Visualizing Case Studies

ABSTRACT

Case studies contain valuable information about development records. They are, however, usually presented as textual documents. Here we begin an initial design for viewing case study data. A prototype for viewing case study data is created using information visualization techniques for the purpose of testing against existing techniques. The evaluations show that the new design is just as usable as existing techniques, making it suitable for future development. Future work may show our design can accomplish tasks both faster and more efficiently.

Keywords

Information visualization, case studies, heuristic evaluation

INTRODUCTION

Case studies document the development record of a project. They provide the user with an insight into what occurred and relevant details of the process. A person can gain valuable knowledge that can be reused in their own projects and allow their own system to be better simply by learning from what others have done.

Problem

Unfortunately, case studies are almost always textual documents with, perhaps, a few pictures. There is no set structure beyond what each individual prefers to offer. Even the CHI format, used for this paper, has only a few requirements and those are mostly about presentation and not content. This leads to problems with all three elements of the Information Visualization mantra: overview first, zoom and filter, details on demand.

An abstract can only provide a brief overview. It tries to squash a long paper into only a hundred words or so. A lot of information is lost with this process. Quality of the abstract also depends on the author’s ability to compress their work, and can vary widely. Therefore the abstract is a poor indicator to rely on for knowing precisely what is contained in the document.

Zooming and filtering have only a loose connection to reading the text document. A reader can certainly filter what they read based on quick glances, but there is no guarantee they have filtered out what they meant to. Without reading a section, one can never be certain of what it contained. Also, since all of the details are presented on the same hierarchical level, zooming makes no sense, as the reader either sees the data or does not see the data. So the only zooming is deciding what page to look at.

Similar to the zooming problem, details are not just provided on demand, they are provided whether the user wants them or not. Again, the user can only chose whether to look at a page or not, and that is the only demand a textual document can meet.

Research Questions

Since a case study contains valuable data, it follows there should be an efficient way of extracting that data. Unfortunately, we are stuck with a pure textual document. As shown in the Information Visualization class, a details only view is the worst way to gain insight into any data.

We propose there must be a better way of viewing this data. In order to establish such a method, we must first answer the following questions:

- Whowillbetheusersofthisvisualization?
- Whatisthedatathatshouldbeviewed?
- Whattaskscanbeperformedwiththisdata?
- Whataresomeoftheinsightsthiswillallow?

After we answer those questions, we can move on with the project goal. That will address our main research question: Is our new design as usable as the interfaces in current use?

Goals

The goal of this project is not to develop a new information visualization interface for case studies. We are not at that step yet. Instead, we wish to come up with an idea for such an interface, and evaluate the usability of that idea versus other approaches now in use. If we can determine that our new idea is not a regression in usability, then we will feel confident in moving forward with the design.

To that end, our new design will only be created in prototype form. This will affect the experiment performed, as explained later.

LITERATURE REVIEW

The visualization of case studies is a new area of research, and little literature exists specifically discussing the implementation or evaluation of such tools. However, literature does exist for such related topics as the need for case-based learning in HCI, the use of information visualization techniques to read and understand electronic documents, the use of timeline interfaces in information visualization, and the user of scenario-based design in the development of a software system. In support of the need for case-based learning, Wixon [6]

argues that the literature surrounding the evaluation of usability methods lacks relevance in terms of real world applications. Software products are produced and tested for usability every day, and it is much more effective to learn from real products as opposed to simulated systems or hypothetical models. A case study approach allows designers of real world products to view past - successful or unsuccessful - examples of other real world products, which are the only examples complex enough to address the issues at hand. Furthermore, a collection of relevant case studies would allow the discovery and analysis of patterns to be generalized across multiple cases.

John [3] suggests the use of case studies in understanding the advantages and disadvantages of usability evaluation methods (UEMs). The case study approach aims to collect a multitude of data types and apply them to an explanation of what happened. In the case of usability studies, what happened refers to the details of the design and implementation of a specific product. The data collected can include numerical results, diaries of user participation, and problem description reports.

John warns that the adoption of the case study approach in usability evaluation also calls for the need to produce case studies effectively. The design and interpretation of case studies is a vital aspect of their use in HCI research. Thus, case study visualization techniques could either help or hinder the case-based approach to usability evaluation.

Henninger, Haynes, and Reith [1] state that, although usability guidelines are a vital aspect of the development of successful new interfaces, these guidelines tend to be either too abstract or too technology-specific. Studies show that, when designing new interfaces, development teams examine and reference concrete examples of existing interface designs more often than they incorporate usability guidelines. Thus, this article suggests an organizational memory approach to developing design guidelines, a system in which a case study of the development of an existing interface is attached to the general guidelines used in that design. The case provides users with examples of how certain guidelines were used, and those existing designs can either be reused or altered to fit the needs of the current project.

The Mimir Design Guidelines System incorporates this case-based usability guideline approach by allowing the user to search for specific cases or view cases related to a certain guideline. Cases include user comments that explain issues and tradeoffs specific to the project. They also provide links to related case studies, helping the user to find all information related to the current project (Figure 1).

Although this approach allows the user to access the information in a case of a project, the study is displayed as a plain text document, which requires the user to read the entire case word for word (Figure 2). The article, however, supports the need for experience-based learning in the realm of interface design, a need that will be facilitated by a case study visualization tool.

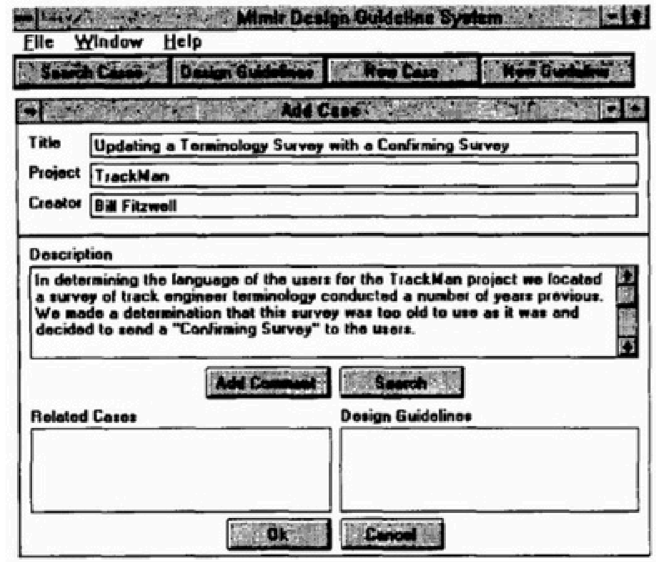

Figure 1. The Mimir interface for adding a new usability case.

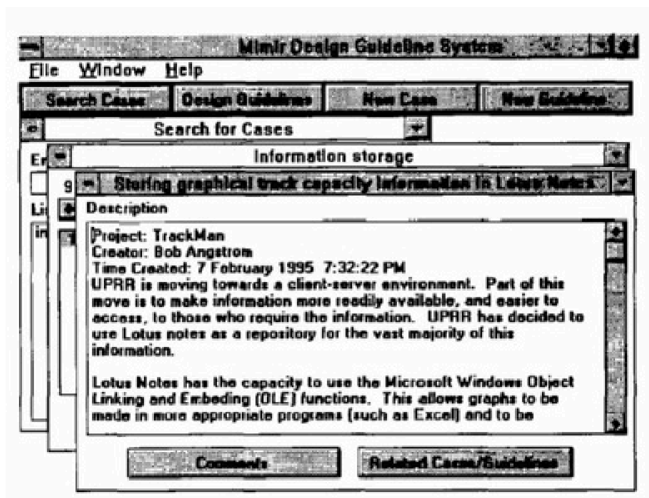

Figure 2. The Mimir interface for displaying a usability case.

Hornbæk and Frøkjær [2] examine the usability of three interfaces utilizing information visualization techniques that support reading of electronic documents (Figure 3). The first interface is a linear sequence of text and pictures, very similar to the way in which documents are traditionally displayed on paper. The second interface incorporates a fish-eye lens, distorting some parts of the document to 25% of their original size and allowing the user to expand specific sections to be read in detail. The third interface involves an overview+detail design in which the entire document is “tightly coupled” onto one pane while another pane displays a readable, detailed version of one part of the document.

Twenty users were asked to read an article using each of the three interfaces, and they were then asked to either answer questions or write an essay about what they read. Of the 20 users who evaluated the three interfaces, 19 preferred the overview+detail design. Results showed that, while use of the fish-eye interface was significantly faster, more correctly answered questions and better essays resulted from the use of the overview+detail interface. These findings may be useful in the design of the detailed levels of a case study visualization.

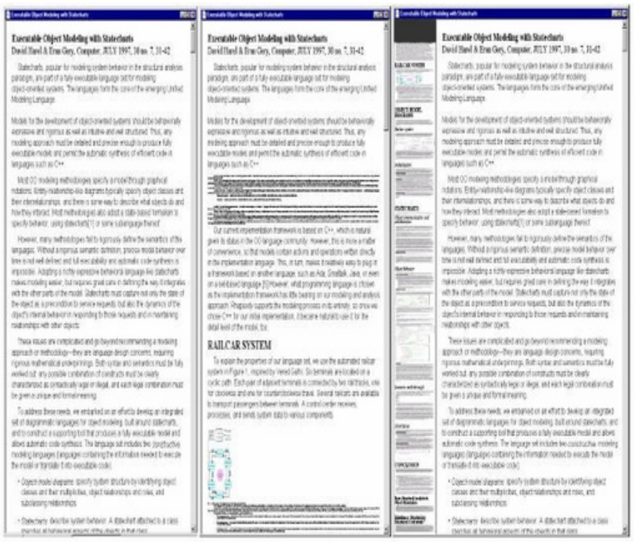

Figure 3. Hornbæk and Frøkjær's linear, fish-eye, and overview+detail interfaces.

Plaisant, Milash, Rose, Widoff, and Shneiderman [4] propose a timeline interface for the visualization of personal histories. This article states that such a visualization provides the information necessary for decision making, and to make an informed decision the user must be able to acquire the entire story, spot trends and anomalies, notice critical events, and recognize relationships. These user needs correspond to those desired in a case study visualization tool.

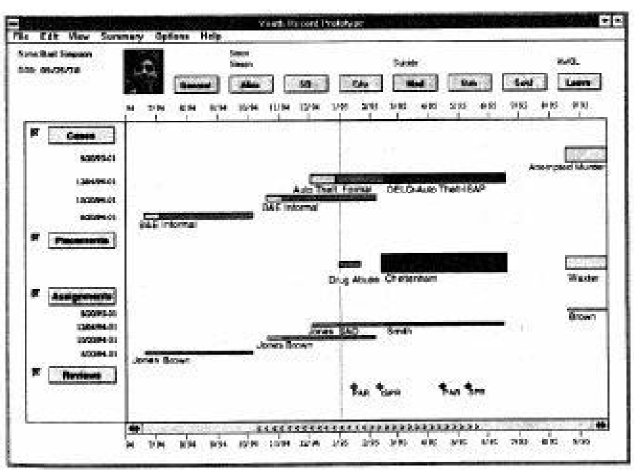

Figure 4. The LifeLines interface.

The timeline interface plots time on the x-axis against various information attributes along the y-axis (Figure 4). Icons, line color and thickness, and alternating background colors express information about case specific events. The interface opens to a one-screen overview of the case and allows one-click access to further details through zooming and detail window techniques.

User feedback for this interface stressed the importance of the overview and the ease of viewing details. Users also suggested providing links to related cases. Finally, large screens facilitate the use of multiple windows, thus keeping the overview visible at all times while displaying details in subsequent dynamic windows. The implementation and evaluation of this timeline interface may be useful in the design of a case study visualization tool.

DESIGN CONSIDERATIONS

Before any prototype is ever created, the research questions must be addressed. We can then take the answers and use them to guide our design.

Users

A large number of people who use case studies are people performing research in an educational environment. Students read them for class and to learn by example. Professors read them to keep current with their field and further their own work. Since this is a familiar environment to the authors, we have chosen to target our userbase towards an educational setting.

Data

While a case study could conceivably contain any variety of content, we have narrowed down what we consider important elements:

1. Abstract
2. Projectvisionanddomain
3. Resourcesandconstraints
4. Milestones
5. Artifacts
6. Developmentprocess-implementations,prototypes, evaluations
7. Lessonslearned

From this list, the best candidates for visualization are numbers 4, 5, and 6. The others are useful, but are perhaps best for classifying the case study amongst a collection of case studies, an idea beyond the scope of the current state of this current project.

Milestones occurring during a development cycle are a way to organize the information based on time. Certain events were started or completed at certain times, and milestones provide these important points within a process. For example, one milestone may be the completion of the first prototype. Artifacts are the tangible work produced from the process.

This includes write-ups on design, prototypes, code, evaluation forms, evaluation data, etc. Development process allows for the organization of the data into categories. This tells the user what was done and for what purpose. Artifacts are connected to the development process in that the process produces artifacts. But the user can not be given just artifacts and expected to know what the process was.

Tasks

There are two general tasks that are important to maintain support for in our prototype. When we look at the Information Visualization mantra, the ability to have an overview and get details on demands sticks out.

Hence, we need a useful overview of the case study. This will direct a user to verify the existence of sections they are interested in and direct their attention to these sections.

Next, the interface should provide efficient access to details. Any artifact the user wishes to view should be easily accessible from the overview with a minimum of interaction. This also integrates the zoom and filter approach to allow for access to the details.

Insights

Our new design should be able to provide all of the insights possible from the standard techniques. These include answering such questions as what occurred, what the design decisions were and why, how the prototypes worked, how the evaluation results led to design changes, and so on. Anything that helps you understand the process that occurred to develop the system is an insight.

INTERFACES

Before we get to the experiment, it is prudent to explain what each interface studied was. We first chose two common interfaces in current use. Note that every interface was given an arbitrary letter label for the purposes of testing.

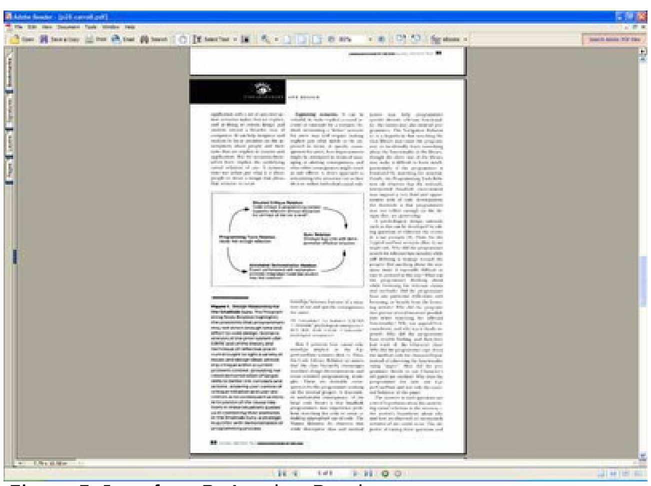

Figure 5. Interface B, Acrobat Reader

First, we chose a standard textual document. Many case studies are presented in standard textual forms. Since the ACM uses PDF files, we chose Acrobat Reader and a PDF file. This is labeled interface B and is pictured in Figure 5.

Next, we created a webpage modeled after the Usability Case Library (UCS) [7]. This organizes the case study information into scenario-based design [5] sections and forces a hierarchy on the data. At the lowest level, users can access artifacts and some details on the project described in the case study. This is labeled interface C and is pictured in Figure 6.

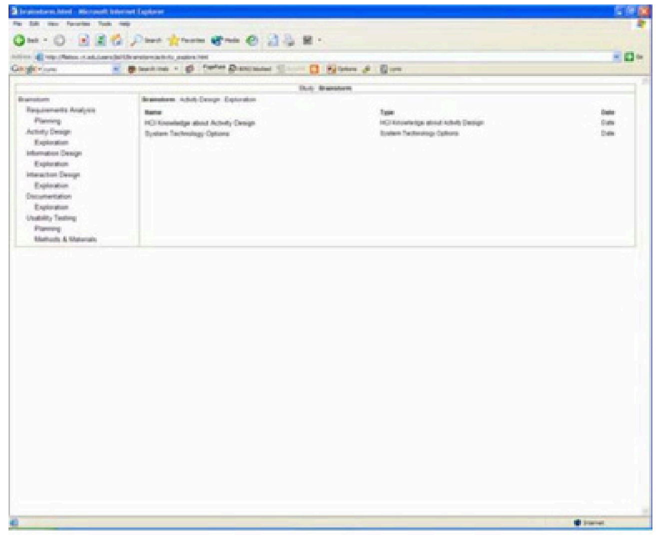

Figure 6. Interface C, webpage (modeled after UCS)

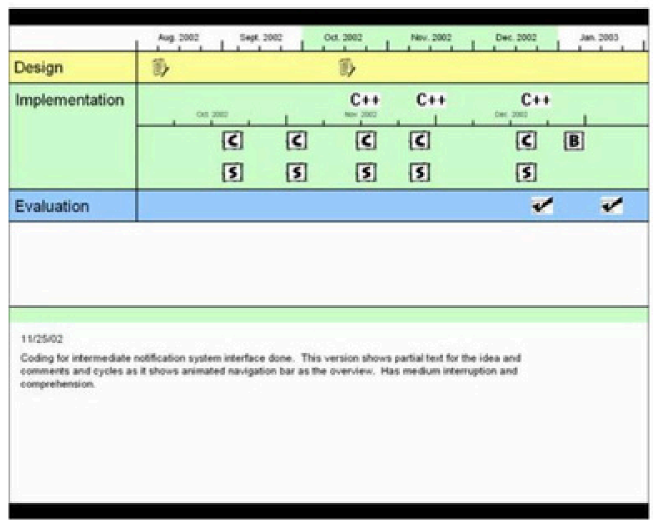

Figure 7. Interface A, prototype in PowerPoint

Finally, we have our own prototype created in PowerPoint. Here, there are three main categories: Design, Implementation, and Evaluation. A timeline stretches across the top of the interface. Icons representing events are placed within the relevant category and at the correct place on the timeline. When a user clicks on an icon, a detailed view of what it represents appears right below the icon. Subsequently, the user sees more icons that represent artifacts within a zoomed-in timeline section just for that selected event. Clicking on these new icons brings up the details in the lower portion of the display. The details range from design documents to prototype screenshots to evaluation materials, and anything else the case study contains.

Notice this prototype starts with an overview process, allows the user to zoom and filter unto a specific event, then provides details on demand. This prototype is labeled interface A and is show in Figure 7.

Each interface contained the same basic data taken from a single case study of an interface. Interface B had the research paper written about the interface, while interfaces A and C used data taken from that paper and the author's knowledge. The actual data was not crucial, as the experiment was testing the interface usability, so it's important that the data did not influence the results.

EXPERIMENTAL DESIGN Hypothesis

The null hypothesis is that there is no difference between any of the three interfaces with their ability to adhere to the heuristics and guidelines we are evaluating with. Our intent was to investigate whether this hypothesis could be disproved through our user testing.

Experiment

We wanted to compare the usability of the three interfaces. Heuristic evaluations [8] provide a useful method to test for adherence to general usability guidelines and to find major problems with the interface. As mentioned earlier, the new interface is only a basic prototype, so precise task evaluation and task timing is not useful for this current experiment.

Therefore, our independent variable is the interface being evaluated. The dependent variable is then the responses to the form we provide based on heuristic evaluation.

We chose a relevant subset of Nielsen's heuristics [8]. Then we created a set of four guidelines that were specifically related to this project and the tasks we wanted to support. These heuristics (for details, see [8]) and guidelines now follow:

1.1 Visibilityofsystemstatus
1.2 Matchbetweensystemandtherealworld
1.3 Usercontrolandfreedom
1.4 Consistencyandstandards
1.5 Recognitionratherthanrecall
1.6 Flexibilityandefficiencyofuse
1.7 Aestheticandminimalistdesign
2.1 Overview:Iwasabletoachieveanoverall understanding of the events that occurred.
2.2 Navigation:Iunderstoodhowtonavigatethecase study using the interface.
2.3 Details:IcanaccessanydetailsthatIwantinan efficient manner.
2.4 Organization:Iwasabletoeasilyunderstandhow the information was organized.

A form was created to collect the desired information. Each user was allowed to use an interface for as long as they desire, and ask questions to the evaluators. They then filled out the form for that interface. For each heuristic or guideline, they rated their response to the statement "The interface follows this guideline" with the choices being among {Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree}. Those response map to integer values {-2,-1,0,1,2} respectively for the purposes of data analysis on quantitative data. This data points to where the interface met certain guidelines and where it failed at meeting certain guidelines.

Next, the user describes, in words, why the chose their response by pointing out specific elements of the interface. This allows for the collection of qualitative data. These comments are usually very useful in learning precisely what is usable and not usable with a system and points the way toward possible improvements.

Each participant repeats this process for each interface as they all evaluate every interface.

For each participant, an order of the interfaces was given. This was different for each new participant and ensured that very few people would evaluate the interfaces in the same exact order. This minimizes any learning curve as a factor in the experiment.

RESULTS AND ANALYSIS

Once the testing phase of the project was completed, the results were compiled and analyzed. To be able to analyze all three interfaces, we compared the responses to each question, calculated correlations, and performed ANOVAs.

The first step was to take the answers for the same question in each interface and compare the results. The total number for each answer to a question was taken from the same question on all interfaces and graphed. This process was repeated for all the questions. The comparisons shown in each graph revealed that, in general, the answers for all three interfaces on the same question tended to be highly similar. This similarity was common among many questions.

In order to verify the similarities among the answers for each question, we calculated how well the results correlated. For a question we took two interfaces to calculate the correlation. For example, for the first question we produced a correlation for the PowerPoint and PDF, PowerPoint and webpage, and the PDF and webpage interfaces.

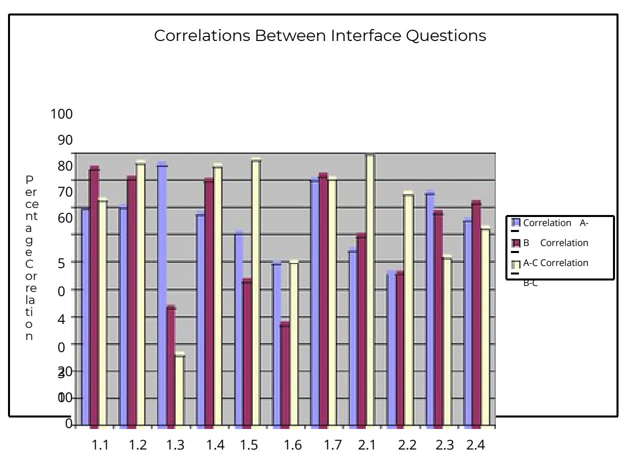

Figure 8. Correlation percentages for all correlations produced

Of the 33 correlations produced, most of the comparisons had data that had a 70% or more correlation (Figure 8). This confirmed our initial analysis in which we had concluded that many of the answers given for all three interfaces were highly similar. Ten correlations were found to be below 70%.

The top three correlations were found when comparing the answers to questions 2.1, 1.5, and 1.2. Question 2.1 asked users to comment on whether they were able to achieve an overall understanding of events that occurred. When comparing the PDF file and the Webpage, the correlation in the data for this question was 99%. Of all the seventeen users, eight users agreed (Figure 9) when answering the question for interface both B and C. Eight and seven users strongly agreed for interfaces B and C, respectively. It is apparent that both interfaces gave users a very good and similar amount of understanding of the events occurring in the design process. Interface A had less agrees and strongly agrees, and more neutrals.

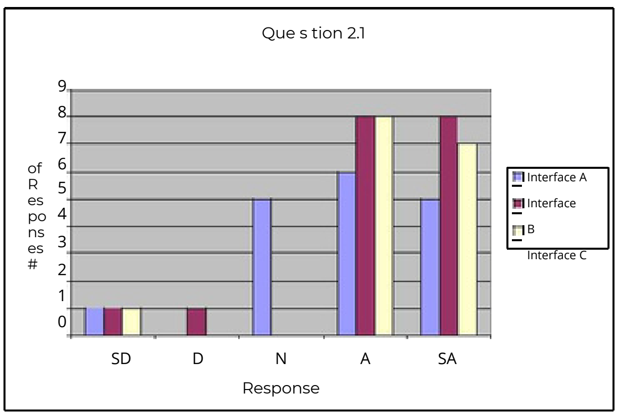

Figure 9. Results for question 2.1

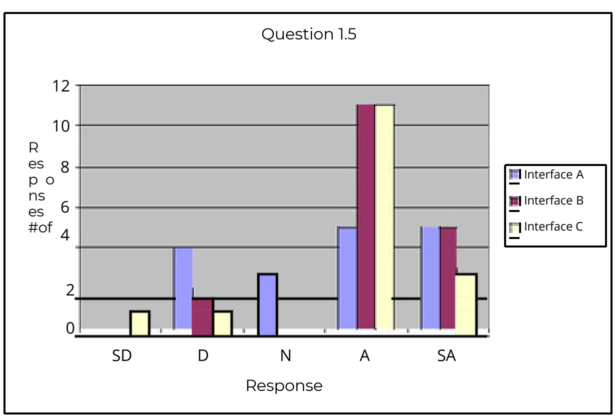

Figure 10. Results for question 1.5

Question 1.5 dealt with whether objects, actions, and options in the interface were visible. When comparing interface B to C, the answers had a 96.9% correlation. For both interfaces B and C, eleven users stated they agreed (Figure 10). Five users said they strongly agreed for interface B while only three strongly agreed for interface C. Once again, a similar pattern between interface B and C is evident. Interface A did not perform as well since it had four disagrees even though it also got five agrees and strongly agrees.

The last of the top three was found in question 1.2, which asked users to comment on whether the system spoke the users' language with words, phrases, and concepts familiar to users. With a 95.9% correlation for interface B and C, most users indicated that they agreed with the statement asked (Figure 11). Interface A did have eight participants agree, but also had four participants that disagree.

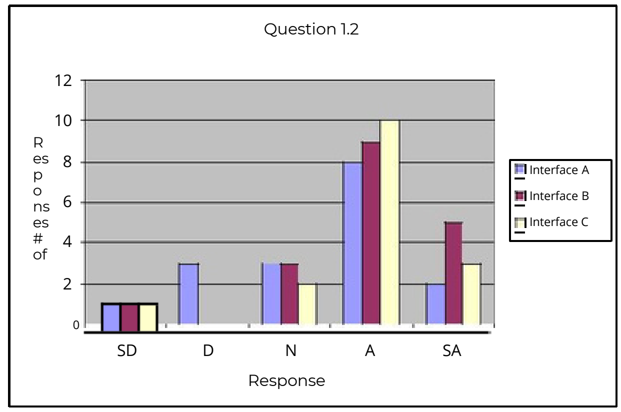

Figure 11. Results for question 1.2

Note that the top three correlations were only found when comparing interfaces B and C. In total, B and C produced five correlations with data that had a 90% or more correlation. Looking at the correlations for A and B, we see that there is only one correlation that is 90% or above. Among the correlations for A and C, there are three. This suggests that interfaces B and C are more similar due to the fact that both are familiar interfaces. Interface A is not a common interface and is slightly different than the other two. If users understood how the interface worked, they usually agreed or strongly agreed with questions. On the other hand, users, who could not understand how a certain part of the interface worked, disagreed more often. This case is apparent in the results of question 1.5 (Figure 10) where we see that there are quite a few agrees and disagrees at the same time.

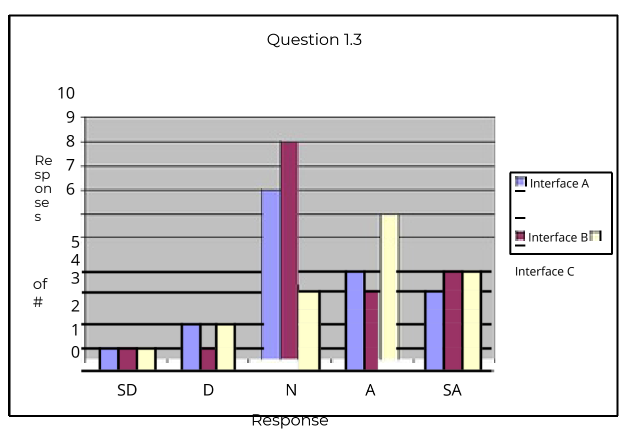

Figure 12. Results for question 1.3

Of the three worst correlations, two happen to involve interface A. When comparing interfaces A and C, question 1.3, regarding support for undo and redo, produced a 42% correlation. While many participants agreed or strongly agreed (Figure 12) with the existence of such support in interface A, the majority were neutral. Interface C, however, had a majority of users agree or strongly agree.

The worst correlation, at 25%, was also found in this same question when comparing interfaces B and C. In general, it seems that interface C had the most undo and redo support. It is interesting to note, however, that many users stayed neutral, implying that they may have not found any support at all in the interfaces.

A 36% correlation was found, once again, when comparing interfaces A and C for question 1.6. Participants felt that accelerators weren't found in interface A, leading to a majority of them to disagree or remain neutral. Interface C, however, had five participants stating that they agreed with the availability of accelerators (Figure 13).

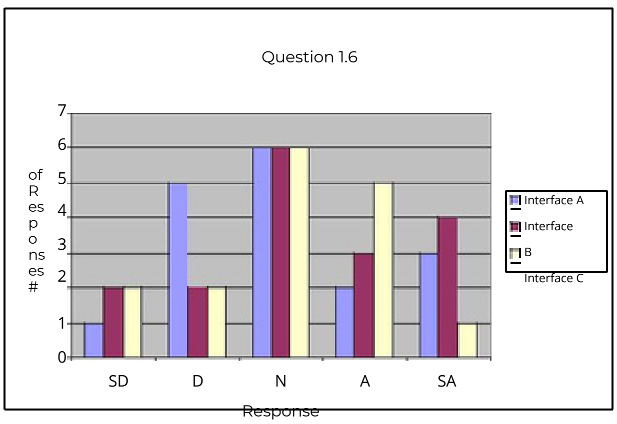

Figure 13. Results for question 1.6

Apart from the correlations, a statistical analysis was performed by using the single faction ANOVA. This was done in order to verify whether our results would show any significant difference between the interfaces. For each question, four ANOVAs were performed. They were done while comparing interfaces A and B, A and C, B and C, and all three together. For all our calculations, *alpha* was set to 0.05.

Once all 44 ANOVAs were performed, we looked at each value for *p*. A value that is less than 0.05 for any of the ANOVAs indicates a significant difference within the data. Each *p* value we looked at, however, happened to be over 0.05 (Figure 14).

The lowest value for p was 0.55. This was found in question 1.2 when comparing interfaces A and B. Interestingly, the ANOVA performed on interfaces A and C for question 1.6 yielded a value of 1.0 even though the correlation for A and C was 36%, as mentioned above.

Such high values for p indicate that we could not find any significant difference between the three interfaces. This result, however, does not mean that all three interfaces are exactly the same. All we can infer is that our method of experimentation did not make the differences apparent. The minor differences we did find from the correlations are not big enough distinctions and thus, do not allow for the discovery of significant differences.

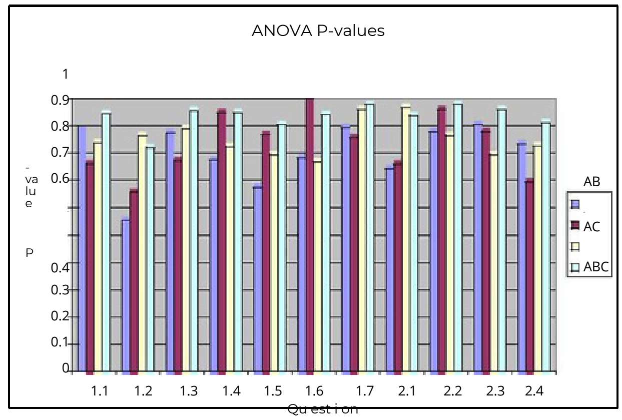

Figure 14. ANOVA P-values

Finally, in our process of data analysis, we looked at the user comments and overall results for each interface separately. For the PowerPoint interface, users indicated that it presented only the essential information and that it allowed for easy understanding of the development process. It did not provide detailed information, but information was easy to access once the goal was determined. Users said they did not know what the icons meant. While this interface did have a number of positive comments, it did have more negative comments than the other two. This is directly reflected in the results for all the questions that were answered for this interface. Compared to the other two interfaces, this interface had the highest number of disagrees (Figure 15). Once again, the issue of an unfamiliar interface is brought up again.

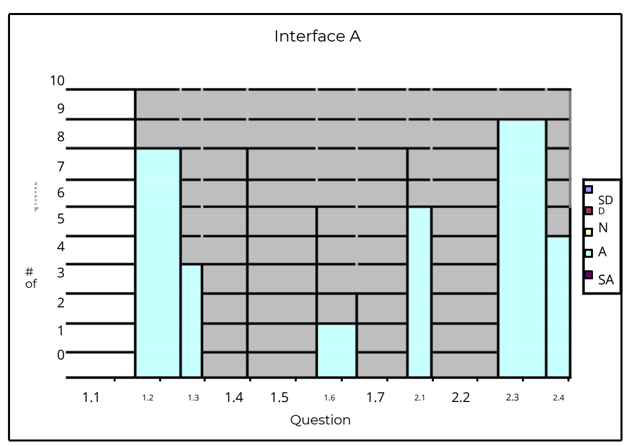

Figure 15. Results for interface A

Interface B, the PDF file, generated a lot of positive comments. They mentioned that it was like reading a newspaper and that it could tailor to a variety of users. Users indicated that reading the paper would give a very detailed idea of what occurred during the development process and that one can understand the paper better than the interfaces. Details, however, were hard to find in the paper. They tried to use the section names and subtitles to try to narrow down where a certain detail may be. This interface did have less disagrees (Figure 16) and had ten more strongly agrees than interface A.

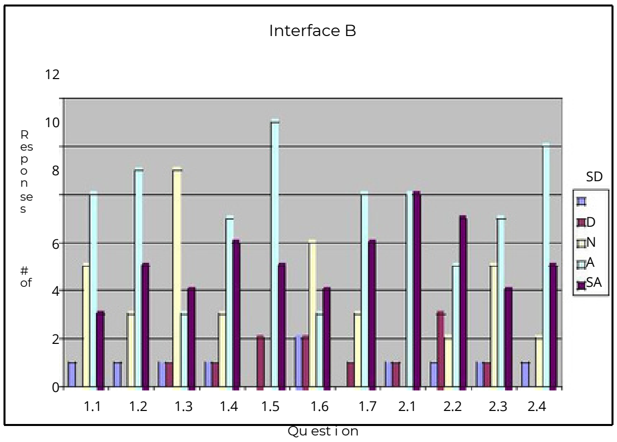

Figure 16. Results for interface B

Interface C, the webpage, was a familiar interface that allowed users to have a categorized view of all the information. They stated that it was hard to generate an error and that it was simpler to use than the PowerPoint interface. They also mentioned that the webpage did not provide enough details. Overall, this interface had the lowest number of disagrees (Figure 17). With the least number of neutrals, it had the highest number of agrees and second highest number of strongly agrees.

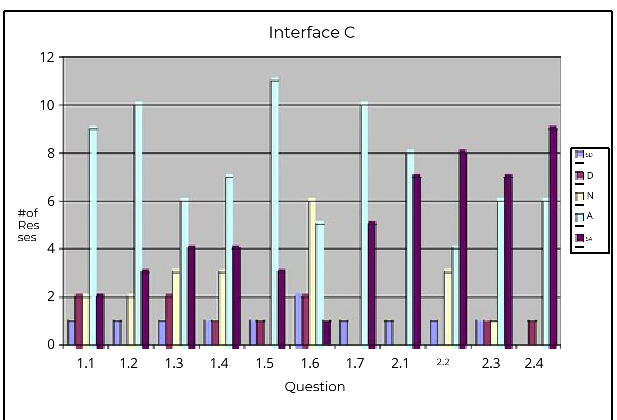

Figure 17. Results for interface C

CONCLUSIONS

The analysis of all the data has shown that all three interfaces have produced similar results. Although some differences were found, the differences were not significant enough. Since no significant difference was found, we can not reject the null hypothesis at this point.

While our own prototype, interface A, apparently did not perform better than the webpage or PDF file, The ANOVAs indicated that it certainly did not perform worse than the two. The webpage and paper are two widely common and accepted forms of an interface. To be able to