Development of a Methodology for Evaluating the Quality in Use of Web 2.0 Applications*

Tihomir Orehovački

University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
tihomir.orehovacki@foi.hr

Abstract. Quality in use is comprised of two seemingly different though interlocking concepts: usability and user experience. Consequently, complementary evaluation of pragmatic and hedonic attributes could significantly affect the acceptance of software applications. However, in the context of Web 2.0 applications this topic has still not attracted enough attention from the HCI community. Therefore we present a research aimed at developing a methodology that would facilitate the analysis and comparison of evaluated Web 2.0 applications.

Keywords: Web 2.0, Quality in Use, Usability, User Experience, Subjective and Objective Measures, Evaluation Methodology.

1 Introduction

Quality in use is considered to be one of the most important factors that affect a wide acceptance of software applications. It is comprised of two seemingly different but rather complementary concepts [6]. Usability refers to the product-centred evaluation of pragmatic attributes through the use of both subjective and objective measuring instruments. On the other hand, user experience (UX) consists of applying subjective measuring instruments on the assessment of hedonic attributes [5]. HCI literature contains a number of different models, methods, and standards aimed at evaluating the quality and usability of software applications like information systems, web sites, web portals and alike. However, the existing research related to the evaluation of Web 2.0 applications in general is fairly scarce. We strongly believe that this is accounted for by the inadequacy of current approaches for their evaluation, as confirmed by several recent studies. According to the research in Hart et al. [4], the popular social networking site Facebook should not be an example of a success story because it complies with only two of ten heuristics originally proposed by Nielsen [7]. Hart et al. also reported that attributes such as ease of use, usefulness and playfulness have a major impact on user behavioural intentions. A study into usability evaluation [11] revealed that YouTube had also met only two heuristics. Thompson and Kemp [14]

---

argued that the emphasis on user experience is one of the main reasons why Web 2.0 applications such as Flickr, Wikipedia and YouTube have a large number of active users. Moreover, they extended the Nielsen's set of traditional heuristics with two additional heuristics although the validity of that modified set has not been empirically confirmed. The findings stated above motivated us to initiate a research on the development of an evaluation methodology that would extend the concept of quality in use in the context of Web 2.0 applications.

2 Research Questions

In line with the above, we outline the following research questions:

a. What are the most important subjective and objective quality attributes that contribute to users' satisfaction and loyal behaviour?
b. How can these quality attributes be measured? In particular, are the existing measuring instruments suitable for evaluating the quality in use of Web 2.0 applications?
c. Can these quality attributes be translated into a set of heuristics that would facilitate the assessment of Web 2.0 applications?
d. Can these quality attributes be converted into a set of guidelines that would contribute to improved design of Web 2.0 applications and their testing?
e. Which methods, techniques and procedures should constitute an evaluation methodology that simplifies an analysis and comparison of Web 2.0 applications?

3 Methodology

Our research is divided into three parts: theoretical background, pilot study and main research. Until now, the theoretical part and the pilot study have been completed.

In the theoretical part, a literature review was prepared according to the methodology proposed by Webster and Watson [15]. An analysis of recent relevant research in the field of web quality and usability assessment resulted in a set of attributes that may have a significant role in the evaluation of the quality in use of Web 2.0 applications [10]. Based on their theoretical similarities, the attributes were initially placed into six different categories: system quality, service quality, information quality, performance, effort, and acceptability. In addition, based on the type of the measuring instrument to be used for data collection (a questionnaire or a tool for logging actual use), the attributes were grouped into subjective and objective quality attributes. The validity and reliability of questionnaire items was evaluated according to the guidelines suggested by Straub et al. [12].

The aim of the pilot study was to test the suitability of the measuring instruments for evaluating the quality in use of Web 2.0 applications. The participants in the pilot study were students who used Web 2.0 applications to perform various educational

activities. In the first experiment, two methods were employed in order to gather both quantitative (questionnaire) and qualitative (retrospective thinking aloud, RTA) data about the perceived quality of cloud based applications for collaborative writing [9]. In the second experiment [8], the research methodology was extended with the logging actual use method. The subjective assessment of Web 2.0 applications for mind mapping was conducted with the use of a questionnaire and RTA while the objective data about mouse movements, mouse clicks, keystrokes and time on task were gathered by means of the Mousotron tool [1]. An analysis of the data obtained from both experiments revealed a list of matching attributes that influence users' satisfaction and loyal behaviour regardless of the type of application used, including ease of use, effectiveness, reliability, interactivity, customizability, and navigability.

The remaining part of the research is structured as follows. Firstly, two different groups of experts (HCI researchers and web developers) will categorize quality attributes using the Q-sorting method [13]. Secondly, a scenario-based evaluation of Web 2.0 applications with different functionalities will be conducted. Data will be collected using a questionnaire (perceived quality in use) and a tool for logging actual use (estimated quality in use). Participants in the study will be recruited from different departments of two Croatian universities. The data obtained from the experts will be used in the development of a conceptual model that will be composed of two sub-models: a measurement and a structural model, respectively. The former will take into account relationships between each latent variable (category) and corresponding manifest variables (attributes), while the latter will take into account the relationships among the latent variables. The validation of the conceptual model will be carried out with Partial Least Squares (PLS) path analysis in accordance with the catalogue of criteria discussed in [3]. Measurement models will be assessed by means of content validity, indicator reliability, construct reliability, convergent validity and discriminant validity. On the other hand, the structural model will be evaluated in terms of its effect size and predictive validity. Based on the outcomes of the Q-sorting and the PLS analysis, a tree with logical relationships between categories and attributes will be specified and elementary criteria for each attribute will be defined. The final estimate (score) of quality in use will be calculated using a stepwise process of logic aggregation of preferences [2].

4 Expected Contributions

The scientific contribution of the research is two-fold. From a theoretical perspective, the knowledge and experience in the field of quality and usability evaluation of Web 2.0 applications will be systematized. The most important quality attributes that significantly affect users' satisfaction and loyalty will be identified. Moreover, it will be determined whether the weight of each attribute varies depending on the Web 2.0 application type and the context of use.

From a practical point of view, the new methodology will offer a number of benefits. Firstly, it is expected to enable the evaluation of Web 2.0 applications using both subjective and objective measuring instruments, which may alleviate possible shortcomings of their separate use. Secondly, through the combination of several data

collection methods, the process of detection of problems in Web 2.0 application usage will be facilitated and accelerated. It is noteworthy that, apart from the HCI experts and users, web developers (programmers, designers, architects) will also take part in the research (which so far, to our knowledge, has been rather rare). Thirdly, based on the results obtained from the practical use of the methodology, a new set of heuristics aimed for the evaluation of Web 2.0 applications will be created. In addition, the research will result in a series of guidelines and recommendations for the development of Web 2.0 applications. Finally, the afore-described methodology will enable the "transformation" of a lot of quantitative data into a single score, which would facilitate the analysis and comparison of evaluated Web 2.0 applications.

References

1. Blacksun Software, http://www.blacksunsoftware.com/mousotron.html (2011)
2. Dujmović, J.J.: Preferential Neural Networks. In: Antognetti, P., Milutinović, V. (eds.) Neural Networks – Concepts, Applications, and Implementations, pp. 155--206. Prentice Hall, New Jersey (1991)
3. Esposito, V.V., Chin, W.W., Henseler, J., Wang, H.: Handbook of Partial Least Squares: Concepts, methods, and applications. Springer, Berlin (2009)
4. Hart, J., Ridley, C., Taher, F., Sas, C., Dix, A.: Exploring the Facebook Experience: A New Approach to Usability. In: 5th Nordic conference on Human-computer interaction: building bridges, pp. 471--474. ACM, Lund (2008)
5. Hassenzahl, M., Tractinsky, N.: User experience – a research agenda. Behaviour & Information Technology. 25 (2), pp. 91--97 (2006)
6. ISO/IEC 25010:2011. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models (2011)
7. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods. John Wiley & Sons, New York (1994)
8. Orehovački, T., Granić, A., Kermek, D.: Exploring the Quality in Use of Web 2.0 Applications: The Case of Mind Mapping Services. Lecture Notes in Computer Science (accepted for publishing)
9. Orehovački, T.: Perceived Quality of Cloud Based Applications for Collaborative Writing. In: Pokorny, J., et al. (eds.) Information Systems Development – Business Systems and Services: Modeling and Development. Springer, Heidelberg (in press)
10. Orehovački, T.: Proposal for a Set of Quality Attributes Relevant for Web 2.0 Application Success. In: 32nd International Conference on Information Technology Interfaces, pp. 319--326. IEEE Press, Cavtat (2010)
11. Silva, P.A., Dix, A.: Usability – Not as we know it! In: 21st British HCI Group Annual Conference on HCI 2007: People and Computers XXI: HCI...but not as we know it – Volume 2, pp. 103--106. ACM, University of Lancaster (2007)
12. Straub, D., Boudreau, M., Gefen, D.: Validation Guidelines for IS Positivist Research. Communications of the Association for Information Systems. 13 (1), 380--427 (2004)
13. Thomas, D.M., Watson, R.T.: Q-Sorting and MIS Research: A Primer. Communications of the Association for Information Systems. 8, 141--156 (2002)
14. Thompson, A-J., Kemp, E.A.: Web 2.0: extending the framework for heuristic evaluation. In: 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, pp. 29--36. ACM, New Zealand (2009)
15. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Quarterly. 26 (2), xiii--xxiii (2002)