Table of Contents

Section 1

Paper for CHI 2010 Student Research Competition

Section 2

Phase III Report

Redesign, Evaluation and Result Analysis

Redesign Rationale

Our low-fidelity system prototype was generally well-received by the participants whom took part in the task-centered walkthroughs we performed in Phase II, and they expressed excitement about the idea of being able to group, annotate and share tabs based on tasks. However, from the point of view of the evaluators and interviewers, we were concerned about the complexity of the tasks that were presented to the participants; given that there are various paths a user can take to achieve the same tasks in a web browser (i.e. there are multiple ways of accessing the bookmarks, such as through the popup sidebar or the persistent bookmark toolbar above the tabs), it was difficult for the task walkthroughs to cover all cases. Such limitations might have prevented our participants to properly and critically evaluate the more detailed interactions associated with our design. Despite these possible deficiencies, we carry through our Phase III implementation by keeping most of the features that we originally came up with in order to more thoroughly evaluate the task-focus browser interface. Namely, these features are grouping, suspension and resumption, annotation and sharing of tasks. Related to these features, we identified two key points that needed be improve upon, which includes the default hiding of task management features and transparent sharing. In the following sub-sections, we detailed the improvement approaches we have taken for the medium fidelity prototype.

Default Hiding of Task Management

While the task-focus approach allows the user to easier organize tabs and is more suitable for complex task switching and management, it might not always be necessary when the users are performing their daily web browsing routines; we intend our interface to not hinder or interfere with the users' regular browser activities.

The initial low-fidelity system prototype relies on a task sidebar to manage all the tasks associated with the users, which can be either active or inactive, as shown in the screenshot of our paper prototype in Figure 1. Participants' feedback suggested that the sidebar should be hid away when task management is not in used. While due to time constraint, such shrink-away feature was not implemented in our medium fidelity prototype, we introduce a default task to our design such that the user can perform all the browsing practices they are accustomed to. Since tasks and tabs need to be actively managed by the users, the default task provides a pathway for the user to return to existing browsing mechanisms without utilizing any of the task management features of our design.

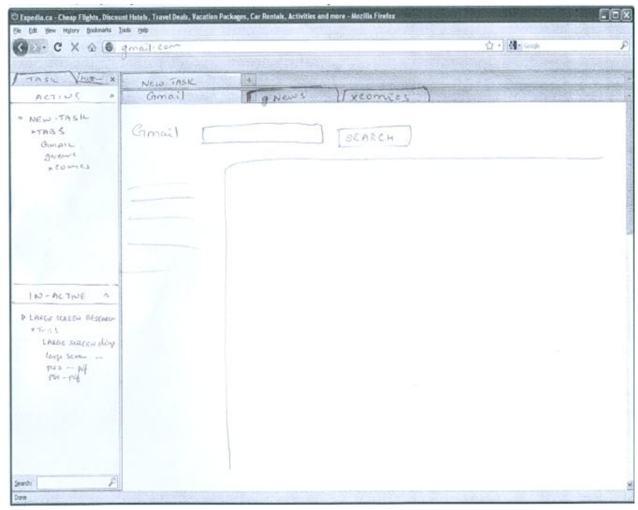

Figure 1. TabFour paper prototype with task sidebar

The default task is essentially the same as any other organized task; as the users feel the need to start organizing the numerous tabs that are being opened in the default task, they can easily rename and annotate, or even create new tasks to better handle the increasing browsing complexity. Figure 2 shows a screenshot of our browser prototype that initializes to the default task.

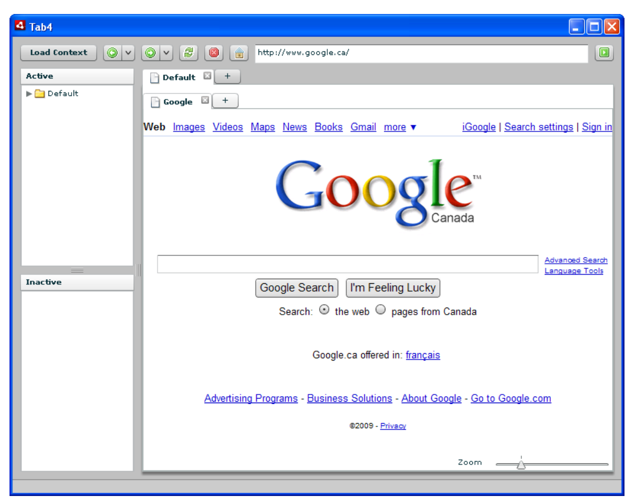

Figure 2. TabFour Browser showing the default task

Transparent Sharing

In our initial design, we intend our system to support sharing of an entire context by saving it as a "context file", which can by transfer through emails or other mean to the receiving party. The context file is essentially a XML data file that contains all the information associated with the task, including opened links, annotations and bookmarks, so the user would be able to share more effectively multiple artifacts associated with a task with other parties. Our participants expressed concerns with the sharing mechanism in several ways. First of all, even though sharing an entire context file simplifies the need to transfer multiple links, it still heavily relies on the emailing or whichever file transfer protocols. Second, what if there are errors or corruptions to the file, or even viruses? Third, these files cannot be easily updated. If two parties were to collaborate on a browsing task, i.e. a research project that requires sharing the same task back and forth multiple times, multiple versions of the context file have to be created and transferred, which can be a daunting and non-trivial procedure.

A large part of our interface redesign surrounds the sharing feature in order to address these concerns. We designed two new mechanisms that allow the user to better share the task information which are drag-and-drop and cloud sharing.

Compared to the concept of a context file, the drag-and-drop design allows the user to more easily transport all the artifacts associated with a task to other applications that support html-format or plain text, such as an email client or a text file. The mechanism simplifies the process of the storage and sharing of browsing information that not only includes the tabs, but also bookmarks and annotations that have accumulated throughout the lifetime of the task. Figure 3 illustrates the drag-and-drop operation from our browser prototype to a web email client.

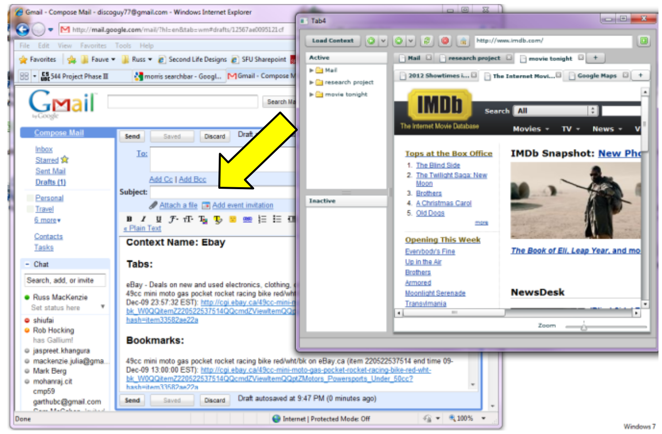

Figure 3. Drag-and-drop sharing

Another interesting sharing mechanism that we hope to explore and evaluate more in-depth in the future is the concept of cloud sharing of tasks. Besides dragging and dropping, we implemented the cloud sharing feature in the prototype that allows users to share the same task information through a simple web-service, referenced by a filename. The same task, along with all the tabs, bookmarks and annotations, can be obtained by other users by referencing to the same file name. The transfer is lightweight, as only the file name needs to be communicated, and updating the task to be shared again is simple, as the same file can be updated through the web. Not only does it promotes easy sharing, eliminating the back-and-forth transfer problem aforementioned, the cloud sharing feature of our prototype should encourage interesting browsing behaviours, such as randomly accessing task names that are of the user's interest to explore what web pages, annotations and bookmarks other users have made in the same task space.

Evaluation

Participants

Convenient samples of 8 participants were selected from amongst friends of the team members, of which 3 are females and 5 are males. All the participants are UBC students and are computer users with average level of computer knowledge. The first participant was intended to be a pilot subject, but since no significant changes were made in the study, the responses from the pilot study are included here. Four participants (P2, P3, P4 &

P5) did the experiments in an observation room specially designed for human subject experiments (discussed below), and all other participants completed the experiments in general settings with one of the team members.

Apparatus

The experiments were conducted by using two connected personal computers. The software system of one of them was specially configured as the working environment for the participant; we call this computer the participant's computer. The other computer was controlled by an experimenter to give simulated tasks and interruptions to the participants; we call this the experimenter's computer.

The Participant's Computer

The participant's computer was running Windows 7 Professional as its operating system. To simulate the working environment of a regular computer user, we created a new clean user account on the computer. Preinstalled software includes Mozilla Thunderbird for receiving email messages, Windows Live Messenger for receiving instant messages, Mozilla Firefox and our TabFour prototype as the web browsers. In addition, we also installed screen capture software Fraps to record real screen motion of the participant for further analysis.

The Firefox web browser was ensured to have no bookmarks or plug-ins installed, which may influence the accuracy of our experiment. The contexts created in the TabFour browser were cleaned and the email messages in Thunderbird were also cleaned after each experiment. Fraps was turned on once the computer was handed over to the participant. We realized that Fraps may cause performance issues to the computer but this will not influence the results of the experiment. See detailed analysis of this in the pilot study section below.

The Experimenter's Computer

The experimenter's computers vary from the team member taking the experiment. While there is no special configuration regarding the experimenter's computers, at least two pieces of software are installed. An email client was installed to send email messages to the participant's computer and Windows Live Messenger was installed to send instant messages to the participant.

Due to the nature of the tasks, the experimenter pretended to be the participant's friend whom is of the same gender, acting as either Mark or Jane. Separate email accounts were created specifically for this purpose, and the experimenter would send email messages and instant messages with the name of the virtual friend to the participant.

The Observation Room

We used the observation room on the 7th floor of the ICICS/CS building. The observation room is equipped with one-directional glass windows, allowing us to observe the actions of the subjects non-intrusively. The observation room has two video cameras fixed in two corners of the room. The experimenters outside the observation room can control the orientation of the camera, zoom in or zoom out remotely. The experimenters can also view the images of the two cameras on the same screen capture window. For the four participants completed their experiment in the observation room, we recorded the video stream of the complete experiment to aid further analysis.

Procedure

The experiment consists of using the two different browsers (Firefox, TabFour) to perform two different trip planning tasks (Orlando, London) that require the participants to search through multiple airline and tourist spot websites to locate cheapest fares and admission fees. One of the trips focused on planning a visit to London during spring break while the other was on a trip to Orlando during Christmas. We counterbalanced the order of the web browsers and the corresponding trips throughout our 8 participants, and we also ensured that different airline websites were used for each trip to avoid any possible learning effects. For each individual participant, the experiment was arranged as follows.

TabFour Walkthrough and Experiment Introduction (5 minutes)

The participants first were given a brief walkthrough of the TabFour browser. One experimenter demonstrated the most common features of the browser but using the features or not was completely up to the participant. The experimenter would briefly introduce the procedure of the experiment, but the details of the tasks the participant was supposed to complete were briefly glanced over.

Task Sets (15 minutes per task)

In a task set using either of the two browsers, the participants first were given an email from the virtual friend, explaining that they are planning a trip and expecting the participants to look up the lowest airfares in a given time window. This is a relatively complex task involving opening multiple airline websites and comparing prices of different dates that is intended to keep the participant occupied throughout the time window. At about the 5th minute from the beginning of the airfare task, the experimenter sent out an instant message through the virtual friend to the participant, asking the participant to lookup the author names of some papers. The experimenter also stated the task was urgent and should be done immediately, to force the participant to interrupt current task and switch to a completely irrelevant context. Right after the participant sent back the results of the second task, the experimenter will then sent another IM message, requesting the participant to look up the prices for some places of interest of their trip destination. The participants were requested to send back the results of both the airfare task and the ticket task in one email at the end. The same procedure was repeated in the second task with the alternate browser and trip destination.

Below are some sample messages sent to the participants.

Message sent to the participants requesting them to lookup airfares in Firefox.

Hey Participant,

Can you look into air fare for the trip we are planning during Christmas to Orlando, Florida? We can leave on any days between December 21st-24th for 7 days. I heard US Airways (http://www.usairways.com) and Alaska Airlines (http://www.alaskaair.com) have some good deals. Can you check the prices on their websites? Don't worry about hotels as we can stay at my aunt's place. Can you get back to me in 15 minutes with the cheapest price (along with the dates) from each of the two airlines? Meanwhile, I will look into places we can go to.

- Mark

Message sent to the participants requesting them to lookup airfares in TabFour. See the underlined part for difference with the previous message.

Hey Participant,

Can you look into air fare for the trip we are planning during Christmas to Orlando, Florida? We can leave on any days between December 21st-24th for 7 days. I heard US Airways and Alaska Airlines have some good deals. Can you check the prices on their websites? <u>For convenience I have started an online task - file name is orlandotickets</u> - Don't worry about hotels as we can stay at my aunt's place. Can you get back to me in 15 minutes with the cheapest price (along with the dates) from each of the two airlines? Meanwhile, I will look into places we can go to.

- Jane

Figure 4. Message sent to the participants requesting them to lookup author names in TabFour.

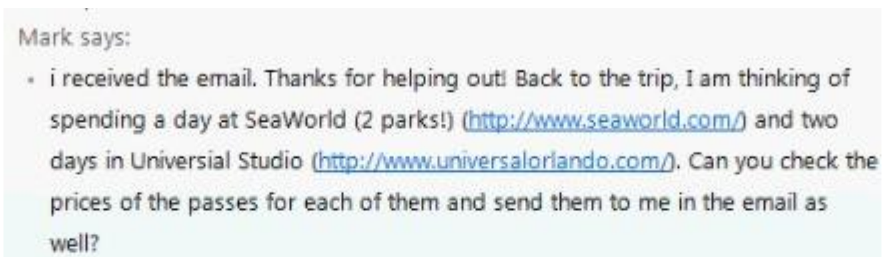

Figure 5. Message sent to the participants requesting them to lookup ticket prices in Firefox.

Evaluation Surveys (5 minutes each)

At the end of each task, the participants were asked to fill out an evaluation survey about how they feel when completing the tasks in the web browser just used. We are mostly interested in the workload required by the design of the primary trip planning tasks and the interruptions, and how well the browser they used supported what they were trying to achieve. After the completion of the two tasks, the participant was also asked to fill up a questionnaire that compares the two browsers. The details of these survey and their results have been discussed in the extended abstract write-up and the results section of the appendix.

Pilot Studies

We conducted two pilot studies. The first one is an informal experiment done by one of the experimenters, which we did not include this in the final result analysis. The second pilot study was done on a real participant in a formal setting and was included in the final results analysis, but the experiment was not carried out in the observation room.

First Pilot Study

Originally, we plan to monitor the screen of the subject through remote desktop software. By doing this, we believe only posed very little influence to the subject's computer. We used TightVNC which is a free remote desktop monitor application. Then we planned to capture the screen of the subjects at the experimenter's computer. However from the first pilot study, we discovered using remote desktop monitoring method has two major defects. First, the network bandwidth was mainly consumed by the remote desktop application and made web browsing noticeably slow. Second, remote desktop monitoring had a very low refresh rate. The quality of the video is not acceptable. Therefore we decided to record the screen on the participant's computer and we found that by doing so, the performance impact was acceptable.

Another change we made from the first pilot study was the order of the interruptions. At first, we planned to give the participants the travel planning first, then the ticket task followed by the paper task at last. However, we found the paper task's interruption effect was not strong enough, as the user was almost finished with the first two tasks by the time the paper task came at the 10^{th} minute mark. Although the paper task was marked "urgent", the users were likely to do it after completed the first two tasks, rendering the interruption meaningless. Therefore we exchanged the paper task with the ticket task and found the two interruptions worked well because although when the ticket task came, it was also approaching the end of the airfare search. As they were more closely related, the users were likely to do them simultaneously.

Another problem we discovered was some websites could not be rendered correctly in TabFour, because some websites were not compatible with WebKit, which is the render engine used in our browser. We eliminated these websites from the tasks and replaced them with alternative websites.

Second Pilot Study

The second pilot study was done on our first participant, under the revised experiment settings and procedures after the first pilot study. After the second pilot study, we made another minor correction to our experiment. Originally for the TabFour browser, only one task of the three tasks used the "online context sharing" feature. For the other two tasks, the user still had to manually copy and paste the link to the browser. Even worse, when the user was selecting the links, sometimes he clicked on a link accidentally, and the operating system opened the default browser (Firefox) for the link. To fix this and to better evaluate user reactions of the sharing context feature. We decided to change all three tasks to use shared contexts by sending the participants only the file name of the contexts to be loaded, hiding all explicit links.

Results

We first present the results of the evaluation questionnaires, followed by the comparison questionnaire. The evaluation questionnaire was administered immediately after each 15-minute session and is modeled on the NASA-TLX workload scale. Subjects rank five statements on a 7-point Likert scale, with 1 as "Very Low" and 7 as "Very High".

The questions were:

1. How mentally demanding was the task?
2. How hurried or rushed was the pace of the task?
3. How successful do you think you were in accomplishing what you were asked to do?
4. How effectively did the browser support what you were asked to do?
5. How insecure, discouraged, irritated, stressed, and annoyed were you?

To analyze these responses we used the Wilcoxon Matched-Pairs Signed-Ranks, a non-parametric test, because the data is ordinal. The matched-pairs signed-ranks are shown in Table 1.

		N	Mean Rank	Sum of Ranks
Q1T4 - Q1FF	Negative Ranks	1a	1.50	1.50
	Positive Ranks	2b	2.25	4.50
	Ties	5c		
	Total	8		
Q2T4 - Q2FF	Negative Ranks	1d	3.50	3.50
	Positive Ranks	4e	2.88	11.50
	Ties	3f		
	Total	8		
Q3T4 - Q3FF	Negative Ranks	2g	3.00	6.00

	Positive Ranks	2h	2.00	4.00
	Ties	4i		
	Total	8		
Q4T4 - Q4FF	Negative Ranks	2j	3.75	7.50
	Positive Ranks	6k	4.75	28.50
	Ties	0l		
	Total	8		
Q5T4 - Q5FF	Negative Ranks	3m	2.33	7.00
	Positive Ranks	2n	4.00	8.00
	Ties	3o		
	Total	8		

Table 1. Table of signed ranks for the five questions from the evaluation questionnaire.

The results of the Wilconox test are shown in Table 2. All results were not statistically significant, p > 0.10. While the null hypothesis, which states that the workload while using each browser is identical, is tenable, we believe that this is almost certainly due to insufficient power rather than a true equality of workloads. Boxplots of the data are given in Figure 1; in particular, the responses Q4 (p=0.135) appears to be different between browsers but no statistically significant result was found.

	Q1T4 - Q1FF	Q2T4 - Q2FF	Q3T4 - Q3FF	Q4T4 - Q4FF	Q5T4 - Q5FF
Z	-.816[a]	-1.089[a]	-.378[b]	-1.496[a]	-.135[a]
Asymp. Sig. (2-tailed)	.414	.276	.705	.135	.892

a. Based on negative ranks.

b. Based on positive ranks.

Table 2. Results of the Wilcoxon Matched-Pairs Signed ranks test. No results were statistically significant.

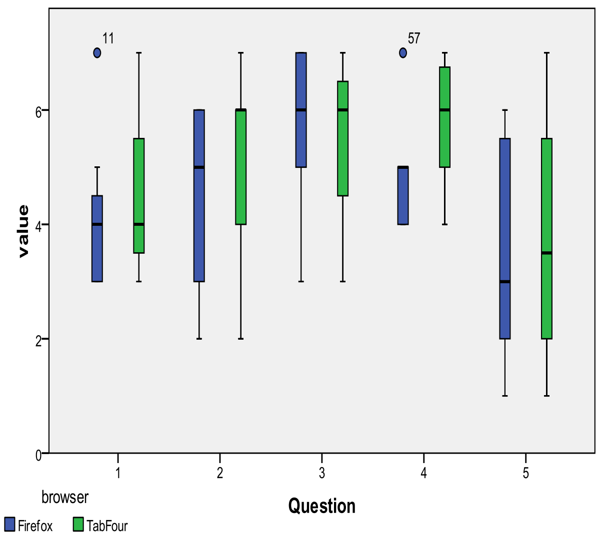

Figure 6. Boxplots of data from the evaluation questionnaire. In particular, Q4 seems to show a difference, although it was not significantly different.

At the end of the experimental session, each subject completed another questionnaire which contained questions comparing the two conditions, as well as general questions about task-based browsing. Questions 2 and 3 were forced-choice yes/no questions, with room for comments. In questions 1, 6, and 7, subjects choose between the TabFour browser and Mozilla Firefox; two subjects selected both browsers for Q6 and noted that they perceived no difference. Questions 4, 5, and 8 were free-response. For precise wording of the questions, refer to the Appendix.

All subjects reported that it was easier to complete the tasks using TabFour. All subjects reported that task-based browsing was useful, and that they would use task-based browsing if it was available in their regular web browser. Also, all subjects reported that it was easier to share webpages with other people in the TabFour browser.

Results were more mixed when subjects were asked in which browser it was easier to return to previously used webpages. The TabFour browser was chosen by 5 subjects, but 2 reported no preference and 1 preferred Firefox, stating that "Firefox lets you see all things at once." In TabFour a user may have to first select the desired task, and then select the desired tab within that task; if only a moderate number of tabs are opened the same result may be achieved by directly selecting the desired tab. We omit analysing this question with a statistical test since they have insufficient power to effectively deal with a data set this small; only a unanimous result is significant.

Qualitative Results

We received a moderate amount of feedback in the comment areas of the questionnaires and from the free-response questions, which is presented here.

Users frequently reported that they found things more organized in TabFour than in Firefox. For example, P1 states that "[TabFour]'s a lot more organized with the subcategories." P4 reports, "I like the nested tabs, more organized," and according to P7, "The organization is great to keep everything in order."

Users also frequently commented that it was much easier to get the shared links from their "friend" through the online shared tasks in TabFour, as opposed to receive a list of links in Windows Live Messenger. Subject P8 "Liked the TabFour browser better than having to open all the new tabs myself in Firefox." P7 reports that "It was a lot easier to use the tasks [Jane] made me" in TabFour. According to P4, the "Context file makes it quicker," and P5 says that "It was easier than trying to separately click and open each link."

User practices were mixed when it came to how they would complete the tasks. The first interruption was given a sense of urgency, to encourage subjects to switch tasks. At least one user, however, continued working on the main task until completion, and only switched to the secondary tasks at the end: he stated that "I always stuck to one task. I ignored Mark." Most users did switch tasks throughout the experiment, and reported that switching was easier in TabFour. For example, P5 reports that in TabFour "It was easier to jump in to a new task whenever I wanted and to leave the first task temporarily unfinished."

Limitations

The chief limitation of our evaluation is that it is extremely targeted on only a few features of the TabFour browser. Participants made heavy use of switching between tasks, and of loading previously shared tasks from the internet. Other features were optional and used only lightly if at all. For example, users saw no value in creating bookmarks since they knew they would only be using the system for a short time. The number of tasks used was still reasonably small and all tasks could fit in the task bar, so users had no need to suspend a task in order to move it to the "Inactive" sidebar. A longer field study could reveal if this feature really does aid resumption of tasks after longer periods such as days or weeks, as hypothesized. Finally, there was no format specified for the users' response during the study; they almost all extracted the relevant information from a page and typed it into an email or instant message manually rather than using the drag-and-drop feature or sharing a task over the internet. One user took a screenshot of the information on the webpage, cropped it in Microsoft Paint, and emailed the image to the experimenter.

Another limitation is that we only used extremely novice users. Each subject was shown how the various features worked by the experimenter, but did not actually interact with the system except when actively working on the tasks assigned to them. Users may behave differently as they develop strategies over time.

All tasks in the experiment were externally driven, and under fairly severe time pressure. It is not clear whether users would be willing to accept the small extra overhead of organizing tasks for their own less structured web browsing needs. Again, this would require observing real world usage over a longer timeframe than was practical for a laboratory experiment.

Section 3

Appendix

Evaluation Forms and Demo Video