Towards a Multi-Modal Methodology for User-Centred Evaluation of Adaptive Systems

Alexander Nussbaumer
alexander.nussbaumer@tugraz.at
Graz University of Technology
Graz, Austria

Christina M. Steiner
christina.steiner@tugraz.at
Graz University of Technology
Graz, Austria

Owen Conlan
owen.conlan@scss.tcd.ie
Trinity College Dublin
Dublin, Ireland

ABSTRACT

This paper presents an approach and methodology for user-centred evaluation of adaptive systems. In contrast to layered evaluation approaches that decompose adaptation into its constituents, our approach conceptualises the quality and benefit for the user into separate evaluation qualities for a comprehensive and multifaceted evaluation. Instruments of different modalities are used to measure these qualities from the user perspective. A service is presented that takes up this approach and enables time- and cost-efficient evaluation by defining and re-using evaluation qualities and instruments, as well as collecting and analysing data based on these definitions. This approach allows to compare different adaptive systems by using the same qualities and adapting the instruments to the specific characteristics of the particular adaptive system.

CCS CONCEPTS

• **General and reference** → *Evaluation*; • **Human-centered computing** → *HCI design and evaluation methods.*

KEYWORDS

evaluation, adaptive systems, evaluation qualities, evaluation model

ACM Reference Format:
Alexander Nussbaumer, Christina M. Steiner, and Owen Conlan. 2019. Towards a Multi-Modal Methodology for User-Centred Evaluation of Adaptive Systems. In *27th Conference on User Modeling, Adaptation and Personalization Adjunct (UMAP'19 Adjunct), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3314183.3323681

1 INTRODUCTION

Several frameworks and approaches have been elaborated in the past, how adaptive systems can be evaluated. For example, different variants of layered evaluation have been proposed, where an adaptive system is decomposed into its main functions and each of them is evaluated individually [1][5]. Other aspects of evaluating adaptive systems, such as empirical evaluation, user-centred evaluation, or goal attaining, are proposed in overview papers [2][7].

In previous work we have elaborated a user-centred approach grounding on the definition of evaluation qualities that are key to the use of an adaptive system. This approach has been applied on the digital library system CULTURA that serves as an adaptive information system for historians. Relevant qualities have been defined and evaluated individually including usability, user acceptance, adaptation quality, visualisation quality, and content usefulness [6].

UMAP'19 Adjunct, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6711-0/19/06…$15.00
https://doi.org/10.1145/3314183.3323681

2 EVALUATION QUALITIES

Evaluation qualities are the key constructs to conduct the user-centred evaluation and to compare adaptive systems and personalisation approaches. They define what should be evaluated and give information about the quality of an adaptive system from a user's perspective. Based on previous work [7][6][4], the following list presents suggestions of evaluation qualities relevant for evaluating and comparing adaptive systems.

Appropriateness. This quality indicates how well an adaptation meets the needs and preferences of the user.
Timeliness. Timeliness is the degree to which the adaptation is done in the right moment.
Personalisation benefit. This quality refers to the user's perceived benefits of being presented with personalised content and information as opposed to non-personalised content.
Purpose. Purpose is the degree how well the system works for the purpose (e.g. learning outcome or goal attainment)
Balance. This quality indicates if the adaptation provides an appropriate balance between guidance and freedom.
Adaptability. This quality indicates if the user can sufficiently influence the behaviour and presentation.
Scrutiny. Scrutiny indicates whether appropriate insight into the user model and adaptation technique is revealed.
Privacy. Privacy refers to users' perceptions of how well personal information is kept private.
Usability. This quality describes the overall usability of the system and its personalised content.
User acceptance. This quality describes users' overall acceptance of the system.

3 EVALUATION MODEL

In order to measure evaluation qualities, three basic types of instruments are defined that provide evidence for a quality (Fig. 1). First, questionnaires are the traditional way of capturing data about the users' opinions after using a system. Second, an immediate feedback instrument is integrated in the system that allows users to provide ratings or comments via so-called *judgets* while using the system. Third, log data of a user's interactions with a system are collected and analysed with respect to the evaluation qualities (so-called *sensors*).

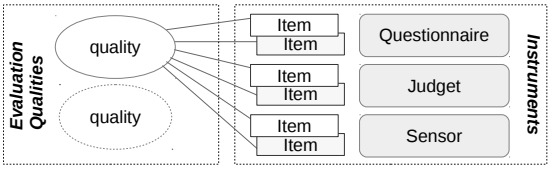

Figure 1: Evaluation model including qualities and instruments

These different types of instruments represent two different dimensions of data collection modalities (see Fig. 2). The *user engagement* modality consists of invasive vs. non-invasive data collection, meaning that the user is either explicitly asked (questionnaire and judget) or data is collected non-intrusively, with users not being interrupted in their tasks (sensor). The *task integration* modality is represented by continuous vs. non-continuous data collection, meaning that data is either collected at specified points in time (e.g. after system use; questionnaire) or during system use or performing a task (sensor, judget). Such multi-modal approach opens up new possibility of data analysis. For example, data collected by different instruments can be compared to each other. In our approach evaluation qualities and different types of instruments together are called *evaluation model* [3].

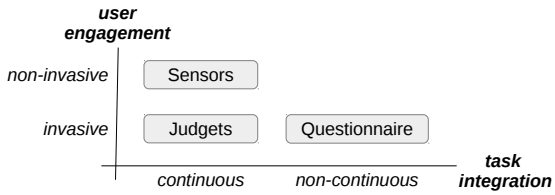

Figure 2: Multi-modal data collection approach

The analysis of questionnaire data is straightforward, following the manual of standard instruments and traditional approaches of opinion research. The analysis of sensor data (log data) needs much more attention. The overall idea to interprete interactions of the user with the adaptive system in relation to particular qualities. For example: *Appropriateness* could be measured by counting how often a user follows a recommendation (in relation to other viewed pieces of information). *Timeliness* could be measured how often a user follows a recommendation provided at the first time. *Personalisation benefit* could be measured by observing, if follow-up actions were taken that are related to a recommended information.

4 EVALUATION SERVICE

An evaluation service (called *Equalia*) has been created that supports the evaluation process including the definitions of the evaluation model (qualities and instruments), data collection, and data analysis [3]. Key features of this service are the formal definition and re-usability of evaluation models that are used to collect and analyse an adaptive system. Due to this flexible approach, it can be loosely integrated with any adaptive system. By using the same qualities and adapting the instruments to the particular features of a system, different adaptive systems can be compared. The overall architecture of the service is outlined in Fig. 3.

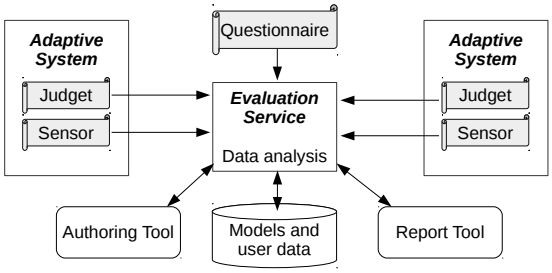

Figure 3: Integration of the evaluation service with the adaptive systems

5 CONCLUSION

This paper presents an approach to formalise the user-centred evaluation of adaptive systems. The key construct of this approach is a set of evaluation qualities that determine the overall quality of an adaptive system from the user perspective. Multi-modal instruments are defined and implemented to measure these qualities. This allows to compare different adaptive systems by using the same qualities, but potentially different instruments to measure them. An evaluation service is established that allows a cost-effective evaluation by re-using existing evaluation models.

ACKNOWLEDGMENTS

The work reported has been partially supported by the PROVENANCE project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825227.

REFERENCES

[1] Peter Brusilovsky, Charalampos Karagiannidis, and Demetrios Sampson. 2004. Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Life-Long Learning* 145, 4/5 (2004), 402–421. https://doi.org/10.1504/IJCEELL.2004.005729

[2] Catherine Mulwa, Séamus Lawless, Mary Sharp, and Vincent Wade. 2011. The evaluation of adaptive and personalised information retrieval systems: A review. *International Journal of Knowledge and Web Intelligence* 2, 2/3 (2011), 138–156. https://doi.org/10.1504/IJKWI.2011.044120

[3] Alexander Nussbaumer, Eva-C. Hillemann, C. M. Steiner, and D. Albert. 2012. An evaluation system for digital libraries. In *Proceedings of the 2nd International Conference of Theory and Practice of Digital Libraries (TPDL 2012). Lecture Notes in Computer Science*, Panayiotis Zaphiris, George Buchanan, Edie Rasmussen, and Fernando Loizides (Eds.), Vol. 7489. Springer, Berlin, Heidelberg, 414–419. https://doi.org/10.1007/978-3-642-33290-6_45

[4] Alexander Nussbaumer, Christina M. Steiner, Eva-C. Hillemann, and Dietrich Albert. 2014. Towards a Service for Evaluating Adaptive Learning Systems.. In *Proceedings of the 22nd International Conference on Computers in Education (ICCE 2014)*, Chen-Chung Liu, Hiroaki Ogata, Siu Cheung Kong, and Akihiro Kashihara (Eds.). Asia-Pacific Society for Computers in Education (APSCE), Jhongli City, Taiwan, 138––140. http://icce2014.jaist.ac.jp/icce2014/

[5] Alexandros Paramythis, Stephan Weibelzahl, and Judith Masthoff. 2010. Layered Evaluation of Interactive Adaptive Systems: Framework and Formative Methods. *User Modeling and User-Adapted Interaction* 20, 5 (Dec. 2010), 383–453. https://doi.org/10.1007/s11257-010-9082-4

[6] Christina M. Steiner, Maristella Agosti, Marc S. Sweetnam, Eva-C. Hillemann, Niccola Orio, Chiara Ponchia, Cormac Hampson, Gary Munnelly, Alexander Nussbaumer, Dietrich Albert, and Owen Conlan. 2014. Evaluating a Digital Humanities Research Environment: The CULTURA Approach. *International Journal on Digital Libraries (IJDL)* 15 (2014), 53–70. Issue 1. https://doi.org/10.1007/s00799-014-0127-x

[7] Christina M. Steiner and Dietrich Albert. 2012. *Tailor-made or Unfledged? Evaluating the Quality of Adaptive e-Learning.* Nova Science, New York, USA, 111–143. https://novapublishers.com/shop/evaluation-in-e-learning/