UTAssistant: a Web Platform Supporting Usability Testing in Italian Public Administrations

Giuseppe Desolda
Università di Bari
via Orabona, 4
70125 Bari, Italy
giuseppe.desolda@uniba.it

Rosa Lanzilotti
Università di Bari
via Orabona, 4
70125 Bari, Italy
rosa.lanzilotti@uniba.it

Antonello Cocco
ISCOM-MISE
Viale America 201
00144 Roma, Italy
antonello.cocco@mise.gov.it

Giancarlo Gaudino
ISCOM-MISE
Viale America 201
00144 Roma, Italy
antonello.cocco@mise.gov.it

Stefano Federici
Università di Perugia
Piazza G. Ermini, 1
06123 Perugia, Italy
stefano.federici@unipg.it

CHItaly '17, September 18-20, 2017, Cagliari, Italy.

Abstract
Even if the benefits of the usability testing are remarkable, it is scarcely adopted in the software development process. To foster its adoption, this paper presents a Web platform, UTAssistant, that supports people, also without skills in Human-Computer Interaction (HCI), in evaluating Web site usability.

Author Keywords
Usability testing; public administration web sites; automatic and semi-automatic tools.

ACM Classification Keywords
H.5.2 Human-centered computing~Usability testing.

Introduction
HCI literature proposes different methods for evaluating usability of interactive systems. Among them, usability testing is generally considered the most complete form of evaluation, because it assesses usability through samples of real users and permits to gather a wide range of qualitative and quantitative data, whose analysis allow to detect possible issues about the system and the interface elements which cause them [7].

Although the benefits of this technique are unquestionable, it is hardly adopted due to the required effort and time needed to set up reliable user testing. A frequently used technique is the thinking aloud in which users are asked to verbalize their actions during the interaction with the system to complete a set of critical tasks [7]. During the task execution, the evaluators can collect quantitative data, e.g. task time/clicks, and qualitative data, e.g. facial expressions and externalized comments. In addition, questionnaires are typically administered to evaluate the users' perceived usability. Then, all the collected data are analyzed to evaluate the software usability with respect to the administered tasks, thus producing statistics like the task completion time or number of clicks to accomplish tasks, task success rate, heat map of the user behavior, questionnaire results and graphs, and so on.

To foster the adoption of usability testing technique, in the past few years, there has been massive growth of usability testing tools. For example, professional platforms like *Morae*, *Ovo Logger* and *User Testing* support HCI experts in most of the evaluation activities [2][9, 11]. Despite their powerful features and automatization of data gathering and analysis, their use is limited by their price and/or the need to install them on dedicate machines (e.g., Morae and Ovo Logger). Simpler and cheaper tools are also available but they typically cover part of the evaluation process, for example the user behavior analysis obtained gathering and elaborating the mouse movements/clicks (e.g., *UsabilityTool*) or the facial expressions and externalized comments recording (e.g. *Camtasia*).

This paper proposes a Web platform, called *UTAssistant*, which assists evaluators, even without skills in HCI, in performing usability testing. UTAssistant automatizes and simplifies the usability testing process in all its phases, i.e. design, execution and data analysis.

UTAssistant

UTAssistant is a Web platform designed and developed within the PA++ Project, an initiative coordinated by the High Institute of Communication and Information Technology (ISCOM) scientific and technical body of the Italian Ministry of Economic Development (MISE), whose main goal is to promote the usability practices in the Italian Public Administration (PA). In particular, the PA++ (A Public Administration + mobile and + usable: design and evaluation of Public Administration Web sites) project involves a grouo of people that are members of the GLU (Gruppo di Lavoro per l'Usabilità, in English, Working Group on Usability). The GLU is related to the Italian Ministry of Public Administration and is working to support people of the PA website staff in performing usability evaluation activities of their websites and other e-government systems. Among the different activities of the GLU, in May 2015, a new version of the eGLU and eGLU-M Protocol 2.1 has been published [1]; its aim is to guide web masters and web editors, who do not have experience on usability and UX evaluation, in performing usability testing of the websites they work on, by committing limited resources in terms of time and people.

UTAssistant has been conceived to support remote usability testing executed by using desktop and mobile devices, from the test design to the analysis of the data collected following the guidelines provided by the eGLU 2.1 and eGLU-M evaluation protocol. UTAssistant has been entirely developed as Web platform. The idea is to

provide Italian PA with a lightweight and simple service that does not require any installation on user devices or special requirements. In fact, thanks to the use of the recent evolutions of the HTML5 standard, the browser gathers data from desktop and mobile peripherals like webcam, microphone, mouse, keyboard, and so on. From a technological point of view this is an important step forward with respect to the state-of-the-art, which can foster a wide adoption of this tool and, consequently, of the usability testing technique. Indeed, the existing software for usability testing, require dedicated devices (e.g. Morae) or cover part of the evaluation process and often do not support the evaluation of web site used on mobile devices that today represent the most common interaction device. In the following sections, it is described how UTAssistant supports *evaluators* in designing a usability test and in analyzing the results, as well as how *participants* are driven by UTAssistant in completing usability tests.

Test Design

The first phase of a usability testing is its design. The most important issue of this phase is the study creation that mainly consists in defining a set of tasks users have to complete during the test execution. After the logging into the system, evaluators create a new study: a wizard procedure guides them in designing the test following three main steps. At the first step of the test design, the evaluators have to define: a) general information about the study (e.g., a title, a short introduction), b) data UTAssistant has to gather while participants carry out the tasks (e.g., mouse/keyboard data log, audio/video from participant webcam/microphone), and c) questionnaires participants have to complete during/at the end of the study. According to the eGLU protocol, UTAssistant actually implements post-questionnaires like SUS [5], NPS [8], UMUX-Lite [6], USE [10]. Another post-questionnaire, not mentioned by the eGLU protocols but also implemented in UTAssinstant are: AttrakDiff, [4] and the NASA-TLX questionnaire [3]. After defining the study general information, at the second step evaluators define the task list. For each task, they have to provide the title, the starting/ending URL, the goal, the duration. Finally, the third step requires evaluators to specify the study participants, by selecting them from a list of users already registered in the platform or by typing their email. Then the study has to be saved. The invited participants receive an email describing all the instructions to participate in the usability testing.

Test Execution

A usability testing execution is typically managed by one or more HCI experts, which assist participants in carrying out the study tasks, gather data (e.g., externalized comments, task times/click) and administer the questionnaires. UTAssistant aims to automatize all the execution steps guiding participants even without the physical presence of evaluators.

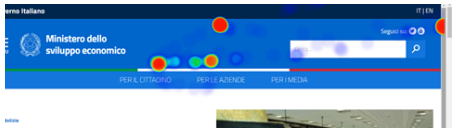
Once the participant receives an email, by clicking the URL reported inside it, UTAssistant is open in a browser on the desktop or mobile device. First, UTAssistant shows general indications about the platform (e.g., a short description of the toolbar with all the useful commands) as well as the privacy policies indicating that data like mouse/keyboard log data and webcam/microphone stream could be captured. Then, UTAssistant administers the tasks one at a time. For each task, the platform shows the goal in a pop-up window. A minimal toolbar is positioned on top of the

Tasso di successo

	Task 63	Task 64	Task 65	Task 66	Task 67	Tasso di successo medio per partecipante
rtecipante 39	0	1	1	0	0	40%
rtecipante 40	0	0	1	0	1	40%
rtecipante 41	0	1	1	1	1	80%
rtecipante 42	1	1	1	1	1	100%
rtecipante 43	0	1	1	0	1	60%
rtecipante 44	1	1	1	0	1	80%
sso di successo dio per task	33.33%	83.33%	100%	33.33%	83.33%	

o di successo medio complessivo (task e partecipanti): 66.67%

ure 1. Summary of the success rate of a formed study.



gure 2. Example of heatmap overlapped to page involved in a task execution.

task web page reporting the current task goal and description, the duration time, the task number, and the buttons to go to the next task, or stop the study. If the user is executing the last task, the "Next task" button becomes "Fill in the questionnaire". All the steps, are completely assisted by UTAssistant that simulates the presence of the HCI expert.

Test Data Analysis

One of the most time-consuming phase of a usability testing is the data analysis. HCI experts have to store, merge and analyze the mouse logs, the video/audio recordings, the questionnaire results, and so on. UTAssistant assists evaluators in performing these activities, thus making easier and lighter the entire analysis. UTAssistant provides different types of tools for the automatic data analysis and to produce information that facilitate the evaluators' work in detecting usability issues.

One of the most important usability metric is the task success rate, i.e. the percentage of tasks that users complete correctly during the study [7]. UTAssistant creates a report showing the general success rate ("Tasso di successo medio complessivo" in Italian as shown at the bottom of Figure 1) and a table reporting the details for each task (column)/user (row).

In addition, for each task UTAssistant shows: its title, the number of pages the user visited to carry out it, the number of notes evaluator made during the analysis, the date and time in which the task starts and finishes, the average time and the average number of clicks all the users needed to complete the task.

A different analysis of the mouse/keyboard logs is provided by the heatmaps, i.e., graphical visualizations overlapped to a web page that show where people click/touch, move the mouse, scroll. It uses hotter colors, e.g. orange or bright red, on areas that get the most clicks/movements and colder colors, e.g. blues and purples, on areas that get the less clicks/movements. For each web page, UTAssistant builds a heatmap that evaluators can analyze to understand if some area of the web page wrongly attracted the users' attention due to some user interface elements. Figure 2 shows an example of mouse heatmap.

UTAssistant automatically analyzes questionnaire data and provides proper statistics and graphs to summarize the results and to help evaluators in understanding the perceived usability, cognitive load and so on.

Conclusion

This paper has presented a web-based platform, called UTAssistant, developed in the project PA++, to support people working in PA website staff in performing usability tests. A preliminary evaluation of the UTAssistant, involving 2 people acting as evaluators and 6 people as participants, showed that evaluators appreciated a lot the support to the evaluation activities, while participants enjoyed the possibility to participate in a user test, without being observed.

Acknowledgements

This work is supported by ISCOM under grant PA++ (A Public Administration + mobile and + usable: design and evaluation of Public Administration Web sites).

References

1. eGLU and eGLU-M Protocol 2.1. Retrieved from http://www.funzionepubblica.gov.it/glu.aspx Last Access June 4th, 2017
2. Morae. Retrieved from https://www.techsmith.com/morae.html Last Access June 10th, 2017
3. Hart S. G. and Staveland L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology 52(1)*, 139-83.
4. Hassenzahl M., Burmester M., and Koller F. (2003). AttrakDiff: A questionnaire to measure perceived hedonic and pragmatic quality. In *Proc. of the Mensch & Computer*. 187-96.
5. Lewis J. and Sauro J. (2009). The Factor Structure of the System Usability Scale. In *Human Centered Design - HCD 2009*, Kurosu M. Ed. Lecture Notes in Computer Science, Vol. 5619, Springer Berlin Heidelberg, 94-103
6. Lewis J. R., Utesch B. S., and Maher D. E. (2013). UMUX-LITE: when there's no time for the SUS. In *Proc. of the Conference on Human Factors in Computing Systems (CHI '13)*. Paris, France, ACM, New York, NY, USA, 2099-102.
7. Preece J., Sharp H., and Rogers Y. 2015. *Interaction Design-beyond human-computer interaction*. John Wiley & Sons.
8. Reichheld F. F. (2003). The one number you need to grow. *Harvard business review 81(12)*, 46-55.
9. Ovo Logger. Retrieved from http://www.ovostudios.com/ Last Access June 10th, 2017
10. Terenzi M., Di Nocera F., and Ferlazzo F. (2006). Firmitas, Utilitas, Venustas: Assessing the Validity of the Usability Evaluation (Us. E.) Questionnaire. *Developments in Human Factors in Transportation, Design, and Evaluation. Maastricht, the Netherlands: Shaker Publishing*, 249-53.
11. UserTesting. Retrieved from https://www.usertesting.com/ Last Access June 10th, 2017