

Evaluation methodologies and user involvement in user modeling and adaptive systems

Cristina Gena

Dipartimento di Informatica, Università di Torino
Corso Svizzera 185, 10149 Torino, Italy
Email: {cgena@di.unito.it}

Abstract. In this paper I would like to present the main issues and conclusions of my PhD thesis, which has been dedicated to the problem of evaluation and testing of user modeling and adaptive systems. Regular and less explored approaches are taken into account and new perspectives and example are here sketched.

1. Introduction

My PhD thesis [4] faced with the issue of evaluation and user involvement in development of user modeling and adaptive systems. During the last few years the international community has underlined the importance of evaluation for a more user-centered approach to these systems [2]. Nevertheless, significant evaluations are not so frequent. Thus, the goal is now making the testing a common practice in the development of such systems and carrying out evaluations in every design phase to report significant results.

In this paper I would like to sketch how regular (Section 2) and less explored (Section 3) methodologies can be applied to the evaluation and the development of user modeling and adaptive systems in order to achieve fruitful results. Section 4 concludes the paper.

2. HCI and information selection process evaluations

The methodologies for evaluating user-adapted systems are generally borrowed by the methodologies used in HCI and by those ones exploited for the evaluation of the information selection process (mainly derived from evaluation in information retrieval systems). The former ones can be classified in *i*) collection of user's opinion, *ii*) user observation and monitoring, *iii*) predictive evaluation, *iv*) formative evaluation, *v*) summative evaluation. For more details see, [8]. The latter ones encompass metrics such as precision and recall, training set and test set, evaluation of the ordering, coverage, MAE and RMSE, reversal rate and sensitivity measures, etc. For more details see [5].

All these methodologies can be exploited during the different design phases of an user-adapted system following a user-centered design approach. However, the consequences are different from those ones achieved in regular HCI systems. First of all, evaluation methodologies should be applied following a layered approach, since, as described in [1], user-adapted systems need an evaluation that differentiates, at least, problems concerning content adaptation and interface adaptation. For instance, in evaluation of a system that provides personalized tourist information onboard cars [4], we evaluated separately *i)* the effectiveness of adaptations by calculating the distances between the user choices and the system recommendations and *ii)* the correctness of interface solutions by means of an usability test. In my thesis I also proposed a layered approach following the three tasks in which Kobsa et al. [6] divide personalized hypermedia applications: *i)* acquisition method and primary inferences, *ii)* representation and secondary inferences, *iii)* adaptation production. For each tasks I proposed the most appropriate methodologies. For details see [4].

This is not a new idea. A similar approach has been already proposed by Paramithys et al. [7]. It is important to underline in this context that, as Paramithys et al. noticed, since the concept of the typical user of a system cannot be applied in adaptive systems, each evaluation has to take into account a particular user having characteristics encoded in some type of user profile and in a particular context of use.

Then, I would like to notice that evaluation methodologies of user-adapted systems can be also used as knowledge sources for the development of adaptive applications. Therefore, the exploitation of evaluation techniques lead to a *generative approach* in the development of these systems. For instance, task analysis can be used to analyze not only the way people perform their jobs, but also how different kinds of users perform their jobs and then modeling the interaction on the basis of these different models. In case of heuristic evaluation, we need to know if an interface works not for a generic user, but how can work for different kind of users. Then, we can design the interface on the basis of the experts' suggestions about these users.

3. Qualitative approaches

The exploitation of qualitative methodologies to the evaluation of user modeling and user-adapted systems is not a new idea, even if quantitative methodologies are largely applied, in particular controlled lab experiments exploiting quantitative metrics.

Preece et al. [8] classify the main qualitative methodologies under the umbrella term *interpretative evaluation*, which can be summed up as “spending time with users”. The interpretative evaluation comes in these flavors: *i)* contextual inquiry, *ii)* cooperative and participative evaluation and *iii)* ethnography.

Facing with qualitative evaluation, I gained insightful inspirations from the theory of *embodied interaction* of Paul Dourish [3]. Dourish defines the embodied interaction as “an interaction with computer systems that occupy our world, a world of physical and social reality, and that exploit this fact in how they interact with us”. While traditional computational model of HCI has been rationally built on a procedural foundation and set out its account of the world in term of plans, procedures, tasks and goals, Dourish's model of HCI places interaction at center of

the picture. By this, he considers interaction not only as *what* is being done, but also *how* it is being done. Therefore, in opposition to the dominant cognitive approach to HCI, based to a goal-driven approach to the human-machine interaction, Dourish resumes both sociology (in particular qualitative-based approaches such as ethnomethodology) and phenomenology for an approach to HCI more oriented to the embedded interaction between human and machine.

Evaluating systems in a qualitative way requires a fewer number of users than quantitative research. Qualitative researchers sustain that qualitative methodologies allow a deeper knowledge of the subjects involved that compensate the less representative sample. In fact, what this kind of methodologies can offer is a more accurate knowledge of the real behavior of a user sitting in front of an interactive system compared to the artificial situation of lab environment.

In case of user-adapted systems, this information can be used *i)* during the development of a user-adapted system by singling out, for instance, the dimension for modeling the users and *ii)* for a system revision after the evaluation. The problem concerns the difficulty of modeling an interaction by taking into account a “situated action” instead of a plan predetermined by “search space” of goals and actions.

The two perspectives seem to be opposite, but I would like to find some point of contact. In particular, my question is how qualitative and embodied approaches can be applied to user-adapted systems.

As Dourish emphasized [3], under a phenomenological point of view, we reach the meaning by acting in the world, so in case of user-machine interaction a subject reaches the knowledge about the system by using the system and interacting with and through it. And also for “the system”, the subject becomes meaningful during the interaction. Another key point of the phenomenological perspective is that experience and interaction come before meaning, while the Cartesian view considers action arising from meaning as the expression of internal mental states. So, the way things are organized shapes our understandings of those things.

A logical conclusion of these observations could be constructing the Knowledge Base of a user-adapted system by observing the real user-system interaction, since we cannot predict the user behavior until she experienced the system. Nevertheless, there are some ways to model the user in advance. The user model could be originated, for instance, by the observation of real users interacting with similar systems or with the system to model, if already implemented in a non-adaptive version. Therefore, interpretative techniques such as contextual inquiry, cooperative and participative evaluation and ethnography can be applied to monitor and to have feedback from the user in every design step.

To extract relevant user dimensions we could, then, analyze work practices and look for common patterns emerging from different users’ actions. On the basis of existing correlations between users and practices, information on how users understand the system can be then exploited to model the users.

To offer personalized recommendations, instead, we could build the system’s knowledge base by monitoring the user choices. Then, we could propose the system’s recommendations to the users and asking them to evaluate such proposals and discussing with them about their choices. Finally, revising the Knowledge Base on the basis of user feedback.

Following Dourish's advices [3], the designer should focus on the ways the user understands the tool and how she uses it in each situation instead of designing the ways to use the artifact. The consequence, in user user-adapted systems, is making the user aware of how adaptation works and adapt the ways in which these facilities are presented on the basis of the user model, since different users understand the tool in different ways. Users having a background in Computer Science have a different approach in understanding how an adaptive system works, compared, for instance, to users having a background in Literature. So, also in this case different suggestions tailored on the basis on user profile are necessary.

In conclusion, the key points of qualitative and embodied approaches are *i*) the importance of user observation in her real context (social, cultural, organizational); *ii*) gathering field data and studying working settings; *iii*) the importance of usage studies that point out the unexpected uses of technology that the designers had never intended; *iv*) the attention to user practices and practices shared in the communities; *v*) user involvement and user participation in the system design; *vi*) the link between meaning and experience.

3. Conclusion

To conclude this brief excursus, I advocate, of course, the importance of evaluation and testing in every design phase of a user-adapted system. Significant testing results can lead to more appropriate and successful systems. According to my point of view, both quantitative (in term of controlled experiments) and qualitative methodologies of research can offer fruitful contributions.

First of all, however, in both cases it is important to correctly carry out the evaluation of the system. In fact, the problem of most evaluations is the non-significance of the results and therefore the absence of generalizations.

Then, the choice between quantitative and qualitative methodologies depends on the point of view of the evaluation: while quantitative research tries to explain the variance of the dependent variable generated through the manipulation of independent variables (variable-based), in qualitative research the object of study becomes the individual subjects (case-based). Qualitative researchers sustain that a subject cannot be reduced to a sum of variables and therefore a deeper knowledge of a fewer group of subjects is more fruitful than an empirical experiment with a representative sample. The goals of the analyses are also different: while quantitative researchers try to explain the cause-effect relationships, qualitative researchers would like to comprehend the subjects under study by interpreting their point of views. While quantitative research tries to explain why subjects behave in a particular way, qualitative research tries to explain how subjects behave in that particular way.

Concerning the methodologies of analysis, quantitative researchers try to validate a theory by falsification, while qualitative researchers try to individuate the so-called *ideal types* (e.g., the behavioral patterns of different kinds of web users, useful to elicit the more suitable interface) through the description and the classification of the collected empirical data and the individuation of typology dimensions. The ideal types are conceptual categories useful to interpret the reality under observation. In

case of user modeling systems, these categories can offer cues to model the features of the users and then adapt the system to these features.

Indeed, the choice between quantitative and qualitative methodologies is not trivial and depends on the aims and the purpose of the evaluation. If we would like to test the impact of different adaptation techniques in a given interface, a controlled experiment can produce useful results. For instance, I tested the different impacts of adaptation techniques applied to a web site by means of a controlled experiment with two different conditions (users interacting with the adaptive version of the site vs. user interacting with the non adaptive-version). See [4] for details. Instead, to discover the features of an interaction useful to model the user, observing users in context and interview them can offer material to build the user model categories. For instance, in a project we are carrying out concerning the design of a web site offering technical information at different levels of details (depending on the different background of the user), we are collaborating with domain expert and final users in a iterative *design-test-redesign* process projects to define the different levels of interaction for given users (e.g., expert, non-expert, etc..).

To sum up, if we would like to discover new categories to model an interaction, a qualitative approach can produce more fruitful results, while if we would like to investigate the impact of already known variables, a quantitative approach can be preferred. However, this is not a rule and both approaches can offer interesting results.

References

1. P. Brusilovsky, C. Karagiannidis, and D. Sampson, *The benefits of layered evaluation of adaptive applications and services*, Proceedings of workshop at the Eighth International Conference on User Modeling, UM2001, 2001, 1-8.
2. D. Chin. Empirical evaluation of user models and user-adapted systems. In *User Modeling and User-Adapted Interaction 11* (2001), pp 181-194.
3. P. Dourish, *Where The Action Is: The Foundations of Embodied Interaction*, MIT Press, 2001.
4. Gena, Evaluation methodologies and user involvement in user modeling and adaptive systems, Phd Thesis, Università di Torino, 2003.
5. N. Good, J. B. Schafer, J. A. Konstan, A. Botchers, B. M. Sarwar, J. L. Herlocker, and J. Riedl, *Combining collaborative filtering with personal agents for better recommendations*, Proceedings of the Sixteenth National Conference on Artificial Intelligence, 1999, 439-446.
6. A. Kobsa, J. Koenemann, W. Pohl, *Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships*, The Knowledge Engineering Review 16(2), 2001, 111-155.
7. A. Paramythi, A. Totter, and C. Stephanidis, *A Modular Approach to the Evaluation of Adaptive User Interfaces*, Proceedings of workshop held at the Eighth International Conference on User Modeling in Sonthofen, Germany, July 13th, 2001, 9-24.
8. J. Preece, Y. Rogers, and H. Sharp, *Interaction Design: Beyond Human-Computer Interaction*, New York, NY:Wiley, 2002.