Visual Application for the Analysis of Web-Based Information Systems Usage: A Preliminary Usability Evaluation

Beatriz Sousa Santos, Florin Zamfir, Carlos Ferreira, Óscar Mealha, José Nunes, IEETA / DET - UA, IEETA - UA, CIO / DEGEI – UA, DeCA - UA, DeCA – UA {bss@ieeta.pt, florin@ieeta.pt, carlosf@egi.ua.pt, oem@ca.ua.pt, jnunes@ca.ua.pt }

Abstract

This paper presents a general description of the methods used in the on-going evaluation of a Visualizer, which is a sub-component of the Web Log Visual Analysis System. We are trying to evaluate some aspects of both the user interface and visualization techniques implemented as part of the prototype. Observation and querying techniques were used with two types of users. A general description of those users and methods used in this evaluation is presented. Preliminary results were encouraging and provided new ideas and information that will, eventually, allow a more complete and formal evaluation of our application.

Keywords: Usability Evaluation, Information Visualization.

1. Introduction

Although there are a lot of information visualization techniques (good examples are included in [1][2]), and many systems that use them to visualize large amounts of information, there are comparatively few studies on the evaluation of those techniques and systems. This is perhaps due to the inherent complexity of this evaluation. However, now that the field of Information Visualization has matured, a wealth of techniques has been developed, and is applied to solve real problems, it is really important to know if these techniques actually work. Furthermore, there is a need to know under what circumstances they should be used, how they compare and what tasks they best serve.

While there is not yet a body of knowledge on information visualization evaluation, we can find in literature some works explicitly using evaluation in the process of designing a visualization system, evaluating specific systems and visualization techniques, as well as comparing alternative visualizations. Examples can be found in [3][4][5][6][7][8]. Moreover, there are also some authors recognizing the importance and making an effort to develop more systematic approaches to the complex problem of evaluation in information visualization [9][10][11]. However interesting these works may be (and we believe that they indeed are), those who want to evaluate visualization techniques and systems, still struggle with a lack of specific techniques and methodologies to conduct the evaluation. Currently, it seems that a reasonable approach could be to adapt the well known and already widely used methods of Usability Engineering [12], taking into account the specificities of the techniques and systems that one is trying to evaluate.

In this paper we describe an on-going evaluation intended to contribute to the development of a visualization based application supposed to help information system managers to understand how the information system of their organization is being used [13]. We have taken into account the work of Freitas et al. [9] to structure our evaluation and we also adapted some usability testing techniques to obtain feedback from users and use that information to redesign our application. In this preliminary stage of the evaluation we decided to use non-experimental techniques as a means of suggesting, clarifying, refining and generating ideas that can be further explored in more controlled conditions [14].

2. Overview of the application

As mentioned above, our aim is to help information system managers of organizations to understand how their internal information systems are being used. Our proposal is related to the visualization of large quantities of information collected inside an institution mainly from the analysis of the web site structure and usage logged
information (obtained either during natural site usage or controlled experiments). Some interesting questions, whose answers might provide insight, can be:
- How is the site used?
- Who is using the site?
- What are the users’ interests?
- What statistical information can be obtained from the log files?
- Which are the areas with problems?
- What usability problems can be identified?

We are trying to answer some of these questions, by providing a visualization application able to capture, compile and present the information related to the structure of a web site and the usage patterns (more details can be found in [13][15]).

Some visualization schemes, which can be used interactively by the user, have been developed in order to represent information in an understandable way. We were able to identify visualization schemes that, in principle, could be useful in the context of our application to support visual inspection of the structure of the site and the user interface design coherence, as well as, the location and usage statistics of page hypermedia connections. These visualization schemes are complementary, each allowing a different way of representing the information that should make it more adequate to serve certain user goals; thus, they should be used in combination with each other and not used one at a time. In order to allow the user to profit from the advantages of these visualization schemes we offer the possibility to choose:
- how many visualization schemes he/she wishes to view simultaneously;
- what multiple views pattern he/she wishes to use;
- where on the screen is the representation obtained using each scheme going to be displayed.

Different views simultaneously displayed to the user are synchronized, i.e. any interaction with the objects of any view is reflected on the other views (e.g., an object selection will be interpreted by all the other views resulting in a change of content or aspect). This synchronization helps the user to explore and understand the information. An example of a screenshot with four different visualization schemes being used is shown in figure 1.

3. Evaluation Method

During the development of the visualization schemes and user interface we have conducted informal evaluation sessions, using simple prototypes, with several people. These people were mainly volunteer Computer Engineering students, which in general had some experience as web developers. These sessions were very useful and allowed us to discard some ideas and refine others. Two members of the team, the ones less involved with prototype implementation, have also made some heuristic evaluations of the user interface features and some informal evaluation of the visualization schemes. After this first evaluation cycle we have developed a more sophisticated prototype including six different visualization schemes and the mechanisms to use them synchronously. Then, we started a second evaluation cycle. In this cycle we have adopted a different, more structured, approach that is briefly presented in the following sections.

The second cycle of evaluation had two major goals: to evaluate the user interface main aspects and to evaluate the visualization schemes. Visualization schemes include the visualization representations and the interaction mechanisms provided to users so that they can interact with data through the visual representation. According to [9], there are two different sets of evaluation criteria for each of these two latter aspects: to evaluate visual representations we can use cognitive complexity, spatial organization, information coding and state transition, and to evaluate interaction mechanisms we can use orientation and help, navigation and interrogation and data set reduction.

We performed several evaluation sessions, with two types of users, using mainly observation and query based techniques to evaluate both the user interface features and some aspects of the visualization schemes. These techniques are widely used in usability evaluation of user interfaces but they have also been considered appropriate to evaluate some aspects of visualizations [17]. The collected data were analyzed using Exploratory Data Analysis.

3.1. Users and Observers

In order to perform the two kinds of evaluation, i.e., user interface and visualization schemes, we have asked for the collaboration of two different types of users: thirty-two Computer Engineering students, currently attending an
introduction course on Human-Computer Interaction, and five professionals that work at the Centre of Informatics and Communications of our University (CICUA). These professionals have several years of experience as web developers, web/network managers and three of them have also attended the same course on Human-Computer Interaction, during their graduate studies. The profile of these professionals made them not only representative of our target users, but also capable of understanding well what kind of feedback we would need from them, in the scope of this evaluation. While the students are not, in general, web developers/ managers, they have a profile that makes them also reasonably suitable as subjects for the evaluation of our user interface, since all of them have experience as web users and computer programmers. Therefore, we have conducted an evaluation more focused on the user interface with the help of the students and an evaluation of the visualization schemes, as well, with the help of the five professionals.

Since the students had been practicing user interface evaluation through different methods, we have decided to profit from their capabilities and ask them to act as observers as well as users. This procedure would allow us to obtain observation data from a larger number of users, provided that we would ask the observers to register simple enough information (since they are not very experienced). We considered this would provide an interesting practice for the students and that was also a motivation to have all the students act both as users and as observers.

Therefore, while half of the students would perform some predefined tasks, the other half would observe them and register times, task completeness, as well as other relevant information. After some predetermined time they would change roles. Obviously, the students that would act first as observers and later as users would have a greater acquaintance with the interface than the others, a different level of awareness, and could be considered as more experienced users. Hence, for the purpose of data analysis, the students were divided in two classes of users: less experienced users and slightly more experienced ones. We should notice, however, that in spite of the fact that we have two groups of users, the performed evaluation is non-experimental, in the sense that there is no control group, nor has been defined any hypothesis. Its main purpose was to gather ideas that can be further explored in later stages of this evaluation.

The basic demographic data of each participating user was collected before the tasks, through a questionnaire, including also some questions meant to assess their experience with information visualization applications and as web developers or managers. Analysing the collected data we found that they were between 19 and 31 years old (median value=21 and two outliers aged 30 and 31), 3 females and 29 males, having no difficulties in colour perception. Moreover, concerning their experience as web designing and evaluation most of them were able to produce web pages of moderate complexity and were acquainted with the basic methods of evaluation. Finally, concerning experience with applications using 2D visualization, most of them use frequently several packages (e.g. Macromedia Suite, Adobe Suite and MatLab).

After completing the tasks, a post-task questionnaire was given to the users aiming the assessment of their satisfaction and opinion on several issues.

3.2. Database

An internal site containing information (such as program, studying material, practical assignments, etc.) corresponding to the Human-Computer Interaction course, offered to approximately forty Computer Engineering students, in 2002/03 at the Department of Electronics and Telecommunications of the University of Aveiro, was developed specifically to collect usage data. The database used contained the information collected during the whole semester both in normal usage and in controlled sessions.

3.3. Equipment

The evaluation sessions were performed in a laboratory classroom equipped with PC computers running the prototype and an SQL Server for the Database.

3.4. Tasks

We have defined a set of tasks for the users to perform during the evaluation sessions that were relatively simple, nevertheless regarded as representative of typical operations end-users will perform with the visualizations and the user interface.

Keeping tasks simple makes it easier to analyse user performance; however tasks should not be so simple that their ecological relevance is unclear (i.e., we have to ask how frequently do those tasks actually occur in real-world tasks, and how significant are they in the overall task solution process).

Each user had to complete ten tasks within a given time window. The tasks were related both to the evaluation of the user interface and the visualization schemes.

The following are some of the tasks:

- Manipulate and navigate among the representation windows;
- Use a menu option or tool button to obtain a given functionality;
- Select a given site;
- Select a given session;
- Find and select a given page of the site;
- Find how many times a link was followed between two given pages;
- Show possible paths between two given pages;
- Find the number of pages corresponding to the shortest path between two given pages;
- Find how many external jumps has a user performed during a session.

The first four tasks of the previous list, are very simple and directed to the evaluation of some user interface features (change viewing conditions or use functionality through buttons or menu options); the other tasks are combined with some of the previous ones in more complex tasks focused on evaluating the performance of the user using the visualization schemes to extract some qualitative or quantitative information through interaction with the data. “Find and select a given page of the site” and “Find how many times a link was followed between two given pages” are tasks oriented to evaluate the interrogation features of the visualization scheme. “Find the number of pages corresponding to the shortest path between two given pages” and “Show possible paths between two given pages” are intended to evaluate the data set reduction feature, according to the evaluation criteria proposed in [9] to evaluate interaction mechanism of visualization schemes.

3.5. Procedure used with the students

A pilot evaluation was conducted with three students (that didn’t participate in any of the sessions mentioned bellow), in order to assess the difficulty of the tasks, their duration and the clarity of the questions. As a result, some modifications had to be made both in the tasks and questions.

![Figure 2 Roles of the students](image)

Four evaluation sessions were performed with the students; one (TS1) with eighteen students and another (TS2) with fourteen students. After these two test sessions, we have performed another two sessions with the same students (TS3 and TS4). Thus, each student participated in two sessions of 1h 45 m. The thirty-two students were organized, during each session, in groups of two students named User#1 and User#2; User#1 would act first as user and latter as evaluator; and User#2 would do the opposite. These roles were attributed randomly. Figure 2 shows the sequence of roles performed by the students.

Before asking the participants to perform the tasks, they were given an overall description of the application and of the visualization schemes, as well as, some details of the user interface. Then, they were debriefed about the evaluation: the tasks they were supposed to perform and observe, the questionnaires they were supposed to answer, and the used scales. In the first session with each set of students, this explanation and question answering lasted approximately 45m; after this period of time, every one admitted they had understood what they were supposed to do and were willing to participate. Subsequently, participants were asked to start the test and perform or observe the tasks during 30 minutes, then change roles.

After sessions TS1 and TS2 we got the impression that some of the students had not fully understood the tasks; thus we decided to perform a second round of sessions (TS3 and TS4) two weeks later, with the same students and following the same organization. This time we didn’t explain the application, we concentrated on the explanation of the tasks and let them practice for 10 minutes, before starting the tasks. This was considered a reasonable approach since our intended users are experienced people. Consequently, in this second round of sessions the difference in experience between the two groups of users, Users #1 and #2, was expected to decrease, given that they all had already participated in one evaluation session, two weeks before.

3.6. Procedure used with the professionals

A different, generally less structured, approach was used for the evaluation performed with the five professionals. In an initial three hour session, with all of them, a more detailed presentation was given about the application and the visualization schemes. Some details of the user interface were also shown. Then, a period of questions and answers was allowed so that the participants could better understand the overall purpose of the application and the type of evaluation we intended to perform. After this, they have installed the prototype on their own computers and were given a period of time to use the prototype on their own.

During this session we have asked users to think how they would use our application in their everyday work and what kind of other people would also profit from the application; then we invited them to describe this through simple scenarios.

A week later, in 90 minute individual sessions, each user was asked to perform a set of tasks thinking aloud, being observed by two members of the developing team. These tasks were basically similar to the tasks given to the students plus some others mainly meant to evaluate how they use the visualization schemes.
3.7. Collected data

In the evaluation sessions performed with the students, each observer student had to register the following data concerning the performance of the user:
- time spent performing the task;
- if the task was completed;
- the answer to the question (for some of the tasks);
- user satisfaction;
- any observations the observer considered relevant.

After completing the entire set of tasks, the users were asked to fill a questionnaire giving their opinion on the visualization schemes and some interface features. Concerning each visualization scheme, the following questions were asked:
- is it familiar and intuitive?
- is it easy to use?
- is it easy to learn?
- does it give adequate feedback?
- does it use an adequate colour coding?
- does it provide adequate interaction mechanisms?

Some of these questions were also asked about the icons used to offer the functionality. All these questions (including satisfaction corresponding to each task) were answered using a qualitative scale:

| 1 | 2 | 3 | 4 | 5 | N |

Where 1 is complete disagreement, 5 is complete agreement and N corresponds to not having an opinion or not wanting to answer. Opportunity to give suggestions or make any comments was given through the inclusion of an open question, at the end of the questionnaire.

Finally, we collected additional opinions and suggestions in informal conversation.

4. Results

In this section we describe the results obtained with the students, which are more quantitative and more focused on the user interface, as well as the results obtained with the professionals, which are more qualitative and focused on the evaluation of the visualization schemes and overall interest of the application.

4.1. Results obtained with the students

As mentioned before, after the first round of sessions (TS1 and TS2) we had the impression that some of the students had not fully understood what they were supposed to do; thus we simplified the fill in forms for the observers as well as the tasks wording. Then, we decided to analyse, as a first approach, the following items collected only during the evaluation sessions TS3 and TS4:
- the time spent in each task;
- if the task was completed;
- if the answer was correct (in some of the tasks);
- the user satisfaction.

Figure 3 shows the box plots obtained using Statistica [16], for the times corresponding to all tasks performed by Users #1 and #2. The median values for the times are 60s, and 32s, respectively; using a Wilcox test we have found the difference between these values significant (p<0.00001). While Users #1, as well as Users #2, had about the same amount of experience with the application before the beginning of this round of evaluation sessions, and we were not expecting to observe a significant difference in performing the tasks, there was a difference in the median time. This seems to mean that Users #2 have learned how to perform the tasks faster just by observing their colleagues. However, no significant difference on the number of correct answers was observed between the two types of users.

As to satisfaction, the median values for Users #1 and #2 are respectively 3 and 4. Analysing user satisfaction and time, task by task (figure 4), we have informally noticed that the users are less satisfied when they take more time to complete the tasks. This result was confirmed as significant in tasks 2, 4, 5, 6, 7, according to the Spearman correlation coefficient.

![Figure 3 Boxplots corresponding to times in all tasks for Users#1 and #2](image)

As to satisfaction, the median values for Users #1 and #2 are respectively 3 and 4. Analysing user satisfaction and time, task by task (figure 4), we have informally noticed that the users are less satisfied when they take more time to complete the tasks. This result was confirmed as significant in tasks 2, 4, 5, 6, 7, according to the Spearman correlation coefficient.

![Table 1 N. of users that completed or didn’t do tasks without any question](image)

<table>
<thead>
<tr>
<th>Task</th>
<th>Completed</th>
<th>Didn’t do</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>3</td>
</tr>
</tbody>
</table>

![Table 2 N. of users that completed correctly, incorrectly or didn’t do tasks having a question](image)

<table>
<thead>
<tr>
<th>Task</th>
<th>Correctly</th>
<th>Incorrectly</th>
<th>Didn’t do</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>24</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>
Table 1 shows, for each task that hadn’t any question to answer, the number of users that completed or didn’t complete the task. This table shows that the great majority of the users were able to complete the tasks. In fact, 97% of the tasks were completed. While for these tasks we cannot know if the users have completed them correctly, for the tasks that have a question to answer we have this information. Table 2 shows, for each of these tasks, the number of users that completed correctly, completed incorrectly and didn’t complete each task. Counting the total number of correct answers to the questions, we have found that a high percentage of users have been able to find the correct information through the visualization schemes: 70 percent of the tasks having a question were completed correctly.

In task 3 and 7 users were supposed to find how many times a given link had been followed, using two different visualization schemes (SiteMap2D and Page Explorer2D). From the 32 users, 24 have completed correctly task 3, 4 users have completed it but obtaining a wrong answer and 4 users didn’t perform the task. As to task 7, 17 users have completed correctly the task, 12 have completed it but obtained a wrong answer and 3 didn’t do the task. This difference in user performance was confirmed as statistically significant using the non-parametric sign test. This suggests that SiteMap2D could support better this kind of task than PageExplorer2D. It is interesting to note that, while the users show a better performance in task 3, their satisfaction is higher in task 7. This result confirms the notion that a higher user satisfaction does not necessarily mean a better performance. This could be related to the total time to complete the task, which is lower for task 7 (median time=30s) than for task 3 (median time=60s).

In task 5 users had to find the number of pages corresponding to the shortest path between two given pages. For this task they had to select the pages using SiteMap2D and then observe the result using another visualization scheme displayed on another synchronised window. From the 32 users, 26 were able to obtain the correct number of pages, 5 obtained an incorrect number (which means that they probably were not able to select the right pages on SiteMap2D), and just 1 user was not able to perform the task. Moreover, the median time to complete it is the 2nd lowest time (median=30s) among all tasks. This seems to suggest that using these two synchronised visualization schemes is reasonably obvious to the users.

Task 9 implied using another visualization scheme (SessionMap2D) in order to find the number of jumps to external pages. In this task, 29 of the 32 users were able to complete the task and obtain the correct number of jumps; only 3 users didn’t complete it. This seems to suggest that SessionMap2D supports adequately this task.

Task 10 was much more difficult, intentionally devised to see if the students would be able to perform a complex task using the application. The median time was 268s and the median satisfaction was 2 (the lowest value of all the tasks). Even so, 16 of the 32 users were able to do it correctly obtaining the right answer. 2 completed it but obtained a wrong answer and 14 didn’t do it. Whereas it was necessary to give a hint on how to perform the task to some of the students, we were expecting worse results in this task.

Figure 4 Boxplots corresponding to times and satisfaction task by task

Even if most of the tasks were relatively simple, these results are encouraging since the students were not experienced users and were able to perform correctly a high percentage of tasks. Also, the median value of the overall satisfaction was 3 for Users #1 and 4 for Users #2.

4.2. Results obtained with the professionals

During the sessions performed until now with these users, we have already collected a lot of interesting low level feedback concerning specific features of the user interface, and visualizing schemes included in the prototype, but also more general, high level information related to the interest and usefulness of the application.

Concerning the user interface, some icons, the colour scales and some dialog boxes, were considered as the weakest points. According to some users the synchronized representations of the same data is the most interesting
feature of the application; however it makes it more complex.

These users generally considered the application as having a great potential and were interested in using it as soon as we could produce a consolidated first version.

5. Conclusions

In this paper we describe the on-going evaluation of a visualization application under development, intended to allow web designers and managers to get some understanding of the usage of their web sites. While we started by an informal evaluation and now we are performing a more structured one, this stage only uses non-experimental methods in order to elicit new ideas and information that will, eventually, allow a more complete and formal evaluation of our application.

We have used observation and querying techniques and asked two types of users to perform a set of tasks meant mainly to evaluate the usability of certain user interface features and of the interaction mechanisms of the visualization schemes.

While 97% of the tasks that hadn’t any question to answer were completed, 70% of the tasks that had a question to answer were correctly completed. Furthermore, the median times for completing the task range from 10s to 60s (except for task 10), values which seem reasonable, considering the type of task and the experience of the users. Finally, the median value of the overall users satisfaction is between moderate and high.

Though we still have to analyse collected data concerning the users’ opinion on the visualization schemes and some user interface features, the obtained feedback already provided a lot of ideas to improve our application. We are encouraged by the preliminary results obtained both with the students and the professionals and intend to perform some more sessions with the professionals to evaluate how the proposed evaluation schemes support them, as they become more experienced users.

We still have to evaluate the visualization schemes concerning just the visual representation; however we have not yet devised any systematic approach. Until now we have judged informally the complexity of the representation and the time it takes to rebuild it after user interaction. Perhaps a graphics designer would help to judge the adequacy of some aspects (e.g. colours, icons, spatial organization and coherence).

Acknowledgements

The authors are grateful to all participants that have graciously collaborated in this evaluation, students and professionals, as well as to Doctor Joaquim Madeira for his help during the evaluation sessions.

References