CS6750 Assignment M4:

Outlooks for Outlook

Michael Tong
mtong31@gatech.edu

Abstract. This analysis evaluates the proposed prototypes from M3 with qualitative, quantitative, or predictive methods. The folder specific email creation is evaluated through a survey, the new meeting response buttons is evaluated with an administered experiment to analyze response times, and a consistent layout is evaluated with a cognitive walkthrough.

Introduction

Moving forward from assignment M3 (Tong 2018a) for the interface redesign of Outlook, the study now transitions into evaluation of the prototypes conceptualized. The two prototypes, new email creation from a folder, the new meeting response buttons are discussed in preparation for a potential evaluation, while the consistent interface across multiple platforms is evaluated with a cognitive walkthrough. The two prepared evaluations will discuss the administering of a survey and experiment to measure the qualitative and quantitative responses of the respective prototype.

Qualitative Evaluation

From assignment M3 the project had identified and described the potential benefit of implementing an additional "new email" option from a specific folder. The rationale for this feature is to allow response emails and conversations to automatically be organized into a folder if the user chooses to do so. In order to

evaluate the effectiveness of this prototype, a survey provides a low-cost, non-intrusive, and analyzable approach.

Since the inspiration for this prototype was conceptualized as a result of the survey from assignment M2 (Tong 2018b), the same general population can be surveyed again to determine if they find the new feature to be productive and beneficial. In addition to a majority (>90%) of users stating that Outlook is their primary email for work, in the free comments, many users specifically stated that they would like better categorization or organization of their emails. As such, the target population for this survey will once again be the student body of Georgia Tech, and data will be recorded and stored through http://peersurvey.cc.gatech.edu/.

From the survey results, the null hypothesis would be that people do not find the new feature helpful and that there is still a desire for improvements to the folder/categorization system in Outlook. The alternate hypothesis would be that people find this feature helpful and that there is a reduction in comments regarding improvements to the folder/categorization system.

Below are the questions that will aim to understand if users find this implementation to be beneficial or a hassle, determine if the request for better categorization or organization options is reduced, and if users would be satisfied with this addition.

1) What do you primarily use Outlook for?
    a. Work.
    b. Academia.
    c. Personal.
    d. Other.
2) Do you use folders to organize your emails?
    a. No.
    b. Yes.

3) If not, explain why.
    a. (Free response)
4) Do you categorize everything or only specific emails?
    a. Everything.
    b. Specific emails.
    c. None
5) Are you content with the folder system?
    a. Yes.
    b. No.
6) If not, explain why.
    a. (Free response)
7) You can now send a message from a folder and all responses will return to that folder, do you think this would be helpful?
    a. Yes.
    b. No.
8) If not, explain why.
    a. (Free response)
9) Would you like to see any specific improvements to the folder system?
    a. (Free response)
10) Additional feedback:
    a. (Free response)

Use of this evaluation addresses the requirements of the data inventory section from assignment M2 through the discussion of continued needfinding for a particular topic. In the requirements definition section, it was discussed how surprising it is that over 90% of the respondents stated that they use Outlook primarily for work, which led to the project shift towards work place productivity. This evaluation addresses these requirements by integrating a potential improvement to the efficacy of the Outlook interface, specifically for a work place environment.

Empirical Evaluation

For the prototype of implementing an additional "hidden button" on emails for meetings and events as discussed in M3, this section will discuss a potential evaluation of the feature. Figures 1 through 3 display the considered feature.

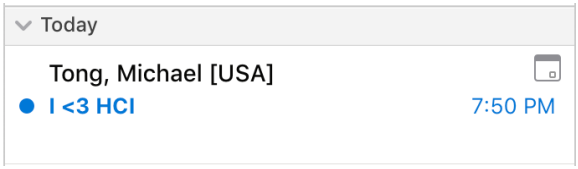

Figure 1: New meeting email, not highlighted

Figure 2: New meeting email, highlighted (current interface)

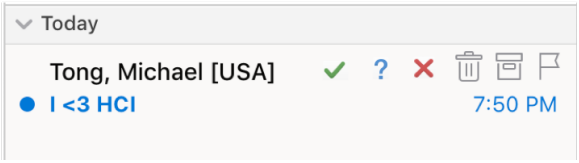

Figure 3: New meeting email, highlighted (proposed interface)

From this evaluation, the goal of the study is to identify if the additional feature improves how quickly the user can accept a meeting or event. The comparison in this situation will be the user's traditional method of accepting a meeting, whether that be clicking the email and clicking accept, or going to their calendar and pressing accept. The comparison metric will be the time it takes for the user to respond to the meeting or event. Since this feature is somewhat difficult to discover initially, a

within-subjects study will have to be administered with the user aware of the new feature. This would then allow the analysis to consider the response time difference per user which can then be used to compare results across study participants. This leads to the null and alternate hypothesis, where the null hypothesis would be that this feature does not show a significant improvement in the time taken to respond to a meeting across a sample population, while the alternative hypothesis would be that the results do show a significant improvement in the response delta.

While a Goldilocks empirical evaluation would be to A/B test the feature through software logging, this project unfortunately doesn't have the support of Microsoft or an experienced hacker. Administering this evaluation without recreating the interface is going to be difficult, but there is a potential solution with some significant assumptions. Since the existing interface is being used, participants in the experimental group will simulate their response with the new buttons by using the existing ones (delete, archive, flag). The experiment will be that the participant is briefed on the changes with visuals on what to expect and has Outlook closed. A meeting will then be sent to the participant randomly, which they have to then select the response that will be titled in the email. Additionally, non-meeting emails will be sent to prevent premature movements. To integrate the new interface into this budget evaluation, the delete, archive, and flag buttons will represent accept, tentative, and decline respectively, which the user will be aware of and have time to operate within the current system.

The second treatment group will be the control group, where participants have to accept the meeting without the new feature. After the study, the sent and returned emails will be reviewed (email source data), and the time deltas will be calculated from when the message first arrives in the inbox to when the response is selected. The analysis of this evaluation will be the time differences between the two methods of meeting responses.

This approach clearly has a significant lurking variable in that the icons do not match the intended response, however, the assumption is that with dedicated buttons

and symbols, the response should be faster, and if this evaluation is able to reject the null hypothesis, then the actual improvements may be more significant than what is discovered. Another lurking variable is that users are aware of the situation and are expecting the meeting request, which may or may not simulate reality accurately. An additional lurking variable is that the email data only tracks sent and received timestamps to the nearest second, which may not be discrete enough to capture the beneficial differences, but this may then result in acceptance of the null hypothesis.

Prediction Evaluation

Assignment M3 discussed the creation of a consistent interface across all platforms and this analysis will now perform a predictive evaluation of this prototype. Since this prototype is fairly inaccessible to redesign and test, a prediction evaluation method is ideal to gather feedback. To simulate the situation, the task addressed will be a user attempting to access the two most commonly used features in Outlook as gathered by assignment M2, namely the email and calendar, and across multiple platforms. These platforms will be the desktop application (MacOS & Ubuntu) and the web (Chrome). Since the mobile application introduces completely different variables, this study will initially focus on only desktop usage. For each of these scenarios the analysis will assume that the user has the application downloaded as well as having an Outlook account. The tasks will then be to view, respond, and create an email, and also to view and create a meeting across both platforms. The operators available to them during this analysis will be moving the mouse, clicking, typing, and viewing. For this study, the operator will be aware of their task ahead of time, as the purpose of this evaluation is to determine if the interfaces differ enough to warrant modifications.

To begin with the desktop application, the user by default is in the email and inbox section, with the folders and inboxes on the left side, list of emails in the middle, and the selected email on the right. Viewing an email is done by clicking on it and responding requires the user to click the “reply” or “reply all” button on the top bar.

Once clicked, a new window pops up which allows the user to compile and format their message, then they are able to click the send button on the top left. Creation of an email requires that the user select the "New Email" icon in the top left of the screen, which opens a new window similar to the response window box that pops up.

With the web application, the user has a similar initial layout, and emails can be selected by clicking on them. Responding however, requires the user to click the "reply all" button within the email, or the drop down arrow next to it to access the "reply" button. Once a selection is made, the original email is replaced with the same editor window that appears with the desktop application, but not as a new window. Creation of a new email requires the user click the "New" button just above the list of emails. Once done the right most column is replaced with a box similar to the response email box but once again within the same screen.

With meetings, the user can view them in the desktop application by clicking the calendar tab on the bottom left of the screen. A screen then pops up with a standard calendar view. For creation of a meeting, the desktop application allows the user to do so by clicking the "Meeting" button on the top of the screen within the email tab, or in the calendar tab. Once selected, a new window pops up which allows the user to populate it with the respective information.

For the web application, the user can view meetings by selecting the calendar tab in the bottom left of the screen, which brings up a similar screen as the desktop application. Creation of a meeting however is done by either selecting the dropdown menu next to "New" in the email tab and then clicking "Calendar event", or clicking "New" in the calendar tab. Both these options open an overlay which the user can populate with event details.

From this evaluation, the study finds that the two systems are fairly similar in function but do have inconsistencies. A big difference is the web application's use of "Events" versus the desktop application having both "Events" and "Meetings". This may cause some confusion as users are not sure if they represent the same goal. The next significant difference is the response interface. In the web application there is no

"reply" or "reply all" button on the toolbar where the desktop application does, and to make matters worse, the desktop application also doesn't have the response buttons displayed on the web application.

Preparing to Execute

From the three evaluations discussed above two will be considered for the next assignment. The prototype of standardizing the layout of Outlook across all platforms is not very feasible to be accomplished by a single person outside of the company, and thus will be the one removed.

The survey is selected due to it meeting the demands of the data inventorying criteria discussed in assignment M2. It presents a low-cost, easily analyzable approach, with high returns in information, and targets the population that conceptualized the prototype to begin with. Additionally, the familiarity and confidence in this approach justifies its candidacy.

Administering the new response button evaluation experiment will also be carried out since it provides a more feasible experiment that the layout standardization. This experiment will also provide quantitative results which will be interesting to analyze and influences the original data inventorying criteria from assignment M2 through enhancement of Outlook's productivity in a working environment.

Overall, both these evaluations can feasibly be performed by a single person over a couple of weeks and both address the requirements defined earlier on in the analysis. They both also provide an interesting opportunity to immediately evaluate the improvements established by the prototypes.

References

1. Tong, M. (2018a). Assignment M3. *OMS CS6750 Human- Computer Interaction*. Washington, DC.
2. Tong, M. (2018b). Assignment M2. *OMS CS6750 Human- Computer Interaction*. Washington, DC.