried out to quantify the impact. The results are summarized in Chapter 5, and a summary and outlook for future work conclude the paper in Chapter 6.

2. Context information

Korkea-Aho [1] described context as: ”Context means situational information. Almost any information available at the time of an interaction can be seen as context information.” To our knowledge, Buchholz et al. have been the first ones to define the Quality of Context (QoC) as “any information that describes the quality of information that is used as context information” [2]. They concluded that the most important QoC-parameters are precision, probability of correctness, trust-worthiness, resolution, and up-to-dateness. In [3], further categorizations and quality modeling of context information are presented. Kim and Lee [4] proposed a method of measuring QoC with two quality dimensions: accuracy and completeness. They measured the raw context data obtained from the sensors, which is not directly applicable for us as the information we deal with is higher level. Manzoor et al. [5] proposed methods for quantifying QoC in terms of up-to-dateness, trust-worthiness, completeness, and significance at senor level and context management system level. They referred to QoC as the value of context information for a specific application.

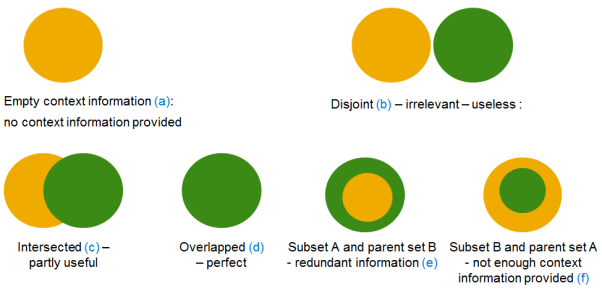

Figure 1: Relationship between available (green) and required (yellow) context information for a specific task and application.

10.21437/PQS.2016-8

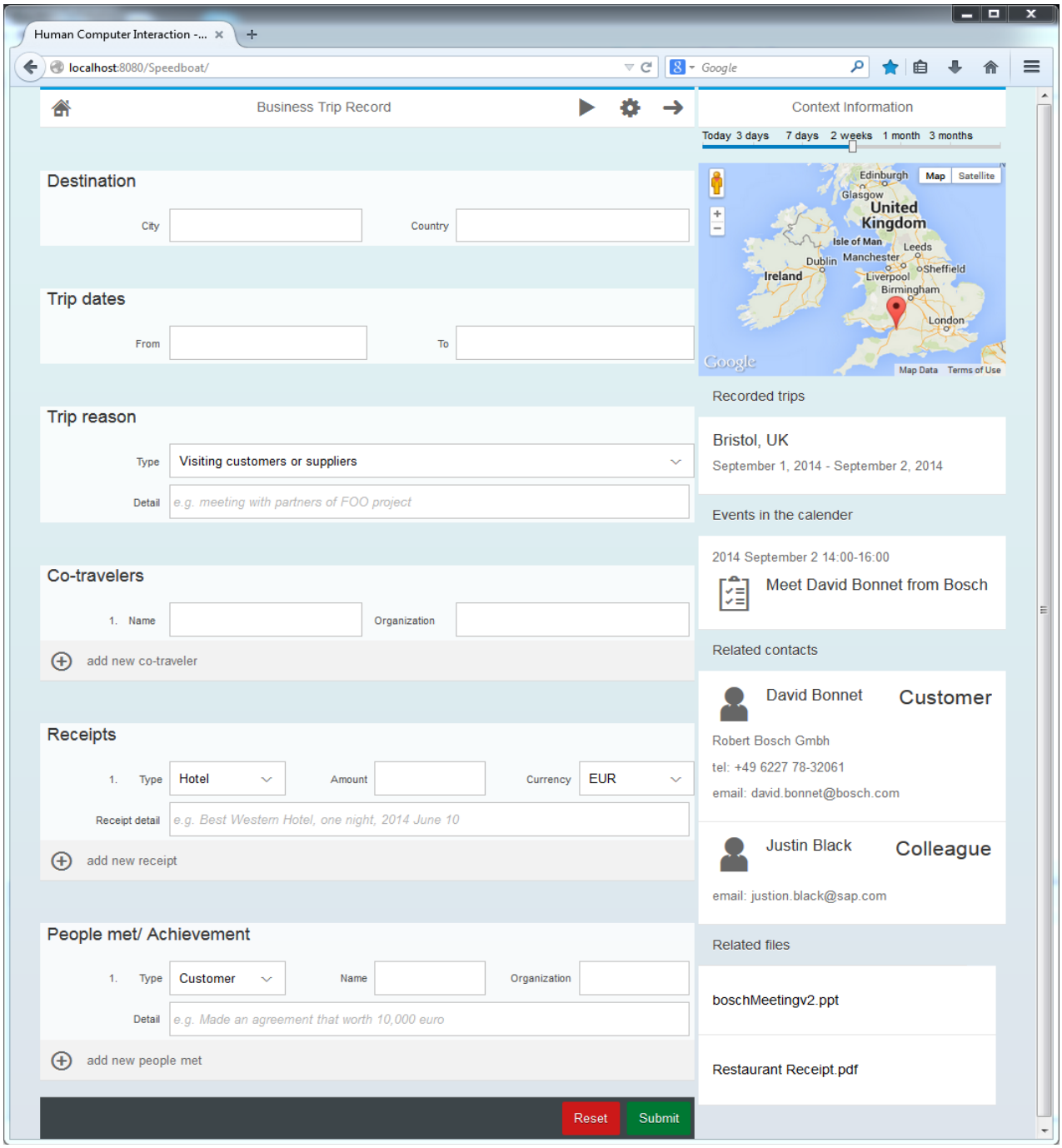

Figure 2: Business trip application with contextual sidebar (right).

For the purpose of our quantitative study, we describe context information in terms of available information offered by the application, and required information for resolving a task. The relationship between these two information sets can be as depicted in Fig. 1. It can be quantified by two parameters: The ratio of the overlapping area compared to the available context information area (yellow) which we call the Raw Information Quality (RIQ); and the ratio of the overlapping area compared to the required context information area (green) which we call the Cover Rate (CR). These two parameters decide the quality of context information. RIQ serves as a measure of distraction (or "accuracy"), and CR serves as a measure of the usefulness of the provided information (or "completeness"), both at the application level.

3. Application and contextual sidebar

The application we consider is a simple form-filling application that aims to help people record their business trips, see Fig. 2. It includes 6 areas "Destination", "Trip dates", "Trip reason", "Co-travelers", "Receipts", and "People met/Achievement" which have to be filled in during a prototypical task, and which can be used to determine the

accuracy of the exchanged information, and the metrics defined for context information quality above.

Context information is presented to the user via a sidebar, i.e. a graphical control element that displays various types of information on the side of an application or desktop user interface. We consider this to be a good way to provide supportive context information, because users are able to view the main application and the context information in a single view. Moreover, it is less intrusive as users can choose if they want to use the sidebar or if they want to close it. In our application, a sidebar was implemented for presenting four main parts of the context information "recorded trips", "events in the calendar", "related contacts" and "related files". The information of these parts can support the user to fill in the form. The application has some "rich interaction" features which include "click item to autofill", "search trips with timeline", "drag and drop", etc., which facilitate the information exchange between the sidebar and the core of the application. In the settings of the application, the rich interaction feature can be switched off for the experiment sessions which do not use any rich interaction.

4. Experimental design and set-up

The quantitative effect of QoC on usability was evaluated in a controlled laboratory study in which 6 different conditions of the application were compared. Condition 0 refers to using the application without sidebar, and conditions 1-5 to using the application with sidebar switched on and "rich interaction features" switched off. The latter conditions differ with respect to their QoC in terms of RIQ and CR (1: 30%/75%; 2: 50%/75%; 3: 100%/75%; 4: 100%/30%; 5: 100%/50%). Compared to Fig. 1, condition 0 corresponds to case (a), conditions 1 and 2 correspond to case (e), and condition 3, 4 and 5 correspond to case (f). The comparison of conditions 1, 2 and 3 helps to analyze the effect of RIQ, whereas the comparison of conditions 3, 4 and 5 shows the effect of CR.

The experimental conditions were distributed to the test participants in a mixed within-between-design, in which each participant had to resolve pre-defined travel reimbursement tasks with three out of the six application conditions. A full within-participant design was not chosen, as this might have caused fatigue to the participants. 24 participants (17 m, 7 f) were recruited, 16 of them employees of a software company and 8 either Master or Doctoral students in computer science or engineering at TU Berlin. 19 subjects were 20 to 40 years old, 5 were 40 to 60 years old. The experiment took about 45 minutes and participants were not paid for their participation.

Participants filled in a pre-experiment questionnaire and were introduced to the application, the sidebar, and the tasks to perform. In each of the three sessions, participants were asked to read a short story about one business trip and afterwards use the application to fill the form based on the story they had read. The stories were designed to simulate the memory of three business trips. Each story consisted of 4 paragraphs and had about 350 words. Detail information of the business trips were nested in the story, and distracting information was added. Each story had a similar structure and contained a similar amount of information. For each story, there was a corresponding form where basic information of that story (name of the character, name of the employer, date, etc.) was stated. The correct results for each story form were used in order to calculate the result accuracy achieved by the participants.

Participants were asked to fill the form as fast as possible and to fill as many fields as possible with the help of the provided information sheet, the contextual sidebar, and the supported file folder. The form filling process lasted roughly 3-10 minutes for each task, and the interactions were logged. Participants were then asked to fill the Rating Scale Mental Effort [6] indicating different levels of mental effort. After all three sessions, the subjects were asked to rate different aspects of user experience using the AttrakDiff [7] questionnaire. In the end, a short interview was conducted.

5. Results

In the following paragraphs, we will analyze the experimental results in terms of accuracy, mental effort and execution time. The average results are displayed in Fig. 3.

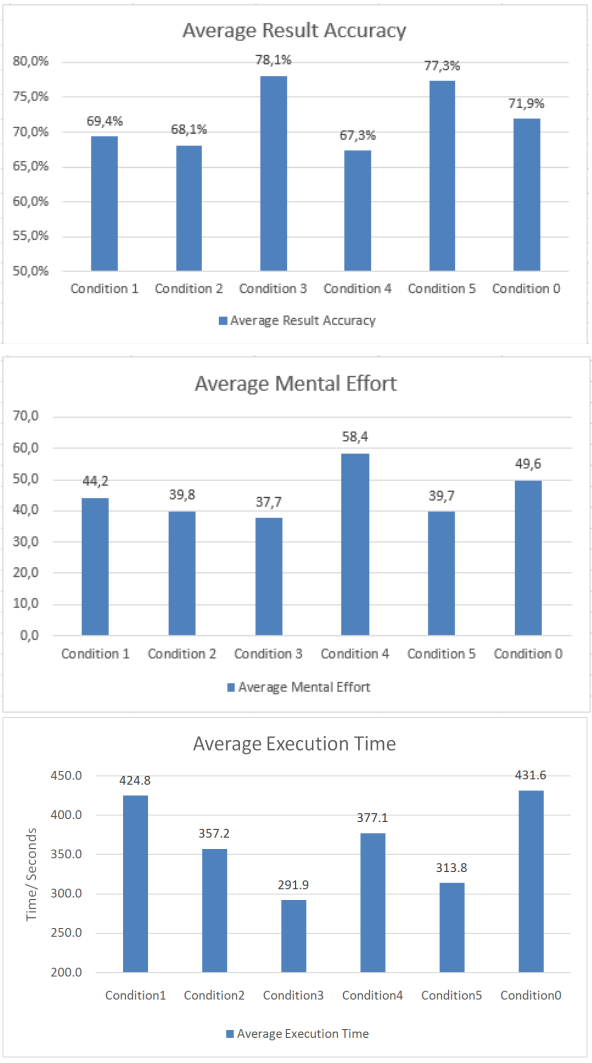

Figure 3: Accuracy, mental effort and execution time per condition

The average result accuracy is shown in Fig. 3, top panel. It can be seen that conditions 3 and condition 5 have higher accuracy than other conditions. However, applying a single factor ANOVA for all conditions, the statistical analysis results showed that there was no significance of the effect of

conditions on result accuracy, p>0.05. Applying a single factor ANOVA only for conditions 3, 4 and 5 (RIQ: 100%, CR: [75%, 30%, 50%]) and using CR as the independent variable and result accuracy as dependent variable, the statistical analysis results showed that there was significance of the effect of CR on result accuracy, F = 3.74, p <0.05. Follow-up paired comparisons tests (Tukey) indicated that condition 3 demonstrated significantly higher result accuracy than condition 4, p <0.05. There were no significant differences of the effect of CR on result accuracy between condition 3 vs. condition 5 and condition 4 vs. condition 5, both p>0.05.

The average mental effort measured by RSME is shown in Fig. 3, middle panel. It can be seen that condition 4 required a higher mental effort than the other conditions. The statistical analysis results showed that there was significance of the effect of conditions on mental effort, F = 3.34, p<0.01. Follow-up Tukey's post hoc test indicated that condition 4 demonstrated significantly higher mental effort than condition 2, condition 3 and condition 5, p<0.05. Between all other pairs of conditions, there were no significant differences on mental effort, all p>0.05. Applying a single factor ANOVA only to conditions 3, 4 and 5 (RIQ: 100%, CR: [75%, 30%, 50%]), the statistical analysis results showed that there was significance of the effect of CR on result accuracy, F = 7.45, p<0.01. Follow-up Tukey's post hoc test indicated that condition 4 demonstrated significantly higher mental effort than conditions 3 and 5, p<0.01. There was no significant difference of the effect of CR on result accuracy between condition 3 and 5, p >0.05.

The average task execution time is shown in Fig. 3, lower panel. It can be seen that condition 3 has less execution time than the other conditions. The statistical analysis results showed that there was significance of the effect of conditions on execution time, F = 3.25, p<0.05. Follow-up Dunnet's post hoc tests indicated that condition 0 demonstrated significantly more execution time than conditions 3 and 5, both p<0.05. Tukey's test indicated that condition 3 demonstrated significantly less execution time than condition 1, p<0.05. Between all other pairs of conditions, there were no significant differences on execution time, all p>0.05. Applying a single factor ANOVA only to conditions 1, 2 and 3 (RIQ:[30%, 50%, 100%], CR: 75%), the statistical analysis results showed that there was significance of the effect of RIQ on execution time, F = 3.74, p<0.05. Follow-up paired comparison tests (Tukey) indicated that condition 3 demonstrated significantly less execution time than condition 1, p<0 .05. There were no significant differences of the effect of RIQ on execution time between condition 1 vs. condition 2, and condition 2 vs. condition 3, both p>0.05.

Although the user experience of the test application was not our focus, a minimal experience should be guaranteed to not affect the results of the study. Thus, we used the AttrakDiff questionnaire to test this effect, but only for the entire application, and not on a per-condition basis. The results in Fig. 4. show that the value of pragmatic quality (PQ) which refers to the usefulness and usability of the application reaches above-average values, whereas the hedonic quality-stimulation and hedonic quality-identity are just at the corner to above-average ratings. The attractiveness (ATT) value is in the above-average region, thus the overall impression of the application seems to be quite attractive.

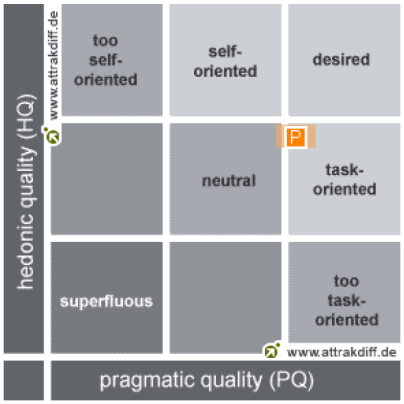

Figure 4: AttrakDiff results of the application.

6. Conclusions and future work

Overall, the six conditions were found to have no significant effect on result accuracy, but significantly affected mental effort and execution time. Sidebar usage had no significant effect on result accuracy and mental effort, but had a significant effect on execution time. The usage of a sidebar with fairly good QoC information improved the overall performance significantly compared to not using the sidebar.

In terms of RIQ, there was no significance of the effect of RIQ on result accuracy and mental effort, but significance for execution time. When CR is fairly good (75 %), there might be a threshold of RIQ that decides if it is worth to use the contextual sidebar or not. The threshold might lie between RIQ of condition 3 and condition 2, that is between 50% and 100%. CR was found to have a significant effect on result accuracy and mental effort, but no significant effect on execution time. When RIQ is perfect (100 %), there might be a threshold of CR that decides if it is worth to use the contextual sidebar or not. The threshold might lie between CR of condition 4 and condition 5, i.e. between 30% and 50%.

To study the effect of RIQ and CR, we applied single factor ANOVAs for parts of the conditions. If we want to analyze the interaction effect between RIQ and CR, a mixed ANOVA should be used. However, a mixed ANOVA could not be applied here, since there were no mixed conditions for different groups of participants, so no repeated measures could be applied. A repeated measures design would enable such an analysis.

From the observation and the log files, we learned that participants tended to ignore the “related files” part in the sidebar. Knowing that the related files were not always correct or relevant, test participants tended to search the files folder directly. In the post-experiment interviews, participants reported that the reason for this phenomenon was that unlike in other parts of the sidebar, they could not see the information directly in “related files” part of the sidebar, but needed to click the link to open a new page. Therefore, when designing similar contextual sidebar applications in the future, it would be very helpful if the outline or key information of related files or related links could be shown in the sidebar. In our study, we switched the “rich interaction” feature of the application off, because we assumed that it would add experimental bias. It would be interesting to check whether the results still hold when such features increase the usefulness of the contextual sidebar.

7. References

[1] K. Henricksen, J. Indulska, A. Rakotonirainy. "Modeling context information in pervasive computing systems," in: Pervasive Computing, pages 167–180. Springer, 2002.

[2] T. Buchholz, A. Küpper, M. Schiffers, "Quality of context: What it is and why we need it," in: *Proc. Workshop of the HP Open View University Association*, 2003.

[3] M.A. Razzaque, S. Dobson, P. Nixon, "Categorization and modelling of quality in context information,", in: *Workshop on AI and Autonomic Communications, held at International Joint Conference on Artificial Intelligence (IJCAI)*, page EJ, 2005.

[4] Y. Kim, K. Lee, "A quality measurement method of context information in ubiquitous environments," in: *Int. Conf. on Hybrid Information Technology (ICHIT'06),* vol. 2, pp. 576-581, IEEE, 2006.

[5] A. Manzoor, H.-L. Truong, S. Dustdar, "On the evaluation of quality of context," in: *Smart Sensing and Context*, pp. 140-153, Springer, 2008.

[6] S. Miller, *Workload Measures*, University of Iowa, 2001.

[7] M. Hassenzahl, M. Burmester, F. Koller, „AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität," in: *Mensch & Computer 2003,*Vieweg+Teubner Verlag, 2003, pp. 187-196.