Usability Evaluation of Web-based Systems: A New Method and Results

Ebitisam K. Elberkawi[1], Naser .F. M. EL-firjani[2], Abdelsalam M. Maatuk[3], Shadi A. Aljawarneh[4]
[1,3]*Faculty of Information Technology, Benghazi University, Libya.*
[4]*Software Engineering Department, Faculty of Computer and CIT, Jordan University of Science and Technology, Jordan*

Address
[1]basma.elberkawi@uob.edu.ly
[2]Nasser195it@yahoo.com
[3]abdelsalam.maatuk@uob.edu.ly
[4]saaljawarneh@just.edu.jo

***Abstract*— One of the most important elements in the system development is the way the system provides an interaction with users through an easy to use graphical user interface. This paper introduces a method for testing and evaluating the degree of usability for web-based systems. The method has been utilized in two phases and applied to the Emirates Airlines Website as a case study. In the first phase a test is implemented on a number of users, and in the second phase the results acquired are converted into charts for the final web-based system evaluation. The results obtained were satisfactory. We have been able to identify the places where the weaknesses and gaps are, and then we draw some recommended solutions to avoid them.**

Keywords*—*Web usability, Evaluation methods, GUI.

I. 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 [1]. 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 [24, 25].

In this paper, we propose an approach for a web-based system testing for usability evaluation. The method, which is implemented on the Emirates Airlines Website, consists 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 final system evaluation. This includes results analysis and recommendations for solving the problems found.

The rest of the paper is organized as follows. Section II provides a background and related work. Section III describes the phases of the method and a description of the testing plan. The discussion of the results is presented in Section IV. Section V concludes the paper.

II. BACKGROUND AND RELATED WORK

Usability is one of the important elements to make product usable. Websites evaluation is determining the quality of the website [2, 23]. 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 [3]. This underscores gross importance of website usability evaluation.

An evaluation is concerned with the gathering of data related to usability of product using a group of users for the tasks and specifying the work environment and context [4]. Generally, two types of website evaluation exist: quantitative and qualitative. Quantitative studies focus on the website quality [5]. The measures of the web pages, such as links formatting and the text elements were introduced in [6], whereas the numerically measurable data, e.g., time-based and traffic-based data have been used in [7]. 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 [8].

An HCI combines gathering data and the intellectuals framework of psychology, using computer tools in creating effective system interface [8]. 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 [9]. Thus, usability is the aspect of website application that the user interacts and gain first-hand interaction with computers [10]. Hence, usability

978-1-5090-5579-1/16/$31.00 ©2016 IEEE

quantitatively and qualitatively measures the design of a system interface for user interaction. According to Nielsen [10], the five key factors of usability are: learn ability, efficiency, memorability, errors, and satisfaction [10].

The most basic method of information acquisition on how user interact with technology and the difficulties faced using the technology is regarded as the usability testing [10, 22]. An approach proposed to improve usability of software applications to interacts with users with an effective manner is described in [16]. The majority of today's computers are designed and centered around the user, as such that gives the same reason as why feedback from the users of the computers cannot be replaced [11]. Usability testing helps in discovering mistakes committed by the 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 [12]. 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 is one of important characters [13]. The use of animation and/or multimedia plug in requirements may affect page loading time [14, 15]. The websites of twenty airline and travel agents in the UK were tested for usability and the result shows that some of the UK's budget airlines were some of the worst to use [3]. Among those websites 50% or less for usability, with Ryanair coming last with just 41%. The British Airways is topped the list with 71%, and closely followed Expedia and Virgin Atlantic 70%. It was observed that the usability issues on Ryanair site include poor search options, such as being unable to search for all flights from the same city rather than specific airports and with flexible departure dates. The lack of options for refining search results was also an issue, while many searches drew a blank, leaving the user at a dead end [3, 22].

A good user interface design typically needs the use of a variety of usability evaluation methods [19, 21]. 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 [20]. 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 [17]. 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 [18]. 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.

III. THE PROPOSED METHOD

In this usability testing method, a summative usability testing plan has been used. The test used was performance measurement. Summative usability testing is the summative 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 users take to complete a specific task.
- The time spent recovering from errors.
- The number of user errors.

Fig 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 many tasks. The *Time* is an important factor, which help us criticize the system. The lowest implementation time leads to a positive evaluation. 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.

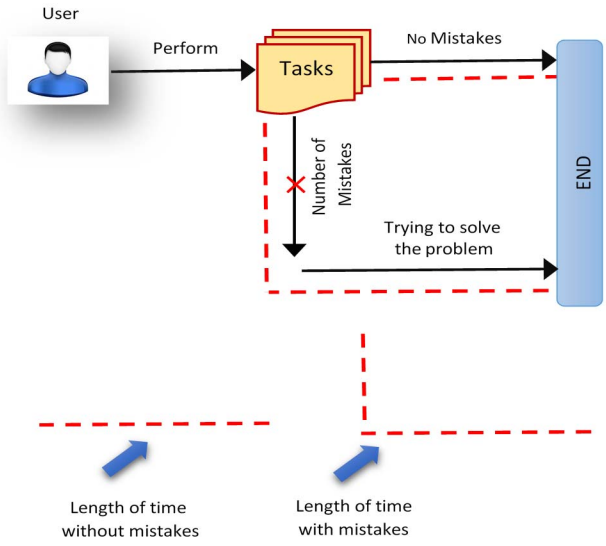

Fig 1 The method usability evaluation process

A. The First Phase

In the first phase, the usability of the chosen web-based system is evaluated according to a test plan. This test plan is implemented by a number of users, during of which we are observing them according to three usability measures selected from Jakob Neilsen usability measures [10].

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 Emirate 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.

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) Traning and Sessions

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

4) 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 I. The usability measures are related to the factors of tasks and times, as follows:

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

TABLE I GENERAL DESIGN OF TESTING TABLE

Task No	Task Name	Total Task Time (min)	Total Task Mistakes (num)	Mistake Time (min)
1	Task 1 Name	X_1	Y_1	Z_1
2	Task 2 Name	X_2	Y_2	Z_2
3	Task 3 Name	X_3	Y_3	Z_3
4	Task 4 Name	X_4	Y_4	Z_4
…..	…….	…..	…..	…..
n	Task n Name	X_n	Y_n	Z_n

B. The Second Phase

In the second phase, the collected data is represented in two charts.

- Chart 1: The total tasks mistakes chart. The input of this chart is the column no (2) in Table I. 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 I. The output is the comparison between the overall time and time of mistakes.

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

IV. RESULTS AND DISCUSSION

In this study, the Emirate 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 II. From the second column in Table II, the percentage of total tasks mistakes are achieved and represented in Chart 1 given in Fig. 2, whereas Chart 2 (given in Fig. 3) that represent the

comparison between the overall time and time of mistakes in is extracted from the first and third columns in Table II.

A. Analysis

Fig. 2 illustrates the number of mistakes, as each of the five tasks (given in Table II) are displayed by percentages. After collected 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.

TABLE II RESULTS OF TESTING TABLE

Task No	Task Name	Total Task Time (min)	Total Task Mistakes (num)	Mistake Time (min)
1	Registration	19	2	4.5
2	Flight Search	6.45	2	0
3	Flight Details	10	2	1.05
4	Ticket Purchase	24	12	15.5
5	Hotel Search	11	3	6.2

As shown in Table II, 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.

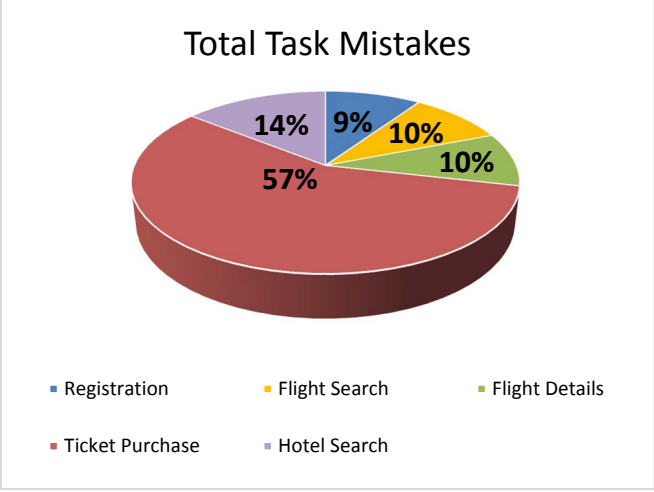

Fig. 2: Percentage of total tasks mistakes

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 is 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) had been 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 some 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. This summary indicates that in terms of human-computer interaction, there exists a considerable amount of problems of usability in the evaluated website.

B. Final Results

Fig. 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 take a large amount of time as it could be 0% as demonstrated by task 2, i.e., Flight Search. 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.

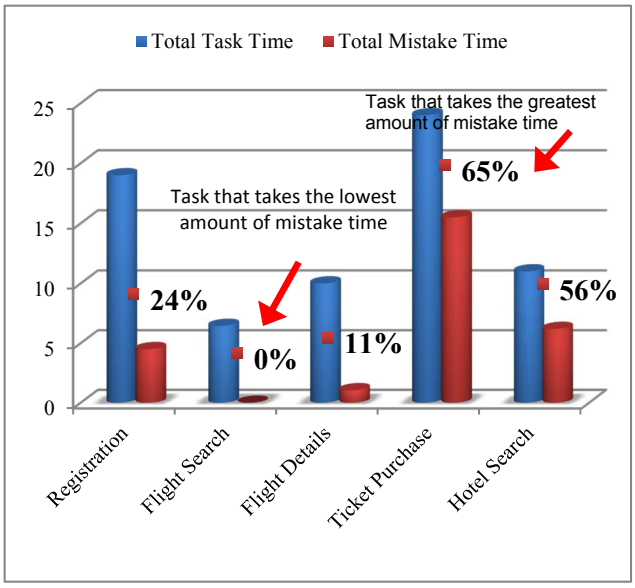

Fig. 3: Comparison between the overall time and time of mistakes

V. CONCLUSION

This paper describes a method for evaluation of a web-based system usability. The Emirate 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 the requirements of the user.

The proposed method is divided into two phases. In the first phase, a system usability 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 Emirate Airline website, the results make obvious 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.

REFERENCES

[1] E. K. El-Berkawi and M. El-Ammari. Development of Graphical User Interface Using Philosophy of Use Case Maps. In *Proceeding of ACIT '11*, 2011.

[2] N. S. Aziz, A. Kamaludin, and N. Sulaiman. Assessing Web Site Usability Measurement. In *IJRET*, vol. 2(9), pp. 386–392, 2013.

[3] G. Chariton. Budget airline website fail usability test. 2009.

[4] J. Preece, , Y. Rogers, H. Sharp, D. Benyon, S. Holland and T. Carey. *Human-Computer Interaction*. Addison-Wesley Longman Ltd., Essex, UK, 1994.

[5] R. Law, S. Qi, and D. Buhalis. Progress in tourism management: A review of website evaluation in tourism research. In *Tour. Manag.*, vol. 31(3), pp. 297–313, 2010.

[6] C. Faba-Perez, V. Guerrero-Bote and F. de Moya-Anegon. Self-organizing maps of web spaces based on formal characteristics. In *Information Processing and Management Journal*, vol. 41(2), pp. 331-346, 2005.

[7] E. Suh, S. Lim, H. Hwang and S. Kim. A prediction model for the purchase probability of anonymous customers to support real time web marketing: a case study. In *Expert Systems with Application*, vol. 27(2), pp. 245-255, 2004.

[8] A. Sarmento. Issues of Human Computer Interaction: Opportunities and Challenges. In *IRM Press*, 2004.

[9] B. Shneiderman, C.Plaisant, M.Cohen and S. Jacobs. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 5th Ed., Pearson Addison Wesley, 2009.

[10] J. Nielsen. *Usability Engineering*. Academic Press, Boston, MA, 1994.

[11] J. Kubie, L. Melkus, R. Johnson and G. Flanagan. User-centered design for productive systems. In *Information Systems Management*," vol. 13(2), pp. 38–48, 2007.

[12] J. Lazar, *User-centered Web development*. Jones & Bartlett Learning, 2001.

[13] S. Anderson. How Fast Should A Website Load?. In *HoBo, Internet Markiting*, 2016 [Online], available at: http://www.hobo-web.co.uk/.

[14] D. Gehrke. Determinants of successful Website design: relative importance and recommendations for effectiveness. In *Proceeding of 32nd Annu. Hawaii Int. Conf. on Syst. Scie. '99*, pp. 1–8, 1999.

[15] K. Nunley. Promoting Your Biz Online: Get Noticed. In *Wealth Building,* p. 47, March 1998.

[16] T. M. Abdelaziz, A. M. Maatuk and F. Rajab. An Approach to Improvement the Usability in Software Products. In *International Journal of Software Engineering & Applications (IJSEA)*, vol.7(2), March 2016. DOI : 10.5121/ijsea.2016.720211

[17] T. Nguyen. Usability Evaluation of Web-Based Learning System. In *International Journal on E-Learning,* vol. 11(3), pp. 281-316, 2012.

[18] P. Chynal. Hybrid Approach to Web Based Systems Usability Evaluation. In *Intelligent Information and Database Systems*, vol, 8397 LNCS, pp 384-391, 2014.

[19] J. Borges, I. Morales and N. Rodríguez. Guidelines for designing usable World Wide Web pages. In *Conf. companion Hum. factors Comput. Syst. common '96*, pp. 277–278, 1996.

[20] J. Melorose, R. Perroy, and S. Careas. Usability Measurement: A Roadmap for a Consolidated Model. In Statew. Agric. L. Use Baseline 2015, vol. 1, 2015

[21] Moftah R. A., Maatuk A. M., Plasmann P. and Aljawarneh S. 2015: An Overview about the Polymorphic Worms Signatures. In *Proc. of ACM Inter. Conf. on Engineering & MIS* (ICEMIS '15). ACM, New York, NY, USA, 4 pages. DOI=http://dx.doi.org/10.1145/2832987.2833031

[22] Elkobaisi M. R., Maatuk A. M. and Aljawarneh S. 2015: A Proposed Method to Recognize the Research Trends using Web-based Search Engines. In *Proc. of ACM International Conference on Engineering & MIS*, (ICEMIS '15). ACM, New York, NY, USA, 4 pages. DOI=http://dx.doi.org/10.1145/2832987.2833012

[23] Maatuk A. M., Akhtar M. A. and Aljawarneh S. 2015: An algorithm for constructing XML Schema documents from relational databases. In *Proc. of ACM International Conference on Engineering & MIS* (ICEMIS '15). ACM, New York, NY, USA, 6 pages. DOI=http://dx.doi.org/10.1145/2832987.2833007

[24] Aljawarneh S., Maatuk A. M. and Akhtar A. 2012: Security Issues in Cloud Computing: A Perspective, In *the Proceedings of the 9th International Conference on Electronics, Computer and Computation (ICECCO'12),* Ankara, Turkey.

[25] Aljawarneh S., Al-Rousan T. and Maatuk A. M. 2012. Security Issues Influencing the Usage of Online Banking in the Arab World: Data Validation, *Security Journal (ISI Thomson)*, pages 10, doi:10.1057/sj.2012.10.