Miniproject Report W07:
Evaluation Report and Requirements

Team: Jellybean

Topic: Using crowd-sourced business recommendations

Interface: Yelp. https://www.yelp.com

Date of Submission: October 18, 2018

V3.0 – Aug 30, 2018

Summary of Evaluation and Results *(~3pg)*

a. Evaluation Goals:

- Efficiency: How quickly can you find what you are looking for?
- Effectiveness: How effective is the search process compared to other websites?
- Learnability: How easy is it to learn to use the search function within the interface including sorting, filtering and navigating search results.

b. Evaluation Summary:

Our team developed four tasks to be performed by participants that utilized Yelp's search function and features to evaluate for our goals. We used triangulation by having two researchers observe the participants in addition to a questionnaire which gathered a range of quantitative, qualitative, objective and subjective data using multiple choice, long answer and likert scales. The observation was conducted in a controlled setting through specific parameters outlined in our coding sheet, afterwards a questionnaire was administered to participants which included self-reported feedback of the task. Our participant pool was composed of 10 UBC students in the age range 17-25 who had varied levels of experience with Yelp. Before beginning the tasks, each participants was made known of the purpose of the study and signed the consent form. All tasks were performed on a computer starting on Yelp.com and another tab with Google.com.

- **Task 1:** For task 1 participants were instructed to find "Jun Sushi" on Yelp.
- **Task 2:** For task 2, participants were asked to search for the same restaurant "Jun Sushi" on Google.
- **Task 3:** For task 3, participants were asked to find any restaurant on Yelp that allowed them to make a reservation, to conclude the task they had to show the observers the reservation button.
- **Task 4:** For task 4, participants were asked to search for the highest rated car dealership in Cleveland Ohio.

c. Evaluation Rationale:

Our team chose to evaluate the participants in a controlled environment so, as researchers we could focus closely on the actions and facial expressions of the participants while they were completing the given tasks and participants were not distracted by their surrounding which could impact the recorded time it took. Participants were also asked to employ the think-aloud technique to give us insight into their thoughts. Additionally they completed a questionnaire, post observation, which provided us with more quantifiable responses. These enabled us to summarise the qualitative data and find patterns between our recorded observations and their responses to the questionnaire. Experience with Yelp was not be required because one of our goals is to see how learnable the interface is so regardless of participant experience the data will still be relevant and provides a comparison to those who use

it often and those who do not. We triangulated our data by having two observers for each participant to increase the validity and credibility of our observations, and administered questionnaires. We are using statistical analysis and finding of patterns to analyse the quantitative and qualitative aspects of our findings.

d. Analysis (outcome of Step 9):

Some observed breakdowns were when participants when instructed to find a business in Cleveland Ohio, would conduct their search not in the designated area but told the observer that they had finished the task. Another issue was setting the location which was observed with participant 5, they repeated the step of entering the location and they self reported having difficulty with setting location. Participant 6 also had difficulty with the location function and stated “this isn’t Cleveland” after inputting their search query. It was also observed that participants had a hard time finding the filter options which was reflected in the questionnaire where 7 out of 10 self-reported issues with setting filters. One user said “a bit difficult to navigate through all the different filters”(P5) and another said they disliked the “layout of the filters”(P10) suggesting a visibility issue with filters.

A common behaviour among participants when searching for the highest rated car dealership did not use the ‘highest rated’ filter initially. They browsed through the list of car dealerships before realising that there was a filter that they could use.

One observed deviation for task 2, was participant 6 who deviated and went to another business aggregation website (TripAdvisor) from google to attempt to find the restaurant's website, which was unsuccessful. For task 3, one participant was clearly frustrated and clicked on restaurants at random from the Yelp search query with no clear goal in mind and did not complete the task.

We designed the tasks such that as the task number increased the tasks themselves also increased in difficulty (with the exception of task 1 and 2 which were for comparison between Yelp and Google and a baseline). We defined difficulty as the task having a larger number of steps and a higher complexity and specificity. Therefore, it was expected that the participants would have the most trouble with the Task 4, visible distress worsened as the tasks got harder as shown in Figure 1. As demonstrated with participant 10 in task 4 , who was rated as showing many signs of distress (rating of 4) said "how am I supposed to do this?","I don’t get it" and "I’m very stressed” during the execution of task 4. The observers’ rating of visible distress was confirmed by the participants’ comments, and their own rating of easiness for task 3 and 4. As shown in Figure 1, their own ratings of ease follow the same trends as the ratings of distress that the observers gave. Furthermore, as shown in Figure 2, the average time it took to complete each task increased as each task difficulty increased, with the exception of tasks 1 and 2, where task 1 took longer on average than task 2. Another reason we expected people to struggle the most with task 4 was because Yelp’s results not sorted by the highest rating, therefore we expected the participants to have some confusion finding and using the filters. Overall, we saw the time, distress, difficulty increased as the tasks became more complex and specific.

We expected that most participants would prefer to use Google. This initial assumption was confirmed by our findings. 70% of the participants preferred Google to Yelp and some of the reasons given were:

1. Google took less time to find the relevant information
2. Yelp allows for fake reviews or preferential exposure for money
3. It's easier to use and more common

Only 30% of the participants preferred Yelp to Google and their reasons were:

1. Got all the info needed in one page vs. different tabs on google
2. Easy to find restaurant and see reviews

Initially, we assumed that those who had used Yelp before would have better average times of completion, better efficiency and effectiveness ratings, and their ability to learn the interface would increase. However, we found that learnability was not affected by prior experience with yelp as eight out of ten participants who had used yelp before expressed high levels of difficulty at an average self rating of 3 when completing task 4 as shown by their questionnaire responses. Generally the participants that did not repeat steps had a better performance and were presumably learning the interface quicker.

A few outlier responses to be highlighted are, one participant did not answer the question regarding their prior experience with websites to find businesses information and for task 3 one participant took 3 minutes and 30 seconds to complete it when the average rate of completion was 44 seconds.

e. Conclusions (outcome of Step 10):

Some critiques to our evaluation process include needing stronger criteria to measure effectiveness in our coding sheet. Because it is a subjective concept when we observed the completion of tasks, more precise criteria would have aided us as observers to better quantify effectiveness. We also evaluated learnability but the coding sheet did not have clear enough parameters to be useful. Overall clearer goals in terms of analysing our data that we could directly relate to our evaluation goals would have been advantageous and helped us utilise all our collected data. A more consistent range of follow up questions would have been more useful than only asking certain ones after each tasks. This way we could have had a self-reported scale of ease across all tasks instead of just task 3 & 4. We also noted some discrepancies in timing the tasks between the two observers, one timer would have helped us avoid this mistake.

Task Examples and Requirements *(Max 1.5 pages)*

a.Task examples:

Susan is an 18-year old university student at UBC in her second year of university studying architecture. She is traveling from Vancouver to Portland for the weekend with a group of four friends, and she would like to make a list in advance of the places she and her friends would like to eat during their two day trip to Portland. Susan and her friends mostly want to eat in downtown Portland at the highest rated restaurants and bars that are not too expensive but are open to eating at other restaurants in different Portland neighborhoods if they are highly \recommended. She likes to research beforehand to guarantee that where she ends up eating will not disappoint. Susan would like to make a list as she will likely not have access to the internet while there, she wants to record the address and phone number of the restaurants she has looked over. She is an organized person but does not have too much time to dedicate to researching with her school workload.

Derrek is a 29-year old marine engineer. His two-year anniversary with his boyfriend, Peter, is coming up in two weeks, and he just discovered that the restaurant he had previously made a reservation for just canceled it for an event they will be holding on the same date. Derrek is desperate to find a nice, expensive restaurant somewhere in Vancouver, ideally more central, that will make a reservation for him in exactly two weeks, after all, he has decided he will propose to his boyfriend on that dinner. He would like to do his search while at work, so that his boyfriend has no idea of what he is planning for their special date. Additionally, he wants to contact the restaurants beforehand to make some arrangements regarding his proposal and ensure that there is availability for his chosen date.

b.Requirements (Step 12):

a) Must include:	b) Should include:	c) Could include:	d) Could exclude:
• Easy (10-second effort) querying and visibly identifiable filtering and selection by a user.	• Effortless (10-second effort) browsing of business categories by a user.	• Precise (10-second effort and one step) location constraint definition by an user.	• Straightforward (5-second effort) identification of top query results from other elements.

c. Justification:

The action of an identifying a visible filter and selection by the user was prioritized under *must include*, since in task #4, presented the most problems would have been aided by filters.

No participants self reported the task as easy and it also was reported to be the hardest task overall with the highest difficulty rating. Additionally, task #4 had the greatest time variability of the tasks, and all participants reported they struggled with the elements relevant in finding query filters and then identifying which filter they wanted to use.

The effortless browsing of business categories by a user was categorized as *should include*; as three out of six users who claimed they had some type of difficulty in task # 3, which required the user to find any restaurant that made a reservation, indicated that the option of finding a restaurant with the possibility for a reservation was ambiguous as nothing in the menu explicitly indicates this as a search option without typing reservations in the search bar or clicking under the restaurant drop-down menu.

Precise location constraint definition by an user was prioritized next as *could include* as while the option of searching a service nearby a location already exists it could be expanded to exclusively search on the location specified; as three people that repeated a step on task #4 when changing the location indicated that they had trouble changing the location as it would reset back to the Vancouver location or could not find exactly a car dealership in Cleveland, Ohio and would choose a car dealership near Ohio.

Finally, at the bottom of the priority list is the straightforward identification of top query results from other elements. This requirement could be excluded as the result query is already neatly laid out with the exception that Yelp places two ads for services before displaying the actual result query but they are very discreetly labelled as these ads blend in with the query results. Therefore a users' inspection of results is delayed as they have to recognize and distinguish what is part of the search and what is not. Three users mentioned the search result layout was confusing, crowded and that the ad placement was deceiving with the search results.

Appendix A

A.1) **Evaluation Instruments**

I. Coding sheet

II. Questionnaire

III. Procedure Walkthrough

I. Coding Sheet

Participant ID:
Observer:
Setting:
Date:
Start time:

	Task 1	Task 2	Task 3	Task 4
Time to complete task				
Repeated step? How many times?				
Areas of encountered difficulty				
State of distress on scale from 1 to 5 (1- no visible signs of distress, 2- barely shows signs of distress 3- mildly distressed 4- many sign of distress 5- extremely distressed)				
On a scale from 1-5 how effectively did the user complete the task? (1- very effectively, 2- pretty effectively 3- somewhat effectively 4- not very effectively 5- not effectively at all)				
Relevant things the subject said (words of frustration, confusion, satisfaction, etc.)				
Did the user improve? Were they able to learn the site?				

II. Questionnaire

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Computer Science
2366 Main Mall
Vancouver, B.C., V6T 1Z4

September 25, 2018

Participant ID: _____ (to be filled out by Researcher)

Experiment Questions

Human-Computer Interaction Course Projects (CPSC 344)

1. What is your age? (Circle one)
    a. 17-25
    b. 26-30
    c. 31-40
    d. 41-50
    e. 50+

2. What is your gender? (Circle one)
    a. Female
    b. Male
    c. Other _____
    d. Prefer not to say

3. In the past, what website have you liked to use to find information on businesses online?

4. Have you ever used Yelp before? (Circle one)
    a. If yes, how often
        i. Daily (everyday)
        ii. Weekly (a couple times a week)
        iii. Monthly (a couple times a month)
        iv. Yearly (a couple times a year)
        v. I have used Yelp, but do not use it anymore
        vi. Other _____
    b. No

5. Please circle which website you preferred when finding reviews on Jun Sushi restaurant (Vancouver, BC)?

a. Yelp b. Google

Why do you prefer this website over the other?

- In the following questions, you will need to evaluate on scale number from 1 to 5. 1 being very easy and 5 being very difficult.

4. Please circle on a scale from 1 to 5, how easily you were able to find a restaurant that you can make a reservation with using Yelp

Very easy	Easy	Neutral	Difficult	Very Difficult
1	2	3	4	5

Please circle (as many as you want) if you had trouble with any of the following steps:

A. Finding the option to find a restaurant with reservations
B. Choosing a restaurant
C. other _____

5. Please circle on a scale from 1 to 5, how easily you were able to find the highest-rated car dealership in Cleveland, Ohio using Yelp

Very easy	Easy	Neutral	Difficult	Very Difficult
1	2	3	4	5

Please circle (as many as you want) if you had trouble with any of the following steps:

A. Setting to option to find “car dealerships”
B. Setting location to Cleveland, Ohio
C. Setting the “most reviewed” search filter
D. Locating the most reviewed rated car dealership in the list
E. other _____

6. After using the Yelp website, would Yelp be your number one choice for reading reviews about a business?
    a. Yes
    b. No
        i. If not, write down one other website that you prefer over Yelp (if any)
        _____

7. After using the Yelp website, would Yelp be your number one choice for finding a business's contact information?
    a. Yes
    b. No
        ii. If not, write down one other website that you prefer over Yelp (if any)
        _____

8. After using the Yelp website, would Yelp be your preferred method of making restaurant reservations?
    a. Yes
    b. No
        iii. If not, write down one other method that you prefer over Yelp (if any)
        _____

9. What did you like about the website's layout? (Leave blank to indicate nothing)

10. What did you dislike about the website's layout? (Leave blank to indicate nothing)

III. Procedure Walkthrough

Procedure Walkthrough

Good afternoon, my name is …. And …., we will be your experiment observers today. The goal of this experiment is to study your interaction with the business search website Yelp. For this study, you will be asked to complete a few tasks online and fill out a questionnaire. Before we begin can you please sign this consent form outlining that you agree to participate in our study and that your participation is completely voluntary and can be withdrawn at any time.
Any questions? (answer the questions)

Begin observation, there will be a computer with yelp.com open... ask participant to :
(do tasks in order)

1. Look up Jun Sushi restaurant (Vancouver, BC) on Yelp (start timer)
2. Look up Jun Sushi restaurant (Vancouver, BC) on Google (reset timer)
3. Find a restaurant that you can make a reservation with on Yelp (reset timer)
    - On the restaurant's page, you have to be able to find a "Make a Reservation" button
4. Find the highest-rated car dealership in Cleveland, Ohio on Yelp (only Cleveland, not nearby cities) (reset timer)
   (expected answer: Enterprise car sales, montrose volvo, etc.)
    - Once found, please take a second to look at the car dealership's hours of operation, address and phone number
5. Complete the questionnaire

(Complete coding sheet)
Thank you for participating in this study. Do you have any questions?

A.2) **Supplementary analysis**

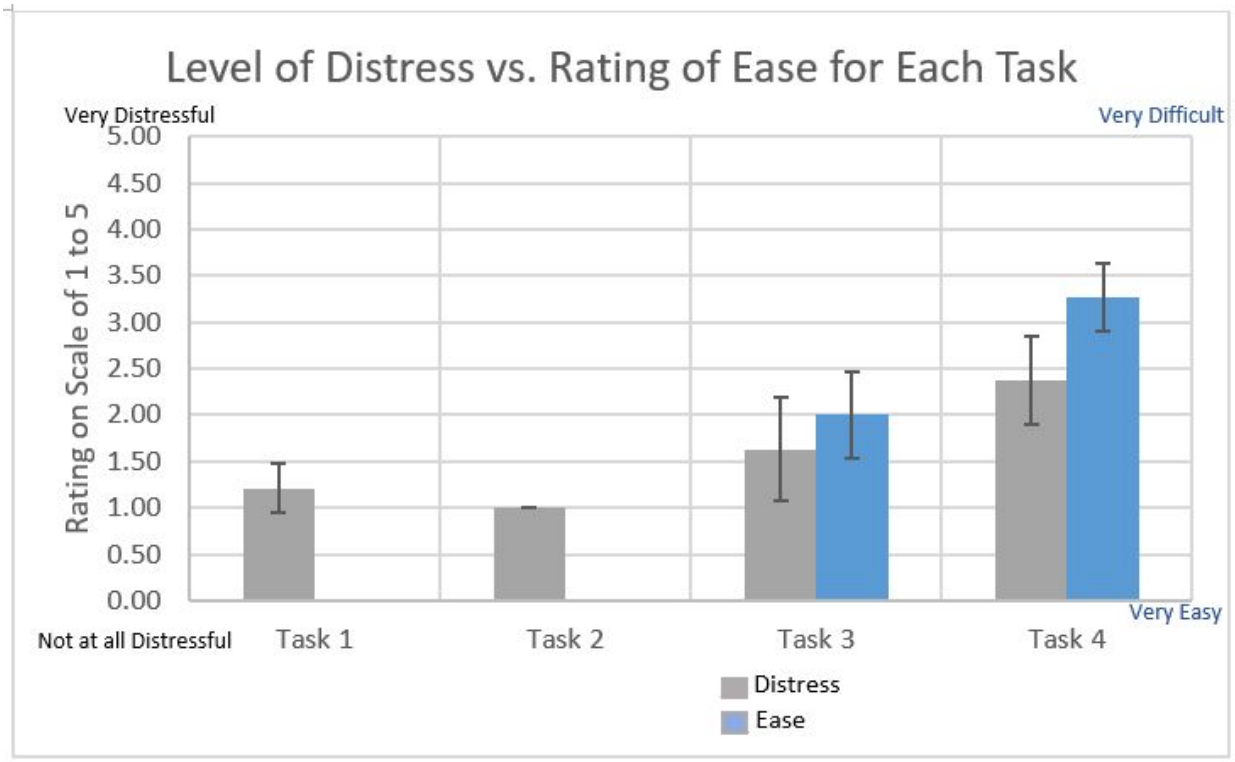

Figure 1. Graph of level of distress for all tasks and participants' rating of ease of tasks 3 and 4, with standard deviation error bars. Both ratings of distress and ease were given based on a scale of 1 to 5, with 1 being the least distressful and easiest, and 5 being the most distressful and the most difficult.

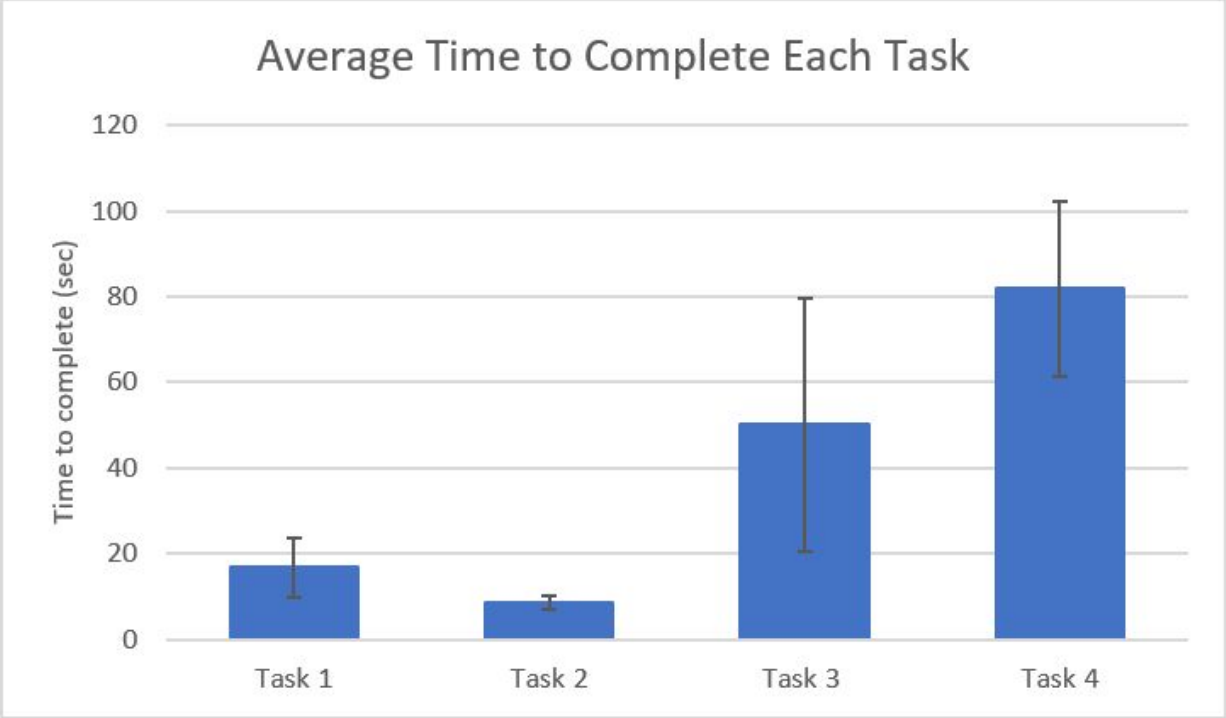

Figure 2. Graph of average time (in seconds) to complete each task, with standard deviation error bars.

Appendix B - Additional Documentation

B.1) Adherence to Ethics Protocol

B.2) Signed participant consent form(s) from the evaluation

B.3) Raw Data