Chapter 24

REMOTE WEB USABILITY EVALUATION EXPLOITING MULTIMODAL INFORMATION ON USER BEHAVIOR

Fabio Paternò, Angela Piruzza, and Carmen Santoro
ISTI-CNR, Via G.Moruzzi 1,56124 Pisa (Italy)
E-mail: fabio.paterno@isti.cnr.it – Web: http://giove.cnuce.cnr.it/~fabio/
Tel.: + 39 050 3153066 – Fax: + 39 050 3138091

Abstract In this paper we describe MultiModal WebRemUsine, a tool for remote usability evaluation of Web sites that considers data regarding the user behaviour coming from multiple sources. The tool performs an automatic evaluation of the usability of the considered Web site by comparing such data with that contained in the task model associated with the pages (which describes the expected behavior of the user). The results of the analysis are provided along with information regarding the user behavior during the task performance. Using such information, evaluators can identify problematic parts of the Web site and make improvements, when necessary. An example of application of the proposed method is also discussed in the paper.

Keywords: Multimodal data, Remote usability evaluation, Web usability.

1. INTRODUCTION

The great penetration of Web sites raises a number of challenges for usability evaluators. In this paper we discuss what information can be provided by automatic tools able to remotely process multimodal information on user behavior gathered from different sources. The collected information ranges from browser logs to videos and eye-tracking data. The approach proposed aims to integrate such data in order to derive the most complete information for analyzing, interpreting, and evaluating the user interactions while visiting a Website. The proposed approach is supported by a tool – MultiModal Web RemUsine, which is able is to identify where users interactions deviate from those envisioned by the system design and represented in the related task model. To this end, it exploits the integration of data coming from such different sources for better understanding potential problems in task accomplishment. Thus, the evaluator is provided with a more comprehensive picture of the actions performed by the user and, consequently, with more information in order to effectively interpret and evaluate the associated user interface. Moreover, the approach proposed has the remarkable advantage to allow evaluators to identify usability problems even if the analysis is performed remotely, which might contribute to keep at minimum the evaluation costs and allows the users to remain in their familiar environments during the evaluation, improving the trustworthiness of the evaluation itself.

2. RELATED WORK

While a Web site can easily be developed using one of the many tools available able to generate (X)HTML from various types of specifications, obtaining usable Web sites is still difficult. Indeed, when users navigate through the Web they often encounter problems in finding the desired information or performing the desired task. With over 30 million Web sites in existence, Web sites have become the most prevalent and varied form of human-computer interface, but, at the same time, with so many Web pages being designed and maintained, there will never be a sufficient number of professionals to adequately address usability issues without automation [2]. For these reasons, interest in automatic support for usability evaluation of Web sites is rapidly increasing [1,6], especially as far as the remote evaluation is concerned, because, on the one hand, it is important that users interact with the application in their daily environment, but, on the other hand, it is impractical to have evaluators directly observe users' interactions.

Some studies [8] have confirmed the validity of remote evaluation in the field of Web usability. Some work [3] in this area has been oriented to using audio and video capture for qualitative analysis performed by evaluators on the result of usability testing. Other works have highlighted the importance of performing a comprehensive evaluation able to take into account data derived from multiple sources, and the consequent need to provide analysts from a variety of disciplines (each using distinct sets of skills to focus on specific aspects of the problem) to work cooperatively, in order to adequately gain insight into large bodies of multi-source data [7]. In our case we focus more on quantitative data and provide the support for an intelligent analysis of such data so as to extract useful information for evaluation goals.

3. THE ARCHITECTURE

Our approach is mainly based on a comparison of planned user behavior and actual user behavior [4]. Information about the planned logical behavior of the user is contained in a (previously developed) task model, while data about the actual user behavior is provided by the other modules (the logging tool, the Web cam and the eye-tracker), which are supposed to be available within the client environment. An overview of the general approach is described in Fig. 1, where we use ovals to indicate data (the colored ovals better highlight the data which are provided to the tool), whereas the rectangles indicate the hardware/software modules aimed at manipulating such data. The eye-tracker provides quantitative data about the gaze of the user during the evaluation session: one of the most relevant measures regards the *scanpaths*, namely the traced routes of the user gaze used to give insights about the navigation strategy followed by the user during the visit of the page. Contextual information is provided by video-based data recorded during the session by a Webcam. The logging tool stores various events detected by a browser, using Javascripts encapsulated in the (X)HTML pages and executed by the browser. When the browser detects an event, it notifies the script which captures the event detected by the browser and adds a temporal indication. Then, a Java applet communicates the log files to the server. The logging tool provides useful information for correctly correlate the data coming from the different sources used in our approach (the eye tracker, the Webcam, etc.), and to this aim some relevant modifications were needed to be implemented. For instance, in order to manage the data associated with the eye-tracker, it is necessary that whenever a scroll event is recorded, also the extent of the shift with respect to the top and bottom corner of the page is recorded as well by the logging tool, so as to reconstruct the actual area that the user was currently looking at. In the same way, in order to correctly manage the correlation between tasks and videos (so as to provide e.g., evaluation about the completion of tasks) it is necessary that the logging tool is able to record the information about starting/ending time of the tasks.

As for the planned user behavior, CTT [5] task models are used to describe it by their graphical representation of the hierarchical logical structure of the potential activities along with specification of temporal and semantic relations among tasks. It is worth pointing out that, with the CTT notation used, the designer might easily specify different sequences of paths corresponding to the same logical behavior just using the same temporal operator, in order to allow the needed flexibility in describing the user behavior: for instance, if two activities should be concurrently performed, (which means that the first one might be performed as the first activity, but also the vice versa is allowed), this behavior is expressed by using the concurrency operator as the right relationship between these two tasks. By comparing the ideal behavior (contained within the task model) with the information coming from logs, videos and the eye tracker, MMWebRemUsine is able to offer the evaluators useful hints about problematic parts of the considered Web site.

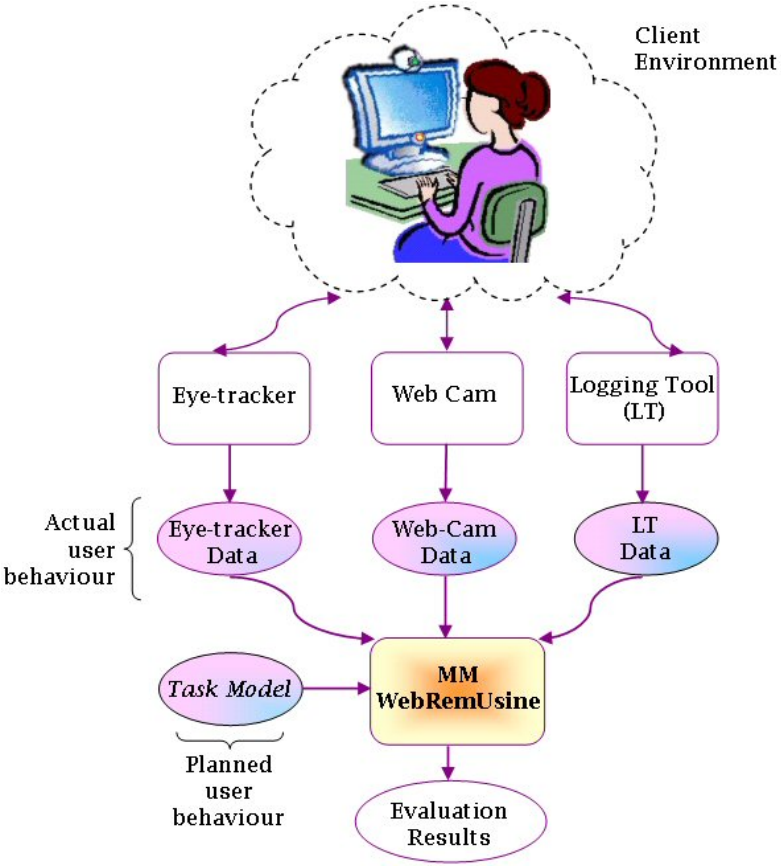


Figure 1. The Architecture of the Environment.

4. THE METHOD

The method underlying the tool is composed of two main phases, the preparation and the evaluation.

4.1 The Preparation

The main goal of the preparation phase is to create an association between the *basic tasks* of the task model and the *events* that can be generated during a user session and recorded in log files. This association allows the tool to use the semantic information contained in the task model to analyze the sequence of user interactions stored in the logs. Basic tasks are tasks that cannot be further decomposed and can belong to three different categories according to the allocation of their performance: *user tasks* are internal cognitive activities and this cannot be captured in system logs, *interaction tasks* are associated with user interactions (e.g., click, change) and *system tasks* are associated with the internal browser generated events.

Three types of events can be saved in the logs: user-generated events (e.g., click, change), page-generated events (associated with loading and sending of pages and forms) and events associated with the change in the target task by the user, which is explicitly indicated through selection from the list of supported tasks. Each event should be associated to one task, a task can be performed through different events (e.g., the movement from one field to another one within a form can be performed using either the mouse, or the arrow key or the Tab key). If an event is not associated with any basic task, it means that either the task model is not sufficiently detailed, or the action is erroneous because the application design does not call for its occurrence. An example of association between a task and an event is, for instance, the association between the task "selecting the home page" and the event "Click on the Home button". Once the association between tasks and events has been carried out, it is possible to move on the evaluation.

4.2 The Evaluation

In the evaluation phase the proper automatic analysis is performed: MMWebRemUsine examines the logged data with the support of the task model and provides a number of results also analyzing the data coming from videos and the eye tracker. Such data can provide useful information especially when it is possible to exploit them in an integrated and cross-checking-based approach (as with MMWebRemUsine), for identifying explanations to any problems users might have encountered during the test.

During the test phase all the user actions are automatically recorded, including those associated to goal achievement. The evaluation consists in analyzing such sequences of actions to determine whether the user has correctly performed the tasks as defined in the task model (the user was able to reach the goals and the actions performed were actually useful to reach them) or some errors occurred (e.g., a precondition error, which means that the execution task order did not respect the relations defined in the system design model). In addition to the detailed analysis of the sequence of tasks performed by the user, evaluators are provided with some results giving an overall view of the entire session considered (such as tasks performed correctly, tasks with precondition errors, how many times a task or an error have been performed, tasks never performed, and pattern of tasks). Such information allows the evaluator to identify what tasks are easily performed and what tasks create problems to the user. Moreover, revealing tasks never performed can be useful to identify parts of the application that are difficult to comprehend or reach. On the basis of such information the evaluator can decide to redesign the site in order to reduce the number and complexity of the activities to be performed.

From the semantic log analysis aimed at comparing the actual behavior recorded by the tool with the ideal behavior specified within the task model various types of results can be generated:

- *Success*: the user has been able to perform a set of basic tasks required to accomplish the target task and thus achieve the goal.
- *Failure*: the users starts the performance of the target task but is not able to complete it;
- *Useless uncritical task*: the user performs a task not strictly useful to accomplish the target task but does not prevent its completion.
- *Deviation from the target task*: in a situation where the target task is enabled and the user performs a basic task whose effect is to disable it. This shows a problematic situation since the user is getting farther away from the main goal in addition to performing useless actions.
- *Inaccessible task:* when the user is never able to enable a certain target task.

A further type of information considered regards the task execution duration, calculated for both high level and basic tasks, which can provide information useful to understand what the most complicated tasks are or what tasks require longer time to be performed. It is worth pointing out that longer execution times do not always imply complicated tasks. In some cases download time can be particularly high, and with this regard MMWebRemUsine provides detailed information, so that evaluators can know its impact on the total performance time. In other cases, when such a long time cannot be explained by a long download, a further cross-checking analysis of additional information provided by the tool (e.g., videos recorded during the session) should be performed in order to find a reasonable motivation for the usability problem (see Section 4.2.3 for an example of it).

The data from videos are important because they can provide more "contextual" information during the performance of a task. Indeed, since the evaluation is remotely performed, the evaluator is not in a position to understand if any condition might have disturbed the performance of a task while the user visits the Web site in his/her own environment. For instance, as we pointed out in the previous section, a long time (or, at least, a time longer than expected) for completing a task might not necessarily be brought about by a usability problem or by a high download time: indeed, it may be caused by some external factors (e.g., interruptions occurring in the user's environment during the session). Another useful information that can be gained from videos are user comments, which sometimes can reveal that users are aware of having performed an error but cannot undo the actions.

In order to provide the evaluator with video-based data, an association between task and video is automatically performed by the tool, thanks to the information regarding the start/end time of the different tasks. Indeed, as the whole session is recorded by a Webcam, through such times it is possible to split the video associated with the entire user session into different fragments related to the completion of the various tasks, together with the possibility to activate/stop the visualization of the related video with a suitable player in the tool. In this way, when the evaluators identify e.g., inexplicably long durations for completing a task, they can easily activate the interested fragment of the video to get further data and investigate about the contextual conditions occurred during the concerned period.

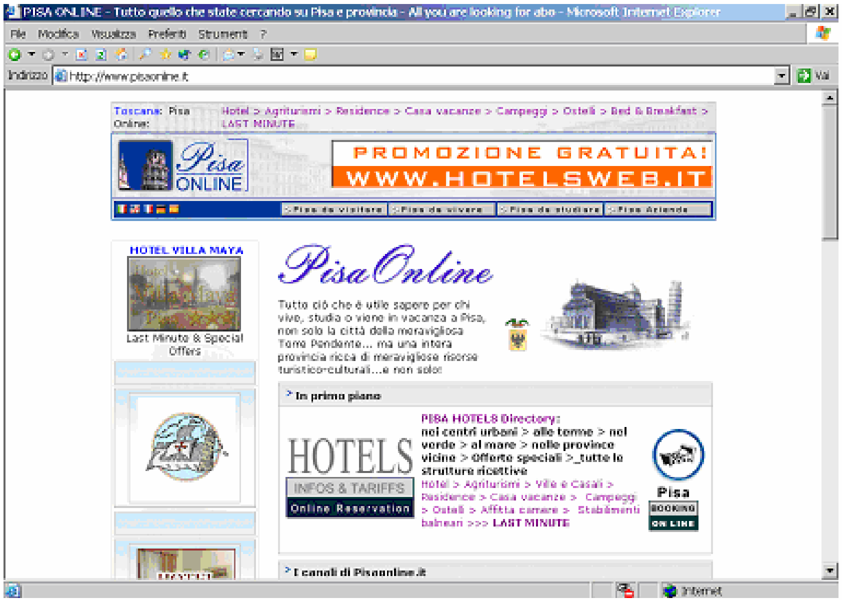
While videos provide more 'contextual' information regarding users, giving the means for correctly interpreting the user's actions, the eye-tracker provides technical measurements and traces of the visual routes followed by users while visiting a Web site. The data provided by the eye-tracker can be interesting "per se" (e.g., the evaluator can understand the areas of the page that attract or not user attention), but they assume even more importance when compared with the user intention (namely: the target task). Indeed, having in mind the objective the user should achieve, it might be relevant to analyze the areas around the links that should be followed in order to reach such goal according to the task model. For instance, it might be relevant to analyze the extent of time the users spent looking at the areas that attracted their attention (duration of *fixations*), as well as the number of fixations. Long fixations might be a sign of user's difficulty in elaborating the information or a sign of high interest in the information. A high number of fixations on certain areas might indicate that the users are confused and are not able to find the information that they are looking for. Moreover, also a long scan path might indicate that the structure underlying the page is rather complicated. All the data have been automatically integrated within the tool, which is able to offer, e.g., for the various tasks, the related video excerpts and the connected data from the eye tracker (i.e., scan paths and fixations).

5. AN EXAMPLE APPLICATION

In this section we show an example of application of the proposed evaluation method and of the related tool, which, in its current version, s mainly aimed at being used for usability tests. The Web site we considered (http://www.pisaonline.it) provides information about Pisa, and in Fig. 2 the homepage is shown. The Website is divided into four main sections: "Pisa da

Visitare" (Visit Pisa), "Pisa da Vivere" (Live in Pisa), Pisa da Studiare" (Study in Pisa) e "Pisa Aziende" (Companies in Pisa). For sake of brevity in Fig. 3 only a simplified version of the task model is visualized, yet detailed enough to highlight the four main tasks for accessing the main sections of the site, together with some tasks that we will refer to in this section.

If we focus more properly on the decomposition of the high level task "Visit Pisa" , it is possible to see that one of its sub-tasks provides access to the "Ulisse" subsection, which inherits the name from the title of an airline magazine offering tourist information about Pisa and providing several information about the town, including data about local products (e.g., information about the white truffle, Fig. 3).



PISA ONLINE - Tutto quello che state cercando su Pisa e provincia - All you are looking for abo - Microsoft Internet Explorer

Indirizzo http://www.pisaonline.it

Toscana: Pisa Online: Hotel > Agriturismi > Residence > Casa vacanze > Campeggi > Ostelli > Bed & Breakfast > LAST MINUTE

PROMOZIONE GRATUITA! WWW.HOTELSWEB.IT

HOTEL VILLA MAYA

Last Minute & Special Offers

Tutto ciò che è utile sapere per chi vive, studia o viene in vacanza a Pisa, non solo la città della meravigliosa Torre Pendente... ma una intera provincia ricca di meravigliose risorse turistico-culturali...e non solo!

In primo piano

PISA HOTELS Directory: nei centri urbani > alle terme > nel verde > al mare > nelle province vicine > Offerte speciali >_tutte le strutture ricettive

Hotel > Agriturismi > Ville e Casali > Residence > Casa vacanze > Campeggi > Ostelli > Affitta camere > Stabilimenti balneari >>> LAST MINUTE

I canali di Pisaonline.it

Figure 2. The home page of the evaluated Web site.

Once having performed the task-event associations it is possible to move to the first step of the proper evaluation phase: the identification of a number of target tasks (the high level activities represented within the task model) to be provided to the test participants at the beginning of the evaluation session. Examples of target tasks considered for the example were "Trova Info su Tartufo Bianco" (Access information about the white truffle), "Trova gradazione alcoolica del Chianti", (Find alcoholic content of Chianti), "Trova ristoranti" (Find restaurant), etc. Once the user selected the interested target task, the environment is in a position to know the intention of the user and automatically identify, within the task model, the *planned* paths that should be followed by the user while carrying out the selected task, which will be used as paragon term for the evaluation. In our test we involved 8 participants aging between 21 and 36. One user selected as target task "*Find information on the white truffle*", which was a subtask of "Access Ulisse" (see simplified task model in Fig. 3). The analysis of this user trying to carry out this activity reported a number of precondition errors. The logging tool recorded several actions performed by the user, which were judged not necessary when compared with the designer's planned path for achieving the task goal (as it is described in the related task model). In addition, the same user was observed pausing a lot looking at the area of the homepage dedicated to the companies in Pisa (a fixation with a relevant duration was registered by the eye tracker), instead of correctly focusing on the "*Visit Pisa*" section which represents the right route for completing the selected task (see the system task model in Fig. 3). From this it might be derived that the user might have misinterpreted "White Truffle" as the name of a restaurant.

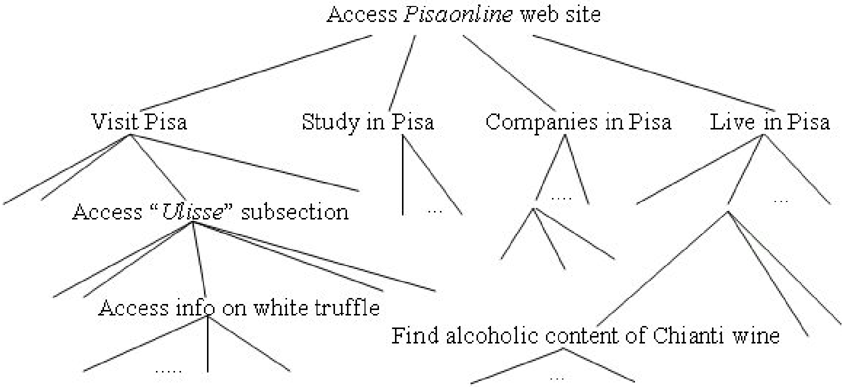

Access *Pisaonline* web site
Visit Pisa
Study in Pisa
Companies in Pisa
Live in Pisa
Access "*Ulisse*" subsection
...
....
...
Access info on white truffle
Find alcoholic content of Chianti wine
.....
...

Figure 3. A simplified version of the task model of the PisaOnLine Web site.

Moreover, once the same user finally realized the correct section on which looking for the concerned information ("Visit Pisa" section), the evaluation still highlighted –through a long scan path- a possible user difficulty in identifying the right link for accessing the "Access Ulisse" section. Indeed, when referring back to the concerned page, the evaluators noticed that, actually, within this page there are three different links for accessing the Ulisse section (they are highlighted by three circles in Fig. 4): a textual link (with label "Ulisse"), another textual link with a different label ("Alitalia Ulisse"), and also an icon with an image associated to Ulisse. To make things even worse further analysis reported that the information available through the last two links is different from the information reachable through the first link.

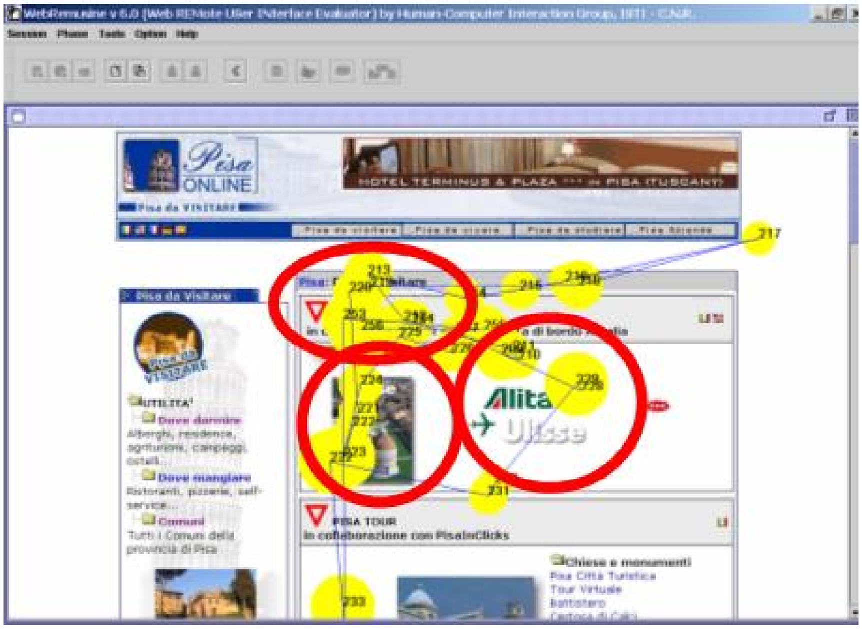

WebRemusine v 6.0 (Web REMote USer INterface Evaluator) by Human-Computer Interaction Group, ISTI - C.N.R.
Session Phase Tools Option Help
Pisa ONLINE
HOTEL TERMINUS & PLAZA *** in PISA (TUSCANY)
Pisa da VISITARE
Pisa da visitare
Pisa da vivere
Pisa da studiare
Pisa Aziende
Pisa da Visitare
UTILITA'
Dove dormire
Alberghi, residence, agriturismi, campeggi, ostelli...
Dove mangiare
Ristoranti, pizzerie, self-service...
Comuni
Tutti i Comuni della provincia di Pisa
a di bordo Alitalia
Alitalia Ulisse
PISA TOUR
in collaborazione con PisaInClicks
Chiese e monumenti
Pisa Città Turistica
Tour Virtuale
Battistero
Certosa di Calci
213
217
220
216
218
219
224
225
226
228
229
221
222
223
231
233

Figure 4. The ambiguity of links related to the section dedicated to "Ulisse".

For another user, who selected "Find alcoholic content of Chianti wine" ("Trova gradazione alcolica del Chianti") as target task, the eye-tracker reported many fixations recorded on the link associated with "Pisa Aziende" ("Companies in Pisa"), rather than, more correctly, within the "Live in Pisa" section, where the link actually is (as you can see from the task model in Fig. 3). This highlighted that the logic followed by users in finding such information was different from that followed by designers.

Moreover, the experiment highlighted that the majority of users did not select the image link associated with the homepage of the PisaOnline Web site (visualized in the top left part of the homepage, see Fig. 5), which was rather surprising due to the relevancy of this page within the entire site. The occurrence of such behavior in almost all users can be interpreted with the fact that the link is rather unclear, and this intuition is reinforced by the image related to the scan path of users on the page (Fig. 5), highlighting that almost all users did not pause on looking at the concerned image link, which might have been confused with a bare decorative image, (especially because it appears on the top part of the page).

In another experiment we analyzed a different site regarding a publishing house and mainly focused on data recorded by videos. Fig. 6 shows the evaluation of task/time performed by the tool, regarding a user who explicitly declared at the end of the task that she was wrong at completing the task.

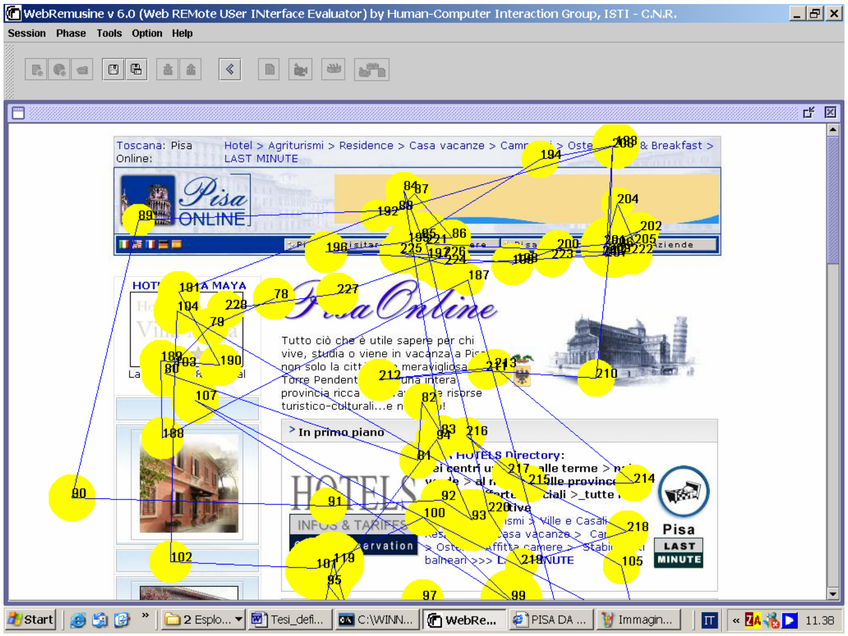
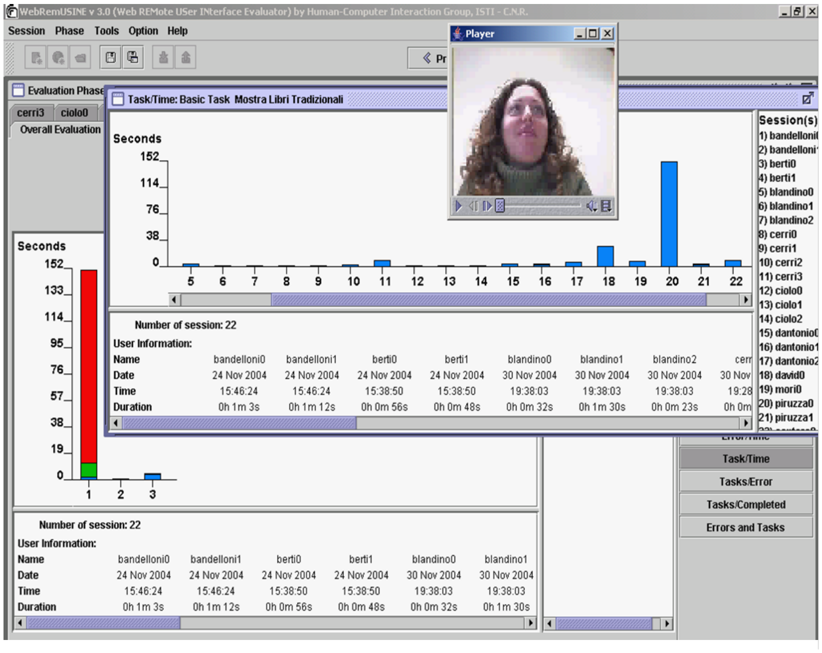

WebRemusine v 6.0 (Web REMote USer INterface Evaluator) by Human-Computer Interaction Group, ISTI - C.N.R.
Session Phase Tools Option Help
Toscana: Pisa Online:
Hotel > Agriturismi > Residence > Casa vacanze > Camp... > Oste... & Breakfast > LAST MINUTE
Pisa ONLINE
Pisa Online
In primo piano
HOTELS Directory:
INFOS & TARIFFS
Pisa LAST MINUTE
Start
11.38

Figure 5. Scanpath of “Find Alchoolic Content of Chianti”.



Session(s)
1) bandelloni0
2) bandelloni1
3) berti0
4) berti1
5) blandino0
6) blandino1
7) blandino2
8) cerri0
9) cerri1
10) cerri2
11) cerri3
12) ciolo0
13) ciolo1
14) ciolo2
15) dantonio0
16) dantonio1
17) dantonio2
18) david0
19) mori0
20) piruzza0
21) piruzza1

Number of session: 22

User Information:

Name	bandelloni0	bandelloni1	berti0	berti1	blandino0	blandino1	blandino2	cerr
Date	24 Nov 2004	24 Nov 2004	24 Nov 2004	24 Nov 2004	30 Nov 2004	30 Nov 2004	30 Nov 2004	30 Nov
Time	15:46:24	15:46:24	15:38:50	15:38:50	19:38:03	19:38:03	19:38:03	19:28
Duration	0h 1m 3s	0h 1m 12s	0h 0m 56s	0h 0m 48s	0h 0m 32s	0h 1m 30s	0h 0m 23s	0h 0m

Number of session: 22

User Information:

Name	bandelloni0	bandelloni1	berti0	berti1	blandino0	blandino1
Date	24 Nov 2004	24 Nov 2004	24 Nov 2004	24 Nov 2004	30 Nov 2004	30 Nov 2004
Time	15:46:24	15:46:24	15:38:50	15:38:50	19:38:03	19:38:03
Duration	0h 1m 3s	0h 1m 12s	0h 0m 56s	0h 0m 48s	0h 0m 32s	0h 1m 30s

Figure 6. Task/Time Information with Video of the user.

The data from the videos were useful to detect some usability problems. For instance, by examining the tasks that were wrongly performed, it was possible to have further details on the facial expressions of the users, who sometimes seemed to be confident of their choices, while other times seemed to be quite confused and doubtful, and this information is important when evaluating the user behavior. Particularly useful information was gained from videos as far as the execution time is concerned, which sometimes seemed to be higher than expected: the analysis of the video revealed that users pause at looking the page, then they happen to comment on it, so this is important to understand that sometimes users are distracted/attracted by portions of the page that are not relevant for carrying out the concerned task, but only by curiosity.

6. CONCLUSION

In this paper, we propose a method, and the associated tool, for remote evaluation of Websites that, through a combination of different sources of data coming from the client side (currently log files, videos and eye tracker data) allows the evaluator to get detailed information about the behavior of the users. Such composite information is the input of an automatic tool that has shown to be effective in providing evaluators with means for discovering possible problematic areas of the Web site. Future work will be dedicated to extending the data detected regarding the user behavior and state, including the emotional state, in order to have a more complete analysis of what happens during user sessions and better identify the potential usability issues.

REFERENCES

[1] Card, S., Pirolli, P., Van der Wege, M., Morrison, J., Reeder, R., Schraedley, P., and Boshart, J., *Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability*, in Proc. of CHI’2001, ACM Press, 2001, pp.498-504.

[2] Ivory, M.Y. and Hearst, M.A., *The State of The Art in Automating Usability Evaluation of User Interfaces*, ACM Computing Surveys, Vol. 33, No. 4, Dec. 2001, pp. 470-516.

[3] Lister, M., *Streaming Format Software for Usability Testing*, in Proc. of CHI’2003, Extended Abstracts, ACM Press, New York, 2003, pp. 632-633.

[4] Paganelli, L. and Paternò, F., *Tools for Remote Usability Evaluation of Web Applications through Browser Logs and Task Models*, Behavior Research Methods, Instruments, and Computers, The Psychonomic Society Pub., Vol. 35, No. 3, August 2003, pp.369-378.

[5] Paternò, F., *Model-Based Design and Evaluation of Interactive Applications*, Springer Verlag, Berlin, 1999.

[6] Scholtz, J. and Laskowski, S., *Developing Usability Tools and Techniques for Designing and Testing Web Sites*, in Proc. of HFWeb’98 (Basking Ridge, June 1998).

[7] Tennent, P. and Chalmers, M., *Recording and Understanding Mobile People and Mobile Technology*, E-social science, 2005.

[8] Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., and Bergel, M., *An Empirical Comparison of Lab and Remote Usability Testing of Web Sites*, in Proc. of Usability Professionals Conference UPA’2002 (Pennsylvania, 2002).