WISH YOU WERE USABLE! HOW TO IMPROVE THE QUALITY OF A MUSEUM WEB SITE

Luca Triacca
TEC-LAB
University of Italian Switzerland
Lugano, Switzerland
luca.triacca@lu.unisi.ch

Davide Bolchini
TEC-LAB
University of Italian Switzerland
Lugano, Switzerland
davide.bolchini@lu.unisi.ch

Nicoletta Di Blas
HOC
Politecnico di Milano

Milano, Italy
diblas@elet.polimi.it

Paolo Paolini
HOC
Politecnico di Milano

Milano, Italy
paolini@polimi.it

Abstract
MiLE (Milano – Lugano Evaluation Method) is an innovative method for evaluating quality and usability of hypermedia applications. MiLE is a methodology combining usability inspection methods and user testing. In particular, this paper focuses upon the Usability Kit (U-Kit) of MiLE concerning museum applications.

THE IMPORTANCE OF EVALUATING THE USABILITY OF MUSEUM WEB SITES

Usability has recently assumed a much greater importance in the internet economy than it had in the past [1], since a web site is an "open product", accessible by anyone who navigates in the WWW. Usability has therefore become a fundamental issue, in every phase of the design process, from the beginning to the end [2].
In particular, Museum Web sites, whose goal is to communicate robust cultural content to a large number of users, have to pay special attention to their usability, or rather quality. Clearly, this is an arduous task for the designers (and in general for all the stakeholders involved in the development of the application): Museum Web sites are of growing complexity, address several targets, deal with complex content, have different communication goals: for all this reasons, they need to be well "usable" and efficient. Evaluating the usability of a web application means to try and answer some crucial questions: e.g., How can we avoid users "getting lost" in the site? How is it possible to improve navigation's effectiveness? What kind of contents shouldn't be missing? How is it possible to know whether the users have learnt anything from the site? The -ambitious- goal is to establish the degree of user satisfaction with the application and consequently a set of guidelines for improving its quality.

EXISTING USABILITY METHODS

Within the field of usability methods it is possible to identify several approaches for evaluating web usability. Among them, the most commonly adopted are *user-based* methods (or *user testing methods*) and *usability inspection methods* (or *expert reviews*) [3].
User-based methods mainly consist of user testing, in which usability properties are assessed by observing how the system is actually used by some representatives of real users [4,5]. User-based evaluation provides the most trusty evaluation, because it assesses usability through samples of real users. However, it has a number of drawbacks, such as the difficulty to properly select correct user samples and to adequately train them to manage also advanced functions of a Museum Web site [3]. Furthermore, it is difficult, in a limited amount of time, to reproduce actual situation of usage. Failures in creating real-life situations may lead to "artificial" conclusions rather then realistic results [6]. Therefore, user-based methods are considerable in terms of time, effort and cost.

Usability Inspections methods is the generic name for a set of methods based on having expert evaluators inspect or examine usability-related aspects of a user interface [7]. With respect to user-based evaluation, usability inspection methods are more subjective, having heavy dependence upon the inspector skills [3]. Their main advantage is the relationships between costs and benefits. In fact, to perform usability inspection does not require any special equipment and the inspector alone can detect a wide range of usability problem.
Generally speaking, we do not find neither among *usability inspection methods* nor among *user-based methods,* a technique fully suitable for museum web sites. Some methods are too "general"; so that they don't provide adequate feedback to help the designers fix the identified problems, while others are too complicated to be applied directly by the museum staff.

MILE METHODOLOGY: A SYSTEMATIC APPROACH TO USABILITY

MiLE (Milano-Lugano Evaluation method) is a new methodology for evaluating the usability/quality of web sites, fruit of a common research carried on by the Politecnico di Milano and the University of Italian Switzerland. It represents one of the most innovative and efficient approaches for evaluating the quality of a web application. MiLE, an extension of SUE usability inspection method [3], is based upon a combination of Inspection (i.e. an expert evaluator, systematically exploring the application) and Empirical Testing (i.e. a panel of end users actually using the application, under the guidance and the observation of usability experts). If this combination of the two methods is not new (several usability methods propose, in fact, a similar combination), the innovation of MiLE comes from the set of guidelines being used for making both inspection and empirical testing more effective and reliable. The inspections carried on using MiLE is task-based, that is, based on the performance, by the expert evaluator and then by the users during the empirical testing, of some "actions" in view of achieving a specific goal. MiLE faces up to the web site inspection emphasizing the need for separating different levels of analysis: navigation, content, graphical design, cognitive aspects and technology. For each level MiLE we have a list of "tasks", that is, generic actions (generic in that they can be applied to a wide range of applications) capable of leading the inspector through the maze of the different parts an application is made of. MiLE, in fact, provides inspectors with some guidelines that draw their attention to the most relevant features of the application.
In extreme synthesis, MiLE introduces two specifics heuristic concepts:

1. *Abstract Tasks*, ATs in short, used for inspection. They are a list of generic actions (generic in that they can be applied to a wide range of applications) capable of leading the inspector through the different parts and levels of the application;
2. *Concrete Tasks*, CTs in short. They are a list of specific actions (specific in that they are defined for a concrete application) that users are required to perform, while exploring the application during the empirical testing.

When it comes to the content's level, where communication issues are stronger, the concept of task is replaced by the more comprehensive concept of *scenario*. In fact, in order to effectively support the evaluation process, MiLE provides the inspector with a reusable set of evaluation tools (U-KIT, the Usability evaluation Kit) depending on the specific domain. The U-KIT is a library of scenarios that help to understand *stories about* use [8]. In fact, a scenario is the description of a concrete episode of use of the application [9]. It is possible to synthetize the concept of user scenario as follow:

User scenario = User profile + Abstract

Obviously, it is unfeasible to define all the scenarios we'd need to cover in details the whole spectrum of potential tasks that could be performed with the application. The evaluator will have to define the most relevant users' scenarios for the specific application he has to evaluate, trying to elicit the site's goal by interviewing the site's stakeholders, that is, all those who have an interest in the web-site: the designers, the institution, the users, the sponsors, the competitors, etc. He then selects from MiLE's list of scenarios ready-made for the application's domain those that fit its purpose best; he may also create new tasks "specially tailored" for the site he's facing, if he feels it necessary.

MILE: THE EVALUATION PROCESS

MiLE evaluation process is divided into five phases:

1. *Modeling the application under inspection*
2. *Performing some selected tasks*
3. *Evaluating the tasks through usability attributes*
4. *Weighting the results according to user profiles and communication goals*
5. *Empirical testing (user testing)*

We will now present of the above phases through its goal, activity and expected output.

1. Modeling the application under inspection
Goal: to give an overview of the web site and to identify the critical areas of the application relevant for a usability evaluation.
Activity: The reviewer draws a high-level mental model – either totally informally or adopting a semi-formal model– of the application under inspection.
Expected output: General schema of the most relevant features of the level under inspection; for ex., the content's structure, the navigational capabilities offered, or the interface elements.

2. Performing the selected tasks
Goal: Assessing the feasibility of some "critical" tasks.
Activity: according to salient user scenarios, the reviewer defines a set of tasks and performs them on the site. For each task, the reviewer assesses whether or not it can be properly accomplished.
Expected output: a task list and a two-value mark for each task (YES: it can be accomplished, NO: impossible to accomplish).

3. Evaluating abstract tasks through usability attributes
Goal: performing the inspection through scoring activity.
Activity: the inspector has to check a list of attributes concerning the different facets of usability/quality (e.g. richness, completeness, etc.). For each attribute (in relation to a specific task), a score must be given.
Expected output: a task matrix which reports the scoring (of each attribute) and the result obtained by every task.

4. Weighting the results according to user profiles and communication goals
Goal: establishing the "real quality" of each critical task with respect to their relevance.
Activity: after the scoring phase is over, the set of collected scores is analyzed through "weights" which define the relevance of each attribute for a specific goal (or, technically speaking, for a "user scenario"). Weighting allows a clean separation between the "scoring phase" (use the application, perform the tasks, and examine them) from the "evaluation phase" in a strict sense, in which the applications' and the stakeholders' goals are considered.
Expected output: a final matrix that shows the overall results obtained by every task. This matrix reports the results according the goal and the requirements of the application.

5. Empirical testing (user testing)
Goal: to empirically verify the most critical tasks identified during the inspection activity.
Activity: to carrying out the user testing in a usability lab. The user accomplishes several critical tasks and reports the results obtained. An inspector controls that the user testing is carried out correctly and gathers the impressions, the satisfaction and the problems of the users (by means of questionnaires, interviews, etc...).
Expected output: A final usability report that shows the results obtained during the user testing.

MUSEUM EVALUATION KIT (U-KIT FOR MUSEUM WEB SITES)

So far, the U-KIT for Museum Web sites is composed approximately of:

- 140 tasks
- 16 attributes
- 20 user profiles

The tasks' list and the user profiles have been defined interviewing several stakeholders involved in the development of Museum application and analysing a large number of Museum Web sites [10].
The attributes are not strictly dependent on the application's domain; rather, the attributes are divided according to the level of analysis (navigation, content, graphic, cognitive aspects).
Following is reported a part of the U-KIT for Museum Web sites (see table1):

MiLE: U-KIT for Museum Web sites		
TASKS LIBRARY		
TASK NUMBER	**DESCRIPTION**	
1	Find information about physical address of museum	
2	Find the city's map and/or area where the museum is located	
3	Find the charge of the ticket	
4	Find information about guided tour and/or special guided tours (special events)	
5	Find information about organisation of events (shows, concerts, etc.) within the "real-museum"	
6	Find information about history of museum collections	
7	Find information about didactic activities organised by museum	
8	Find information about Press-office material	
USER PROFILES		
U. PROFILE NUMBER	**PROFILE**	**DESCRIPTION**
1	High-school teacher	He surfs the site to find content which can help him for preparing a visit to the real museum.
2	Art-lover	He visits regularly the real museum. Exploring the virtual museum he searches information about current exhibitions and the calendar of events.
ATTRIBUTES		
ATTRIBUTE NUMBER	**NAME**	**DESCRIPTION**
1	Accessibility	The information is easily and intuitively accessible
2	Completeness	The user can find all the information required
3	Currency	The time scope of the content's validity is clearly stated. The info is updated.

Table 1: short example of U-KIT for museum websites

CONCLUSIONS

The general distinctive features introduced by MILE can be synthesized as it follows:

- Efficient combination of inspection and empirical testing
- Use of Tasks as guidelines for inspection;
- Use of Attributes, as a way to detail scoring;
- Use of weights, as a way to translate scores into evaluation;
- Use of user profiles in order to assign weights;
- Creation of a U-KIT (Usability Kit) specially tailored for Museum web sites;

In conclusion, it is possible to appreciate the method on a number of issues:

a) The level of detail: we evaluate a specific task and score/ weight a detailed list of attributes for each; a global evaluation of the site comes only as a final result. In this way we have two advantages: we can give to the application's designers a very precise feedback, so that they can efficiently correct faults and defects (for ex, with respect to a collection's description, we may find out that the text is good, but hard to find in the application); we can easily compare different inspectors' results of the same task (for ex, if there are discrepancies in the results, we can see which are the "crucial" attributes on which the scores and/ or weights diverge and investigate the reasons);
b) Through weights it is possible to take into account the specific objectives for the (portion of the) application.
c) Global concise evaluation can be obtained by combining all the micro-level evaluation processes; we have to compare the tasks' results and assign a weight to each of them. In this way we can obtain both a score and a weighted score either for a specific section or for the whole site. We could also extrapolate all the scores obtained by the attributes considered while performing the tasks, and assign a weight (that is, again, a relevance!) to each, and see whether the whole section/ application is "rich" or "accessible" and how important it is to be so.
d) Given the scoring, we could apply to it different systems of weights, in order to take into account different user profiles.

References

[1] Nielsen, J., *Designing Web Usability*, New Riders, 1999.

[2] Brinck, T., Gergle, D., Wood, S.D., *Usability for the web*, Morgan Kaufmann, 2002.

[3] Matera, M. et al., *SUE Inspection: An Effective Method for Systematic Usability Evaluation of Hypermedia*, IEEE Transaction, Vol.32, No. 1, January 2002.

[4] Whiteside J., Bennet J., and Holtzblatt K., *Usability engineering: Our experience and evolution*, in Handbook of Human-Computer Interaction, M.Helander, Ed. Amsterdam, The Netherlands, North-Holland, 1988, pp.791-817.

[5] Dix A., Finlay J., Abowd G., and Beale R., *Human Computer Interaction*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1998

[6] Lim K.H., Benbasat I. and Todd P.A., *An experimental investigation of the interactive effects of interface style, instructions, and task familiarity on user performance*, ACM Trans.Comput.-Hum.Interact., vol.3, no1,pp1-37, January 1996.

[7] Nielsen J., Mack R., *Usability Inspection Methods*, Wiley 1994.

[8] Carroll, J., *Making Use – Scenario-based design of Human-Computer Interactions*, MIT Press, 2002.

[9] Cato, J., *User-Centred Web Design*, Addison Wesley, 2001.

[10] Di Blas N., Maria Pia Guermandi, Carolina Orsini, Paolo Paolini, *Evaluating the Features of Museum Web Sites*, in Bearman D. & Trant J. (eds), (2002) Museums and the Web 2002. Selected Papers from an International Conference, Archives & Museum Informatics, Pittsburgh, USA: 179-185.