A METHOD FOR WEBSITE USABILITY EVALUATION: A COMPARATIVE ANALYSIS

Naser F. M. El-firjani, Ebitisam K. Elberkawi, Abdelsalam M. Maatuk
Faculty of Information Technology, Benghazi University, Libya.

ABSTRACT

Graphical user interfaces design in software development process focuses on maximizing usability and the user's experience, in order to make the interaction for users easy, flexible and efficient. In this paper, we propose an approach for evaluating the usability satisfaction degree of a web-based system. The proposed method has been accomplished in two phases and implemented on an airlines website as a case study. In the first phase, a website usability test is implemented by a number of users, and then the results obtained are translated into charts for a final web-based system evaluation in the second phase. The results achieved were satisfactory, since the places where the weaknesses and gaps in the website are identified and recommended solutions to avoid them are drawn. The authenticity of the results have been confirmed by comparing them with user opinions acquired from a questionnaire, which proves the precision in which the website is rated.

KEYWORDS

GUI, Web usability, Web evaluation method.

1. INTRODUCTION

The success of a product is based mainly on user satisfaction. Many products have been proved ineffective even though they met all scientific and technical design aspects. The only setback is that a lack of user satisfaction. In the development of systems, the software designers should focus on the requirements of the user so that to increase the compatibility between GUIs and other components of the system to ensure that the final product meets user requirements [9]. On other hand, the evaluation usually is carried out by users. This provides a direct incite on how actual users in a real situation use the system in practice. Thus, we have to measure the difficulty degree of systems to determine the gaps and problems that need to be resolved and involve users in such process of measurements. As a result, a method for testing the usability of systems with enhanced role of the user is necessary. The usability testing is a way of assessing the degree to which an interactive system is easy and pleasant to use with a view of identifying usability problems and/or a collection of usability measures/metrics [2, 3].

Web usability has several definitions and characteristics. It is a combination of characteristics oriented to the user, which are easiness of learning, high speed of user task performance, low user error rate, subjective user satisfaction and user retention over time [24]. Usability is a measure of the ease with which a system can be learned or used, its safety, effectiveness and efficiency, and the attitude of its users towards it [15]. The measure of the ease might be an extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, rate of errors, memorability and satisfaction in a specified context of use [25].

In this paper, we propose an approach for a web-based system testing for usability evaluation. The method, which is implemented on the United Arab Emirates (U. A. E.) Airlines website consists DOI : 10.5121/ijwest.2017.8301
of two phases. In the first phase, the system is evaluated by users regarding usability according to a predefined test plan, which involves some tasks, each of which has a time to be performed in. In the second phase, the results acquired from the first phase are converted into charts for initial results of system evaluation. As a proof of concept, we have compared the initial results achieved by our proposed method with the results obtained from a questionnaire in user’s point of view. The questionnaire has been designed based on two aspects: (1) general evaluation, and (2) interface and interaction designs of the website. The questionnaire results demonstrate the user’s dissatisfaction, which have been the same findings of the proposed method.

The rest of the paper is organized as follows. Section 2 provides a background and related work. Section 3 describes the phases of the method and a description of the testing plan. The implementation and the initial results are presented in Section 4. The validation and comparisons analysis are presented in Section 5. Section 6 concludes the paper.

2. BACKGROUND AND RELATED WORK

Usability is one of the important elements to make product usable. Websites evaluation is determining the quality of the website [5, 16]. Online booking flight is one of websites that provide services to customers, and should be enough usable. Booking flight online makes an excellent user experience in Human Computer Interaction (HCI) on the website interfaces exceedingly. Airline operators may not worry about usability issue but in situations where there are plenty of competition costumer can find a competitor’s website when the booking process become too difficult. This underscores gross importance of website usability evaluation.

An evaluation is concerned with gathering data related to usability of product using a group of users for the tasks and specifying the work environment and context [15]. Generally, two types of website evaluation exist, i.e., quantitative and qualitative evaluations. Quantitative studies focus on the website quality [14]. The measures of the web pages, such as links formatting and the text elements were introduced in [10], whereas the numerically measurable data, e.g., time-based and traffic-based data have been used in [25]. However, in qualitative studies, the indices of the website quality are evaluated without generating indices or scores. For example, the combination of branding HCI and usability could possibly be used to enhance websites evaluation [23].

An HCI combines gathering data and the intellectual’s framework of psychology, using computer tools in creating effective system interface [23]. The work of man and computers are being together understudied by the HCI and usability studies to ascertain effective interaction between human and machine. Usability, universality and usefulness are expected outcomes of the HCI in technology [24]. Thus, usability is the aspect of website application that the user interacts and gain first-hand interaction with computers [20]. Hence, usability quantitatively and qualitatively measures the design of a system interface for user interaction. The five key factors of usability are: learn ability, efficiency, memorability, errors, and satisfaction [20].

The most basic method of information acquisition on how user interacts with technology and the difficulties faced using the technology is regarded as the usability testing [20]. An approach proposed to improve usability of software applications to interacts with users with an effective manner is described in [1]. The majority of today’s computers are designed and centred around the user, as such that gives the same reason as why feedback from the users of computers cannot be replaced [12]. Usability testing helps in discovering mistakes committed by users when interacting with system’s interfaces. The selection of the users that truly represent the entire user population in accomplishing given testing tasks is needed. During a usability test, a target user population should be selected and recruited [13]. The test setting can be done through usability laboratory experiment or in a workplace. The web-based usability testing can be also used as a
remote usability test where the user and experimenter might be located remotely from each other. The time for page response during tests is one of important characters [4]. The use of animation and/or multimedia plug in requirements may affect page loading time [11].

A good user interface design typically needs the use of a variety of usability evaluation methods [6, 18]. One of such methods is the end-user think aloud protocol and the heuristic evaluation method. The heuristic evaluation method is one of the most useful method and the least expensive one [17]. However, it requires to use additional software to observe the participants from a distance. End-user think-aloud protocol method is based on asking participants to say out loudly what they are thinking about when using a website or an application software. The results obtained from this method are close to what is experienced by users, even though, the environment is not natural to the user. A usability evaluation method for web-based learning systems using a set of empirically-supported usability factors has been described in [21]. At some stage in the evaluation, the method allows for the prioritization of usability problems to be dealt with in system improvement. A combination of various elements of several usability methods are enhanced with a mechanism to evaluate the use of the system by the users and described in [7]. This is to determine the usability of a tested system by diverse metrics and practices in a single test.

In contrast to existing methods, the proposed method described in this paper depends on evaluating the usability of the systems based on two factors, i.e., time and mistakes. In fact, these two factors are the solid foundation for understanding the user reaction. For the mistakes factor, we have used two different viewpoints to obtain the most accurate results. The first viewpoint is the number of mistakes per task, whereas the time that the mistakes require to be resolved in has been the second viewpoint. Then, the time is compared with the total time required to complete the task. Hence, we can acquire the results of the evaluation process through the obstacles facing the user of the system, and this gives us the results that more accurate than the once where the two factors were not used.

3. THE PROPOSED METHOD

In our usability testing method, a summative usability testing plan has been used. The test used was a performance measurement. Summative usability testing is a comprehensive evaluation of a product with representative users and tasks designed to measure the usability (defined as effectiveness, efficiency and satisfaction) of the complete product.

Usability evaluation assesses the extent to which an interactive system is easy and pleasant to use. This technique was used to obtain the quantitative data about test participants’ performance, when they performed the tasks during usability test. The test was conducted in a formal usability laboratory and the data has been collected accurately and possible unexpected interference minimized. The users were given a pre-test training. A list of prepared tasks was provided to the users. The users were observed while the experiment was running on. The aims for the usability testing in terms of usability attribute (e.g., easy to learn, efficient to use, easy to remember, few errors, subjectively pleasing) have been defined. The various components of the aims were balanced and their relative importance were decided. Usability issues were quantified by measurements such as:

- The time that users take to complete a specific task.
- The time spent recovering from errors.
- The number of user errors.
Figure 1 shows the method usability evaluation process, which is based on two factors: Task and Time. The system to be evaluated is divided into sub-parts, each of which is called a Task, so that each system is a combination of several tasks. The Time is an important factor, which help us criticize the system. The lowest implementation time leads to a positive evaluation.

![Usability Evaluation Process Diagram](image)

Figure 1: The method usability evaluation process

The number of mistakes lead to increase in the time it takes to complete a task. Thus, the longer time taken to accomplish a task, the more likely there is a negative review from the user. Hence, the factor of time is of the utmost importance, as of which it has been the main focus. The method consists of two phases. In the first phase, the system is evaluated by users regarding usability according to a test plan. In the second phase, the data results obtained are translated into charts for system evaluation.

### 3.1 The First Phase

In this phase, the usability of the chosen web-based system is evaluated according to a predefined test plan. This test plan is implemented by a number of users, during of which we are observing them according to three usability measures according to Neilsen usability measures [20].

#### 3.1.1 The Test Plan

The purpose of the test plan, which is applied in the first phase is to identify how users navigate and perform the tasks of the system to be tested. We have used the GUI of the U. A. E. Airline website to be tested. The test covers both the navigation and the contents of the system. The test was held in a laboratory in the University of Utara at Malaysia. Laptops with good operating systems and browsers have been used in the test.

#### 3.1.2 Participant Roles

- **User**: is the person who implements the test.
- **Facilitator**: is the person who helps the users.
- **Observer**: is the person who notes the performance of the users.
3.1.3 Characteristics and Experience of Users

A pre-test questionnaire has been designed and used to collect users’ demographic data and information on prior knowledge and other interesting information about the usability evaluation. Five users have been invited for conducting the usability tests in the same time. Two of the users had some experiences on using airline reservation system but have no experience on U. A. E. Airlines website system. All the users were students from University of Utara Malaysia: One BSc student, Two MSc students, and Two PhD students.

Four out of the five users used for the testing the website were between 21-30 years old while only one user in the range of 31-40 years old. This indicates that all the users were matured age-wise, and none of them was a minor.

Two of the testers, have been using the computer for between 0-3 months before the test time, and three users have used the computer for more than 12 months before the time of the test. This is indicative of the fact that majority of testers were experienced and have good knowledge in the use of computer. All the five users were familiar with web browsing. Two users have used the web browsers between 0-3 months while three users have use web browsers for more than 3 months. This clearly shows that most of the users have good experience with web browsing and are not novices. The testers use the three common web browsers, i.e., Internet Explorer, Chrome and Firefox. Each of the users used the browser that they are familiar with. Two users used Internet Explorer, two users used Chrome, and one user used Firefox.

3.1.4 Ethical Considerations

The following ethical issues were considered:

- **Confidentiality:** users were anonymous and their identities have not be revealed, and their data used strictly for research purpose.
- **Protection:** The user were in a safe and protected place.
- **Participation or Withdrawal Option:** The users consents were sought to solicit their participation and they have been free to withdraw from the test at any time or to stop their data from being included in the analyses after the test.

3.1.5 Training and Sessions

A pre-test training is given to the participants where a list of tasks have been implemented (five tasks in the U. A. E. Airline website). The test sessions took about 90 minutes. About 30-60 minutes were reserved for scheduling participants between sessions, setting and resetting the test environment. Sessions with observer(s), are briefly reviewed while allowing users who might arrive a little late.

3.1.6 Evaluation Metrics

The following metrics were considered during the website evaluation:

- **Successful Task Completion:** In each scenario, users were required to obtain a specific data to be used in a typical task. When users indicate that they have found the answer or completed the task goal, the scenario is considered successfully completed.
- **Mistake:** Deviations that are observed at the completion of a scenario that vary from the targets expected of the scenario were noted.
- **Time on Task:** The time that each user takes to complete each task has been noted.

The results of this phase are assembled as shown in Table 1. The usability measures are related to the factors of tasks and times, as follows:
Table 1: General Design of Testing Table

<table>
<thead>
<tr>
<th>Task No</th>
<th>Task Name</th>
<th>Total Task Time (min)</th>
<th>Total Task Mistakes (num)</th>
<th>Mistake Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Task 1 Name</td>
<td>X₁</td>
<td>Y₁</td>
<td>Z₁</td>
</tr>
<tr>
<td>2</td>
<td>Task 2 Name</td>
<td>X₂</td>
<td>Y₂</td>
<td>Z₂</td>
</tr>
<tr>
<td>3</td>
<td>Task 3 Name</td>
<td>X₃</td>
<td>Y₃</td>
<td>Z₃</td>
</tr>
<tr>
<td>4</td>
<td>Task 4 Name</td>
<td>X₄</td>
<td>Y₄</td>
<td>Z₄</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N</td>
<td>Task n Name</td>
<td>Xₙ</td>
<td>Yₙ</td>
<td>Zₙ</td>
</tr>
</tbody>
</table>

- The time \(X_n > 0\) (in minutes) that users take to complete a specific task called "Total Task Time".
- The time \(Z_n \geq 0\) (in minutes) that is spent in recovering from mistakes called "Mistake Time", where \(X_n \geq Z_n\).
- The number of user mistakes \(Y_n \geq 0\) is called "Total Task Mistakes".

3.2 The Second Phase

In the second phase, the data collected in Table 1 is represented in two charts.
- Chart 1: The total tasks mistakes chart. The input of this chart is the column no (2) in Table 1. The output is the percentage of total tasks mistakes.
- Chart 2: The overall time and time of mistakes. The inputs of this chart are the columns no (1) and (3) in Table 1. The output is the comparison between the overall time and time of mistakes.

At the end of this phase, each of the system tasks would be clear enough to be evaluated. Moreover, by making the required comparisons, we can achieve the final evaluation results of the system usability.

4. Implementation

4.1 Data Collection And Analysis

In this study, the U. A. E. Airline website usability is evaluated based on five tasks, which are: (1) Registration, (2) Flight Search, (3) Flight Details, (4) Ticket Purchase and (5) Hotel Search. The results are assembled and presented in Table 2. From the second column in Table 2, the percentage of total tasks mistakes are achieved and represented in Chart 1 given in Figure 2, whereas Chart 2 (given in Figure 3) that represents the comparison between the overall time and time of mistakes is extracted from the first and third columns in Table 2.

Table 2: Results of Testing Table

<table>
<thead>
<tr>
<th>Task No</th>
<th>Task Name</th>
<th>Total Task Time (min)</th>
<th>Total Task Mistakes (num)</th>
<th>Mistake Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Registration</td>
<td>19</td>
<td>2</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>Flight Search</td>
<td>6.45</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Flight Details</td>
<td>10</td>
<td>2</td>
<td>1.05</td>
</tr>
<tr>
<td>4</td>
<td>Ticket Purchase</td>
<td>24</td>
<td>12</td>
<td>15.5</td>
</tr>
<tr>
<td>5</td>
<td>Hotel Search</td>
<td>11</td>
<td>3</td>
<td>6.2</td>
</tr>
</tbody>
</table>
4.2 RESULTS AND DISCUSSIONS

Figure 2 illustrates the number of mistakes, as each of the five tasks (given in Table 2) are displayed by percentages. After we collected and analysed the data, we have found that the fourth task (i.e., Ticket Purchase) has the highest number of mistakes compared with the other tasks. On the other hand, the first three tasks are roughly equal in the number of mistakes.

As shown in Table 2, a total of 19 minutes were spent by all users in task 1, representing of the total task evaluation time. In task 2, 6.45 minutes were spent. Task 3 was performed in 10 minutes, whereas 24 minutes taken in task 4, and 11 minutes was spent in task 5. The task 4 is ranked as the highest time, which has been more difficult to the users during test implementation. In addition, in task 1, a sum of 2 mistakes have been made by users, representing (10%) of the total task mistakes. In task 2, 2 mistakes (10%) were also made, as same as with task 3 (9%). In task 4, 12 mistakes were made (57%). Lastly, in task 5, 3 mistakes (14%) were made by users.

In this evaluation, more mistakes were made by users in task 4. The tasks 1, 2 and 3, had only 2 mistakes each. This implies that the task 4 may be less effective and less efficient in use with regard to the total time spent on fixing mistakes. Moreover, the error in task 4 was the highest with 15.5 minutes spent on mistakes fixing. However, this time represents 65% of the total time of the task. In contrast, in task 2, two mistakes that arise were resolved immediately as the time spent for this task to be fixed was 0%.

With these results, it could be concluded that the task 4 was particularly a problematic and uneasy to the users. It is less efficient in terms of time and less effective with respect to usage. With respect to tasks completion, all tasks were completed by all users except only two users. The only tasks that were not finished (or completed) were task 3 and task 1. This suggests a task completion rate of 0.6. The reason for the none completion of tasks 1 and task 3 by users has been because of access difficulty. The users found it difficult to gain access to the relevant page of the interface to accomplish the tasks. On the average, 14.09 minutes were spent on each task. In addition, an average of 4.2 mistakes were made on each tasks, while an average of 5.2 minutes were spent in fixing mistakes in each task by users. In terms of HCI, there exists a considerable amount of problems of usability in the evaluated website.

Figure 3 shows the comparison between the overall time and the time of mistakes for each of the five tasks. We found that the fourth task is the one that requires the largest time needed for trying
to fix mistakes. However, the mistakes in other tasks have not taken a large amount of time as it could be 0% as demonstrated by task 2, i.e., Flight Search.

In summary, amongst all the facts we have received from the two charts, we can rightfully conclude that the fourth task is the most complicated in the number of mistakes and in the time needed for those mistakes to be resolved. In contrast, the second task is the easiest to accomplish and saves a lot of time. Based on this, it is recommended to follow the way that is used in the second task design, including the GUI, which plays an important role in both the simplicity and sophistication of executing the different tasks system. Moreover, it is recommended to divide the fourth task into multiple secondary sub-tasks as the amount of work needed for it is too big compared to the second task.

![Figure 3: Comparison between the overall time and time of mistakes](image)

5. Validation and Comparisons

For method validation, we have compared the initial results achieved by our method with the results obtained from a questionnaire in users' point of view. This section describes the results of a post-test questionnaire presented to the five users. The purpose of the questionnaire is to recognize the user opinion on the U. A. E. airlines website from two points of view (1) General evaluation and (2) Interface and interaction designs of the website. The following subsections describe some of the questionnaire results. Further details about the questionnaire can be found in [19, 8].

5.1 General Evaluation

For a general evaluation, some question items were designed with a 10-point scale following a semantic differential scale questionnaire design as given in Table 3.
1. Do you feel that you found the needed information?
The mean of this item is 5.20 and the standard deviation is 2.78. The scale ranged from “not at all satisfied” to “very satisfied”. This mean score indicates a slight level of users satisfaction for the website’s interface in terms of interaction. The 5.20 is slightly above the criterion average of 5, but is not far above the overall criterion mean score of the scale (5.00).

2. Is the acquired information accurate?
This scale is also a 10-point semantic differential scale, with an criterion overall mean of 5. The mean of this item is also 5.00 with a standard deviation of 2.55. This implies that the accuracy of the acquired information is around average. Again, we can say it is not far above the overall criterion mean score of the scale (5.00).

3. Does the page title match the content information?
The title of the page does match the content information, though by a small margin, as can be seen from the mean of the item score. The mean is 5.20 and the standard deviation is 2.49.

4. How easy is it to find specific information in this website?
As agreed upon by the users, it is moderately /slightly easy to find specific information in this website on the average. This implied from the mean score of this item which is 5.80. The standard deviation is also, 1.30.

5. Is the information used in this site easy to understand?
The mean of this item is 3.60 and the standard deviation is 2.30. This reveals that on the average, when compared to the overall criterion mean of the scale, the information used in this site is not easy to understand, as the mean score of 3.60 is far below the overall criterion mean score of the scale (5.00). We noticed that the mean results are all surround the middle result, which is around 5 or fractionally above. There is also one, i.e., Is the information used in the site easy to understand?, which has a mean of 3.60 and the item explains how easy it is to understand the information on the website and the important aspects that lead to the users satisfaction or dissatisfaction.

5.2 Interface and Interaction Designs
For interface and interaction design evaluation, some question items were designed with a 10-point scale following a semantic differential scale questionnaire design, as shown in Table 4.

Table 4: Interface and Interaction Designs

<table>
<thead>
<tr>
<th>No</th>
<th>Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How easy is the site’s navigation?</td>
<td>5.00</td>
<td>2.92</td>
</tr>
<tr>
<td>2</td>
<td>How easy is it to read the texts in this website?</td>
<td>4.80</td>
<td>2.95</td>
</tr>
<tr>
<td>3</td>
<td>Are the fonts, sizes, colors of texts appropriate and consistency?</td>
<td>4.60</td>
<td>2.97</td>
</tr>
<tr>
<td>4</td>
<td>Are the tables, charts, or graphics on the website is readable?</td>
<td>5.60</td>
<td>3.65</td>
</tr>
<tr>
<td>5</td>
<td>Compared to what you expected, how quickly did the tasks go?</td>
<td>5.00</td>
<td>2.83</td>
</tr>
</tbody>
</table>
1. How easy is the site’s navigation?
The mean of this item is 5.00 and the standard deviation is 2.92. The mean is about the same as the overall mean. This implies that on the average, this website’s navigation may be easy.

2. How easy is it to read the text in this website?
The mean of this item is 4.80 and the standard deviation is 2.95. On average, it is not easy to read the text in this website, this is as evaluated by the users.

3. Are the fonts, sizes, colors of text appropriate and consistent?
The mean point of this item is 4.60 and it is standard deviation of 2.97. This indicates that on the average, when compared to the overall mean of 5 of the scale, the fonts, sizes, colors of text in this website interface are not appropriate and consistent as observed by the users.

4. Are the tables, charts or graphics on the website readable?
The tables, charts or graphics on the website are readable on the average, this is so, as the mean point of this item 5.60 and the standard deviation is 3.65. However, the observed readability is slightly above the overall average.

5. Compared to what you expected, how quickly did the tasks go?
The mean score of this question is 5.00 and the standard deviation is 2.83. When compared with the overall average of 5.00, it may be gone quickly, though, not certainly.

The results obtained from the questionnaire and those achieved by the method are indistinguishable. The questionnaire results demonstrate the users dissatisfaction, which have been the same results of our method. Based on this findings, it cloud be concluded that the proposed method are encouraging in evaluating websites similar to U. A. E. airline website.

6. CONCLUSION

This paper describes a method for evaluation of a web-based system usability. The U. A. E. airline website is used as a case study. The evaluation of the GUI of the system is carried out by users, so that it provides a direct incite on how actual users in real situation use the system in practice. The study offers a novel method that can be used in evaluating the GUI of a web-based system in order to ensure that the final product of a system designer meets user requirements. The proposed method is divided into two phases. In the first phase, a system usability is evaluated in accordance to a test plan, while the second phase explains the data received from the first phase.

Using different five tasks considered in the evaluation of the usability of the airline website, the results make obvious that the number of mistakes made while using the system. The tasks were: (1) Registration, (2) Flight Search, (3) Flight Details, (4) Ticket Purchase and (5) Hotel Search.

The total mistake time achieved is displayed in percentages. According to two output charts extracted from data analysis, it could be concluded that the fourth task is the most complicated in the number of mistakes and in the time needed for those mistakes to be resolved. In contrast, the second task is the easiest one to accomplish and saves a lot of time. For validation, the results obtained by the method are proven by a questionnaire. The questionnaire has been designed based on two aspects: (1) general evaluation, and (2) interface and interaction designs of the website. The questionnaire results demonstrate the users dissatisfaction, which have been the same findings of our method. Based on this analysis, we cloud conclude that our method results are encouraging in evaluating the website.

REFERENCES


