

Alfred: A Visualization Tool to Facilitate Decision Making Through Large Scale Comments

Aditya Narula

UMass, Amherst

anarula@umass.edu

Mahmood Jasim

UMass, Amherst

mjasim@cs.umass.edu

Shanu Vashishtha

UMass, Amherst

svashishtha@umass.edu

ABSTRACT

The advent of internet has led to a massive increase in the user content generated on a daily basis. In online shopping websites, in particular, every product comes with not only the manufacturer's and product information, but also user comments from the users who have purchased the product before and new users looking to buy the products. The users need to browse through a plethora of comments before deciding whether to purchase a product. The existing process is time consuming as well as cumbersome. To make the process streamlined, we have designed and developed a new web-based prototype called, Alfred. It is a visualization tool which presents users with a clear interface where they can explore to gather relevant product information in limited time. Alfred also provides a comparison feature where the users can compare between two products simultaneously. We performed a within-subject laboratory study where we evaluated Alfred by 18 participants who compared Alfred with industry-leading online shopping website, Amazon. Our results show that Alfred provides better categorization and comparison features as compared to Amazon while providing richer information paving the way towards faster comparison between products and decision making.

INTRODUCTION

With the advent of internet, there are ubiquitous online platforms to share opinions, ranging from blogs and news outlets to social media and review sites. As a result, the amount of user opinions generated is massive. One can easily find hundreds if not thousands of comments reviewing a product on Amazon.com, a restaurant on Yelp.com or any job experience on Glassdoor.com. While the rise in the volume of content generated is constructive, it takes enormous effort and time to draw meaningful conclusions from them. Our observations show that an average user browses through at least 20-30 comments before arriving at a decision. Each of the users in these scenarios act as an independent decision maker.

To circumvent this issue, there is a growing need for a solution to help users understand the opinions and extract useful

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI'16, May 07–12, 2016, San Jose, CA, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 123-4567-24-567/08/06...\$15.00

DOI: http://dx.doi.org/10.475/123_4

information in a quick and efficient manner. In our work, we approach this problem through our proposed web-based prototype tool - Alfred. Our proposal is to develop a solution incorporating text analytics and data visualization, which makes it convenient for online platform users to incorporate the gathered views in their decision making process. Through this work, we intend to make this step of the decision making process an effortless exercise.

Our prototype Alfred built as a web browser interface helps users in performing common tasks like browsing through comments, extracting a gist of those comments and compare between products based on these comments. We present a description of our iterative design process which led to the final prototype. In this literature, we have focused on exploring the comments of an online shopping system. We performed a within-subject user study to evaluate Alfred by comparing our prototype with Amazon - arguably the leading online shopping website today. Our results demonstrate clearly that our system helps users perform the listed tasks while reducing the time it takes for them to do it simultaneously on Amazon.

LITERATURE REVIEW

Researchers have proposed various visualization systems to view large scale user comments generated throughout the last decade in several domains. There is an abundance of such systems where opinions of platform users can be succinctly presented visually, but only a few have come with generalizable solutions. Most notable of these include Opinion Observer [5], a tool for analyzing and browsing online opinions. It helps compare the sentiments of users related to different features of a particular product and also provides a count of positive and negative comments of specific product features. The users of Opinion Observer can choose from multiple products and compare them on a point wise basis. To understand people's opinion on products, Carenini et. al. [4] used template based text summarization to help understand public input. In a more focused environment, OpinionSeer [7] uses a visualization-centric opinion mining technique that considers uncertainty for faithfully modeling and analyzing customer opinions for hotels. Aside from these standalone platforms, Amazon [1] allows users to post comments, provide a rating for that particular product as well as tag other user comments if they found it to be helpful or not. Some helpful keywords are presented as well to the user which is a culmination from thousands of comments already posted reviewing that particular product.

Although, most of these systems are good for providing a quick overview, they do not provide a robust way to explore

deeper. While Amazon provides users the options to view all the raw comments, it cannot extract a quick summary for their use. OpinionSeer uses an unfamiliar visual encoding which is difficult to use and although informative, too complex to grasp for a user. Opinion Observer on the other hand fails to capture nuanced information and suffers from an unintuitive interface. Moreover, user comments are usually descriptive in nature while its interface is suited for short comments only. As a result, the exiting tools fail to be comprehensive in presenting the user comments lucidly so that the users can make informed and faster decisions about buying or not buying a particular product.

NEED FINDING

For the Need finding stage, we started off with an initial Unstructured Observation. The following were our original focal points which we chose. All of these were scenarios where user generated comments is used by other users to make a final decision - Comments in Dining Commons, Hotel reservation, Restaurant Selection and Online Shopping. We carried out these early observations in an informal house setting, a public social gathering and in a public space respectively. From the notes made by us, a set of observations were made and then, a brief questionnaire was prepared for the interviews. These notes also formed a basis for the Survey questions which had some common questions related to the broad topic and some specific questions to each of the focal points. Participants(17) ranged from colleagues, friends to random people who agreed to take part in the survey in a public place. Each one of us carried out one interview from the questions prepared.

Some of the findings at this stage of the process were that - The interviewees would prefer a platform that removes the profane part of the comments while bringing out the important parts that require action, 59% of the respondents considered the labelling of experiences on the websites as positive or negative helpful, 50% of the respondents gave a preference for Visual info-graphic depicting comment statistics and a drill down to specific comments, 67% of the users browsed through 2 pages of comments before deciding a hotel and 83% agreed that those comments can be presented in a simpler fashion to them.

The key takeaways for us as tool designers were to present users with all the comments while at the same time giving importance to visual representation of the information in the design process. We also had to emphasize important features associated with them. In the same context, filtering inappropriate and extraneous content will make it easier for the users to traverse through the comments.

SYSTEM

We describe the different stages of the prototype stage sequentially below.

Low fidelity Prototype

We created our low fidelity prototype using paper, color pens and sticky notes. A video walkthrough of doing the task examples using the paper prototype can be found here [2]. Refer to 1 for a brief visual explanation of performing the task examples

using the paper prototype. The landing page is the 'Search for Products' page which allows the user to enter and search for products. If the product category (e.g. Laptops) is not in the database, an error is generated and the user must search for another product. Hitting the 'GO' button, will take the user to the product view as shown in the figure. The products are displayed as tiles on the left window which can be scrolled through. Each Tile has a complete clickable area and a comment icon on the top right. It also has a radio button on the top left corner to select products for comparison. Clicking anywhere on the tile for a product will generate a 'Summary View' which provides an aggregate view of all comments for that product based on 3 features: Sentiment, Vulgarity and Star ratings. The Sentiment is shown in the form of a Donut chart to show the distinct sentiment and the appropriate number of comments in each category (Positive, Negative and Neutral). The Vulgarity is shown in the form of a bar chart with gradient color scheme such that X-axis represents percentage vulgarity in comments and the Y-axis represents the number of comments in that bin. The Star ratings feature shows the average ratings for the product over all comments. All these feature visualizations are clickable and clicking on them takes the user to an 'All Comments' view.

The 'All Comments' view can also be reached by clicking on the comment icon on each product tile. Moving on to the 'Comments View' where the user can see all comments and the features of each comment listed next to them on the left. The user has an option to browse through the comments, flag them for printing later and also see keywords at the top of the page.

Coming back to the 'Product' view, the tiles have radio buttons which can be clicked to select the products for comparison. Clicking on the 'Compare' button takes the user to the 'Compare Products' view which allows the products to be compared using a drop-down and based on particular features, again selected using a drop-down.

The 'print' icon exists on the bottom right corner of each page. It becomes grey when disabled and blue when enabled on a screen. Clicking on it takes you to the page which has the flagged comments for printing.

Medium Fidelity Prototype

We created the medium fidelity prototype in Axure RP 2. Only a single version of the medium fidelity prototype was created since we have already created and experimented different versions of the system with the paper prototype. Based on the results from the user testing of the paper prototype, we incorporated the feedback while creating the medium fidelity prototype.

We built up on our paper prototype for this stage of the process. Our medium-fidelity prototype has a 'Home Screen' where one can choose amongst the products. Upon selecting the product and clicking on the 'Go' button, the user is taken to the 'Product view' page. All the products are listed on the left. Upon selecting the product, attributes related to the product comments are displayed on the right side. The attributes include the overall sentiment which is present as a pie chart, overall vulgarity bins which is present as a bar chart and the star rating at the bottom. There is an 'All comments' icon

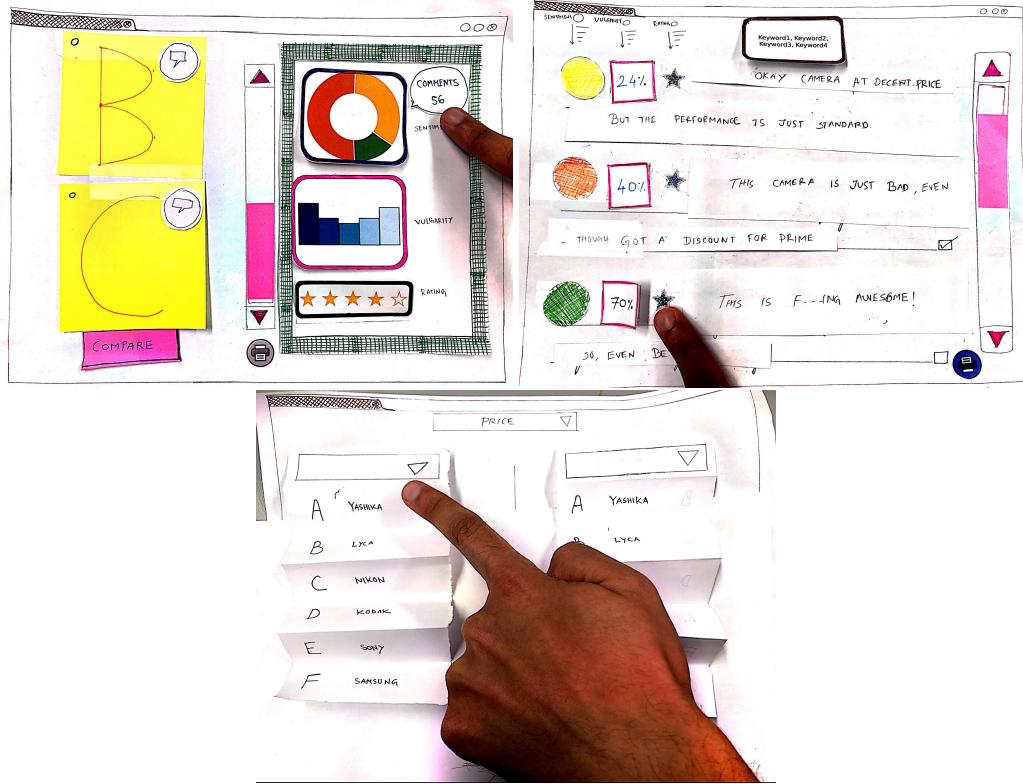


Figure 1. Paper Prototype with the Product view, All Comments view and Compare view in clockwise order

which takes a user to the 'All comments' view and displays all the comments related to that particular product. Alternatively, if the user wants, they can also view the comments sorted according to the star rating, sentiment and the vulgarity score for those products. These open on separate tabs for the user to browse through the comments. At the bottom of the page, we have a 'Print' icon which can be used to print the comments which a user finds useful for their purposes. The last feature we support in the prototype is to compare between products based on comments. We choose the two products to be compared and then, in the 'Compare view', the user can select the criteria to compare the comments of those products.

At this stage, we received the feedback that our 'Product view' design led to space wastage on the right and we could use it to display some of the comments there itself. Also, our vulgarity bin colour coding scheme was difficult for users to follow and they asked us to be more consistent while using those across the views. A minor issue was that the selected product on the 'Product view' didn't highlight which made it difficult for users while interacting with the system.

Final prototype

Based on the feedback we received, we set the following goals for our final prototype design -

1. Our final prototype should provide extensive information
2. It should allow product comparison
3. It should help categorize the comments in an efficient manner

4. It should provide a helpful visual interface

Our final prototype - which we are calling Alfred - is geared to help any user perform all the three task examples we have been tackling from our personas. We created the prototype with a single product category - cameras. Addition of new product categories is a simple code replication process and hence, as quickly more products can be listed on our system. In our home screen, we show two different categories- books and tools - which will be added soon. Also, we do not have a limitation on the number of products that can be listed on our system. We however would consider placing a search bar to find categories of products.

We describe Alfred's individual views in detail here -

Product View: It displays the product list on the left, the option to go to All comments, the product's comments sentiment, vulgarity bins and the star rating in the middle. Towards the right, we display individual comments which come up when they are clicked. When clicked on summaries we generate corresponding comments under the "comments" section. We color code to match the font-color with the selected part of the graph(s) to highlight the selection. A title header containing the product name was also added to reflect which product is selected.

Comments View: In the comments view, where we display all comments for the selected product, we have added checkbox selection to allow users to identify useful comments for printing when they require.

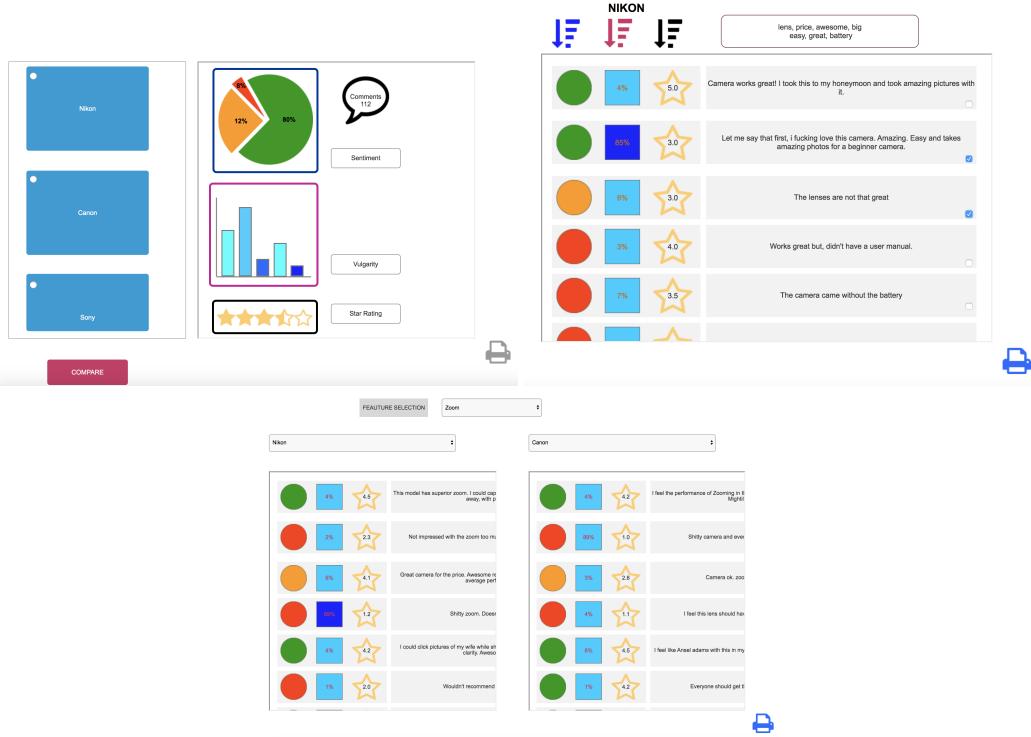


Figure 2. Medium fidelity Prototype with the Product view, All Comments view and Compare view in clockwise order

Compare View: In the Compare view, we provide two options to select the products and at the top specify the comparison criteria which by default is All. From the pull-down, the user can select the product for which s/he wants to compare the comments. For all the comments, we display the sentiment, vulgarity and the star rating at the beginning of each comment.

Our prototype has a visual interface which has been developed using JavaScript and HTML/CSS and can be viewed in any web browser. We provide mouse interaction based support in the system. The first layer is via buttons which can be clicked to open new views. Alternatively, the Print button displays the print view with all the comments for printing, Compare button takes you to the compare view for comparing the products and the About view displays the system information for Alfred. In the product view, one can hover over the chart elements to gather total number of comments belonging to a particular category. Upon hovering, a color change is also shown in those elements. These chart elements are again clickable and display all the comments corresponding to that particular chart element when clicked.

In our final prototype, we followed the design principles we covered in class. We followed the Constraints rule where the features that were under process of being added were grayed out while the yellow ones were used to display the active product categories in our system. We used the Constant feedback iterative approach where we improved upon our prototype from the low fidelity prototype towards the medium fidelity prototype and at last, the final prototype. We also incorporated a consistent colour coding scheme where green corresponded to positive comments, red to negative ones and yellow to neu-

tral comments. Our comments section was modeled according to a To-do list which reflects the natural mapping process in the design.

DESCRIPTION OF EXPERIMENT

We conducted our user test with 17 participants. The studies were conducted either in the participant's home environment or in the Computer Science department. In both cases, the environments were controlled to simulate lab settings.

Participants

As described in Test-1 report, we recruited our participants using the word of mouth snowballing technique [3]. A total of 18 participants consented to take part in our experiment. We asked them to fill a short Google Form which captured their demographics. All the participants were students from the UMass Amherst campus. Most of them belonged to the age group of 21-30 while the rest were from the group 31-40. Two-thirds of our participants were male while the other third were female. Also, three-fifths of our participants purchased products on a monthly basis while 4 did so on a weekly basis. The other 3 purchased products on an yearly basis. A large majority of our participants(16) shopped online on Amazon.

Summary of Data Collected

After our participants completed the demographic questionnaire, we described them the tasks for them to complete on Amazon as well as our prototype Alfred. The tasks included browsing through all comments related to a product till the point they felt comfortable in making a decision whether to purchase the product or not, extracting a summary of all the



Figure 3. Prototype with the Homepage, About section, Product view, Product view with positive comments, Product view with mildly vulgar comments, Product view with 5 star comments, Product view with flagged comments, Comments view with all comments, Compare view and Print View in clockwise order

Table 1. Post-Task Questionnaire

Question Id	Question
Q1	How was your overall experience with Alfred/Amazon?
Q2	Alfred/Amazon helped me find useful information from the comments.
Q3	Alfred/Amazon helped better categorize the comments.
Q4	Alfred/Amazon helped me compare the products using comments.
Q5	What do you think about the visual interface of Alfred/Amazon?
Q6	How easy was it to use Alfred/Amazon?
Q7	I prefer Alfred/Amazon over other online shopping website interfaces.

Table 2. Post-Study Questionnaire

Question Id	Question
Q1	Which system provides a better experience?
Q2	Which system presented more information from comments?
Q3	Which system allows better comparison of comments?
Q4	Which system allows better categorization of comments?
Q5	Which system has a better visual interface?
Q6	Which system was easier to use?
Q7	Which system do you prefer?

comments in terms of sentiment as well as star ratings and comparing the comments of two different products. While the users were performing the tasks, we measured the time they took to perform each of the tasks using an external stopwatch on our mobile phones. The participants were not aware of us timing their actions. We did so to avoid the Hawthorne effect[6].

We administered two post-task questionnaires after our participants completed the three tasks we asked them to accomplish either on Alfred or Amazon. Initial assignment of the system to begin the task was completely random. The task questionnaires were focused on asking the users how was their overall experience with the system, if they were able to perform the intended tasks and if it was easy for them to perform the task on the system. Once they filled out these questionnaire, we conducted a final post-study questionnaire which asked the users for an explicit comparison between the two systems in terms of ease of use, support for tasks and their visual interface. All of these were collected online using Google Forms. The questionnaires are presented in Tables 1 and 2. A short unstructured interview followed the entire study inevitably where almost all the participants provided their informal feedback on Alfred and their struggles with the existing Amazon setup.

Procedure

We started the study by collecting demographic information as well as measuring the participants familiarity with online shopping websites using a pre-study questionnaire. After the questionnaire, the participants were asked to perform tasks (**T1 - T3**) on two conditions (Alfred and Amazon). The two conditions were assigned randomly to the participants. After the tasks were performed on each condition, the participants completed a post-task questionnaire which collected their subjective opinion on the conditions. Following the post task questionnaires, we asked the participants to fill up a final post-study questionnaire which asked for their preference among the two controls. We performed a quantitative analysis based on the time it took for the participants to complete each task.

The participants didn't know that they were being timed while the tasks were being carried out by them. This was done to minimize the Hawthorne effect that could have crept in during the study. We also conducted a short unstructured interview asking the participants' feedback on Alfred and any suggestions that they might have about it.

Tasks

The participants of our user study were asked to perform three tasks on Alfred and Amazon. These tasks are informed based on our observations, interviews, personas along with the proposed goals. The tasks are designed to verify the hypotheses as described in a future section. The tasks were as follows:

T1. Start by searching a product category. Select a product from the list of products. Go to all the comments related to the selected product. After exploring the comments, identify the ones you find important and you think will help you make the decision whether to purchase this product.

T2. Start by searching for a product category. Browse the comments related to the product based on specific features of those comments. While exploring the comments, look at their features displayed on the systems. Identify the ones you find important and you think will help you make the decision whether to purchase this product.

T3. Start by searching for a product category. Select multiple products based on their features. Compare between the two products' comments. Decide which product you would rather purchase based on the comparison.

Hypotheses

We had the following Hypotheses during our experiment:

H1. Alfred provides more information from comments compared to Amazon.

H2. Alfred allows better comparison of comments than Amazon.

H3. Alfred provides a better visual interface than Amazon.

H4. Alfred helps to find important comments faster than Amazon.

RESULTS

We compiled the results from the post-task and post-study questionnaire as well as the task timing based on the amount of time participants took to perform the three tasks. We collected the responses of the post-tasks questionnaire on a 5-point Likert scale. The responses were normalized and paired T-tests were performed on the collected responses. The results are presented in Figure 4 A and B, where A presents the mean Likert scores for post-study questionnaire responses and B presents the mean time required in seconds for performing three tasks.

The results show that the difference between means for Question 3 is statistically significant which asks the users about better categorizing the comments, surfacing richer information that can be accessed in a convenient way. The p-value was

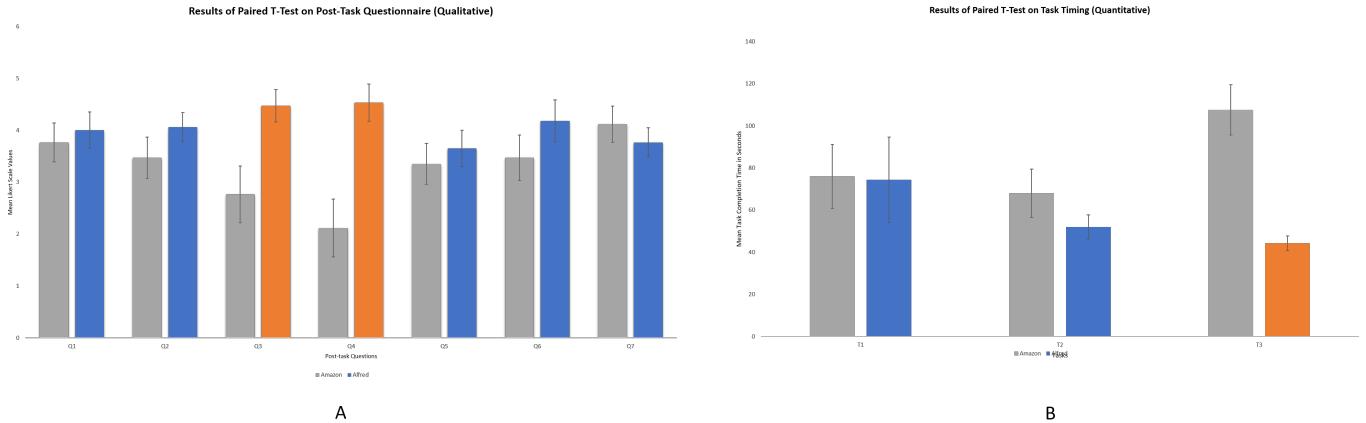


Figure 4. Analysis of the Experiment.

Table 3. Result of Chi-Squared Test from the Post-Study Questionnaire

Question Id	Chi-Square	p-Value
Q1	.059	0.808
Q2	9.941	0.002 (<0.05)
Q3	17.000	0.000 (<0.05)
Q4	9.941	0.002 (<0.05)
Q5	.059	0.808
Q6	2.882	0.090
Q7	2.882	0.090

0.0014(<0.05) between Alfred and Amazon for Question 3. Hence **H1** was supported.

The results also show that the difference between means for Question 4 is statistically significant which asks the users about comparing between two products based on comments. The p-value was 0.0003(<0.05) between Alfred and Amazon for Question 4. This proves that Alfred provided better comparison functionalities between the two conditions. Hence **H2** was supported.

However, Question 5 asked the users about their preference on the visual interfaces of the two controls. The p-value between the two means was 0.2558(<0.05) showing that the users preferred Amazon's visual interface over Alfred's visual interface. Hence, **H3** was not supported. We argue the reason behind it is the familiarity of the users with Amazon. One participant (P3) mentioned,

I am using Amazon for a long time and I just know where to look at when I am searching for a product or looking for a comment

The learning curve required to get familiarized with Alfred might have also contributed as suggested by another participant (P11) who mentioned,

Some of this [Alfred's features] might need some getting use to [...] like, when I am looking at sentiments, I need to think about how to process this [...] to fully understand what the comment says about the product.

Finally, when evaluating which system provided a faster service to the users, we performed paired T-test on the time required by the users to performed the tasks as shown in Figure 4 B. The results show that although the time to perform Tasks 1 and 2 (**T1** and **T2**) was not significantly different with p-values of 0.4102(<0.005) and 0.2753(<0.005), the time to perform task 3 (**T3**), which asks the users to compare between two products was significantly different with a p-value of 0.0433(<0.05). Hence, we claim that **(H4)** was partially supported. We also posit the reason behind (**T1** and **T2**) not being supported is also the familiarity of using Amazon. However, Amazon does not provide a proper comparison option between multiple products based on comments. Hence, it took users a significant amount of time to compare between two products. One participant (P5) mentioned,

The compare view is new. Comparing on Amazon is so cumbersome. [...] This has to be Alfred's highlight!

The post-study questionnaire pitted Amazon and Alfred against each other asking the users which one did they prefer. We performed Chi-squared significance test to compare between the two conditions. The resultant p-values of the Chi-squared test for the post-task questionnaire are presented in Table 3. The results show that for Q2, Q3 and Q4, the participants found Alfred to be significantly better than Amazon. These questions asked the users about which system presented more information, provided better comparison and provide better categorization mechanisms respectively. These results aligns with hypotheses **H1** and **H2** being supported.

DISCUSSION AND LIMITATIONS

We received a range of comments and constructive feedback from our pilot and user studies. Initial feedback on medium fidelity prototype helped us improve on the high fidelity prototype's design and interactions. An important design change suggestion was that when we click on specific parts of the visualization, the filtered comments corresponding to that part opened in a new tab on the browser. Based on the suggestion, we decided to make use of the white space on the right side of the screen and show the comments based on the click there. The other upgrades we made while creating

the final prototype were related to the color schemes as the previous colors did not portray the meaning well and we also improved on the contrast and choosing a better gradient scale for the vulgarity bins. Users found the Compare View to be extremely useful in Alfred, and spend considerable time on the page to compare products on the basis on comments filtering on features selected from the dropdown.

In terms of the interactions, we added interactions on each sector of the visualizations. The hover feature and corresponding color change helped users understand that the areas were clickable. Clicking on each of the stars showed the comments having the particular star ratings.

The feedback we received was related to the different areas of the application. For the look and feel, the users felt they required a little more explanation to get an idea of what is happening where. We also found that our products on the product list lacked image thumbnails and hence were difficult to find by the users. Users also asked for a search bar to find a particular product category. Users also wanted extra features such as showing common keywords extracted from the comments.

CONCLUSION AND FUTURE WORK

The results show that Alfred provides richer information containing sentiment, star ratings, vulgarity and the review text itself providing better insight into the user comments. It also provides better categorization than its contemporaries that allows users to explore user comments with ease and convenience. Finally, the users found Alfred's Compare View as useful in making decision about which product to buy when conflicted. The feature was popular among users also because Amazon does not provide such a feature and this provided a novelty to the user.

In addition, users were appreciative of the visualizations and comments on the same page as compared to Amazon where users need to scroll multiple pages down to reach the comments section.

We could not establish that Alfred provides a better interface due to Amazon's extensive familiarity among the users.

For the future work, on a priority basis we need to include the image thumbnails for each product. Other changes would be to better categorize comments after performing sentiment analysis on the comments. We also aim to extend the usability and application of Alfred to other domains where user's decisions are based on large number of comments.

REFERENCES

1. 2019. Amazon. (2019). <https://www.amazon.com>
2. 2019. Paper Prototype Video. (2019). <https://drive.google.com/file/d/1554ZaB5H0io1jG1fno3ldmxQ9MOM4Ua8/view?usp=sharing>
3. Patrick Biernacki and Dan Waldorf. 1981. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research* 10, 2 (1981), 141–163.
4. Giuseppe Carenini and Lucas Rizoli. 2009. A Multimedia Interface for Facilitating Comparisons of Opinions. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI '09)*. ACM, New York, NY, USA, 325–334. DOI: <http://dx.doi.org/10.1145/1502650.1502696>
5. Bing Liu, Mingqin Hu, and Junsheng Cheng. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*. ACM, New York, NY, USA, 342–351. DOI: <http://dx.doi.org/10.1145/1060745.1060797>
6. Rob McCarney, James Warner, Steve Iliffe, Robbert Van Haselen, Mark Griffin, and Peter Fisher. 2007. The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology* 7, 1 (2007), 30.
7. Liu S Au N Cui W Zhou H Qu H. Wu Y1, Wei F. 2010. OpinionSeer: interactive visualization of hotel customer feedback. In *IEEE Transactions on Visualization and Computer Graphics (Volume: 16 , Issue: 6 , Nov.-Dec. 2010) (IEEE)*. <https://ieeexplore.ieee.org/document/5613449>