Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Towards an Evaluation Service for Adaptive Learning Systems

Alexander Nussbaumer[1*], Christina M. Steiner[1], Eva-Catherine Hillemann[1], Dietrich Albert[1,2]
[1]*Knowledge Technologies Institute, Graz University of Technology, Austria*
[2]*Department of Psychology, University of Graz, Austria*
*alexander.nussbaumer@tugraz.at

Abstract: This paper presents a service that supports the evaluation of adaptive learning systems. This evaluation service has been developed for and tested with an adaptive digital libraries system created in the CUTLURA project. Based on these experiences an approach is outlined, how it can be used in a similar way to evaluate the features and aspects of adaptive learning systems. Following the layered evaluation approach, qualities are defined that are evaluated individually. The evaluation service supports the whole evaluation process, which includes modelling the qualities to be evaluated, data collection, and automatic reports based on data analysis. Multi-modal data collection facilitates continuous and non-continuous, as well as invasive and non-invasive evaluation.

Keywords: Evaluation service, layered evaluation, adaptive learning systems

1. Introduction

Evaluation is an important task, because it reveals relevant information about the quality of the technology for all stakeholders and decision makers. It involves collecting and analysing information about the user's activities and the software's characteristics and outcomes. Its purpose is to make judgments about the benefits of a technology, to improve its effectiveness, and to inform programming decisions (Patton, 1990). The evaluation process can be broken down into three key phases: (1) Planning, (2) collecting, and (3) analysing (Cook, 2002). However, conducting evaluations is usually a time consuming task. Besides planning and data collection, it requires a lot of time to analyse the collected evaluation data. Including log data in the evaluation further increases the evaluation task and also makes it more complex. In order to address these aspects needed for a sound and systematic evaluation and to reduce workload for the evaluator, a holistic conceptual and technical approach has been created and based on that the evaluation service Equalia has been developed (Nussbaumer et al., 2012). This approach and the developed service has been tried out and tested in the context of the digital library project CULTURA (http://cultura-project.eu/). This paper suggests that this approach and service can also be applied to evaluate adaptive learning systems.

2. Evaluation Approach and Conceptual Design

In order to evaluate adaptive learning systems, Brusilovsky et al. (2004) propose a layered evaluation approach. Instead of evaluating a learning system as a whole, they suggest to evaluate the core components, which are user modelling and the adaptation decision making. This approach has been extended in the GRAPPLE project, where also other aspects are evaluated, for example usability, user acceptance, and adaptation quality (Steiner et al. 2012). A similar 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, visualization quality, and content usefulness (Steiner et al., 2013). Though applied in digital libraries, these aspects are also relevant in adaptive learning systems.

The general goal of the evaluation service is to support the whole evaluation process, consisting of planning the evaluation, carrying out the evaluation, as well as analysing the data and creating reports

(Fig. 1). The evaluation model is the core part of the conceptual approach. It allows for explicitly modelling what and how should be evaluated. Therefore, the evaluation model formally represents the evaluation approach and thus represents the evaluator's expertise. It consists of two parts. First, the quality model is an abstract model that defines what should be evaluated. It defines evaluation aspects (such usability (Brooke, 1996), user acceptance (Davis et al., 1989), or recommendation quality), which express the qualities of a system including its content. Second, the survey model defines the items for measuring these quality aspects. Items might be concrete questions, but can also be specifications, how tracking data covers the user's behaviour (for example, how often a user follows a recommendation).

Figure 1: Evaluation Process as supported by the Equalia evaluation service.

The data collection approach consists of two main aspects. First, the data collection instruments are based on the evaluation model and thus related to system qualities that should be evaluated. Second, three different types of instruments are defined (questionnaires, sensors, judgets) that allow data collection on different dimensions, namely invasive and non-invasive, as well as continuous and non-continuous data collection. Questionnaires are the traditional way of capturing data about the user's opinion. A different way of collecting evaluation data is realised with judgets, which are little widgets integrated in the system to be evaluated where users can give immediate feedback (e.g. ratings). Software sensors are instruments that establish a continuous and non-invasive evaluation method. Sensors are not visible to the users, but monitor and log the interaction and usage behaviour, and collect evaluation data in this way.

An important feature of the evaluation system is the generation of automatic reports from the collected evaluation data on the basis of the underlying evaluation model. A report is made upon a survey model by aggregating all participants' data related to the respective survey model. The data from different sources (questionnaire, judgets, sensors) are compared according their relations to quality aspects. Thus overall scores for each quality are calculated.

3. Application in Adaptive Learning Environments

In order to apply this approach and the Equalia service in adaptive learning environments, the most important step is to identify which qualities should be evaluated and how the measurement can be accomplished. Beside the aforementioned qualities, such as usability, usefulness, and content quality, in typical adaptive learning systems, the qualities outlined in Tab. 1 are of specific interest:

Quality Name	Explanation	Measurement
recommendation quality	how well are learning resources recommended to the learner	(1) ask the learner about appropriateness; (2) track how often learners follow a rec. resource
visualisation quality	how useful are visualisations for the learner	(1) ask learner about usefulness of vis. (2) track how often learner uses a visualisation
collaboration quality	how good is the collaborative support to interact with peers	(1) ask learner about benefit of coll. mechanisms; (2) track how often learner uses coll. features

Table 1: Possible evaluation qualities and measurement methods in adaptive learning systems.

Integration with an adaptive learning system is easy, because the Equalia service is an independent Web service that can be loosely coupled with the system to be evaluated (Fig 2). It exposes a REST interface to collect evaluation data from different sources. Judgets and sensors have to be integrated in the adaptive learning system capturing the user's opinion and monitoring the user's system interactions and sending them to Equalia. Furthermore, Equalia provides a Web interface for creating and managing the evaluation models, for generating questionnaires, and for creating evaluation reports.

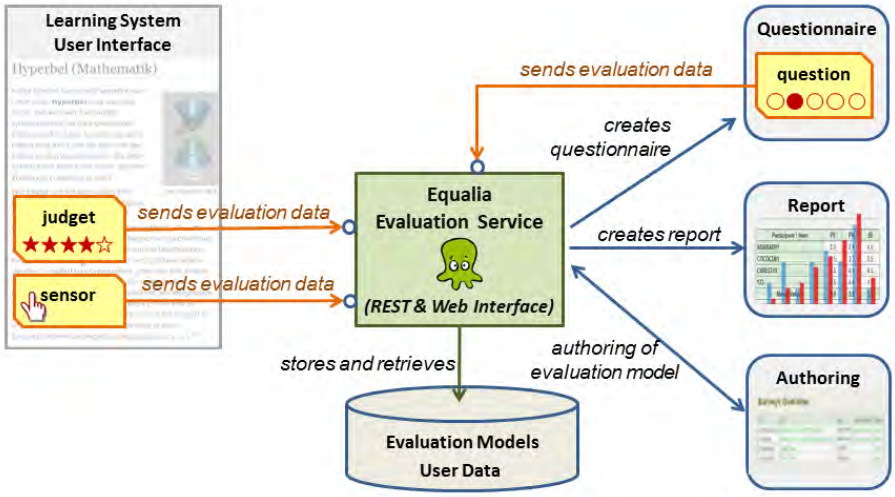

Figure 2: Integration architecture of Equalia with an adaptive learning system.

4. Conclusion and Outlook

This paper presented the evaluation service Equalia that can be used to support the evaluation of adaptive learning systems. Based on the experiences made with this service to evaluate an adaptive digital libraries system, we propose to use it in a similar way for adaptive learning systems. The most important necessary steps include the identification of the qualities to be evaluated, the injection of judget and sensor code (functionality) in the system to be evaluated, and the authoring of the survey model (questionnaires, judget questions, interaction behaviour). Then the evaluation can be conducted more or less automatically.

Acknowledgements

The work reported has been partially supported by the CULTURA project, as part of the Seventh Framework Programme of the European Commission, Area "Digital Libraries and Digital Preservation" (ICT-2009.4.1), grant agreement no. 269973.

References

Brooke, J. (1996). SUS: A "Quick and Dirty" Usability Scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester & a. L.McClelland (Eds.) Usability Evaluation in Industry, pp. 189-194. London: Taylor & Francis.

Brusilovsky P., Karagiannidis C., & Sampson D. (2004) Layered Evaluation of Adaptive Learning Systems. International Journal of Continuing Engineering Education and Life-Long Learning, 14, pp. 402-421.

Cook, J. (2002). Evaluating Learning Technology Resources. Retrieved from http://www.alt.ac.uk/docs/eln014.pdf on 1 June 2014.

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science, 35(8), pp. 982-1003.

Nussbaumer, A., Hillemann, E.-C., Steiner, C.M., & Albert, D. (2012). An Evaluation System for Digital Libraries. In Zaphiris, P. Buchanan, G., Rasmussen, E., and Loizides, F. (Eds.), *Theory and Practice of Digital Libraries*. Second International Conference, TPDL 2012. Lecture Notes in Computer Science vol. 7489, pp. 414-419. Berlin: Springer.

Patton, M (1990). Qualitative Evaluation and Research Methods. Thousand Oaks, CA: *Sage*

Steiner, C.M., Albert, D. (2012) Tailor-made or Unfledged? Evaluating the Quality of Adaptive eLearning. In: Psaromiligkos, A. Spyridakos, and S. Retalis (eds.) Evaluation in E-Learning, pp. 113-143. New York: Nova Science.

Steiner, C.M., Hillemann, E.-C., Nussbaumer, A., Albert, D., Sweetnam, M.S., Hampson, C., & Conlan, O. (2013). The CULTURA Evaluation Model: An Approach Responding to the Evaluation Needs of an Innovative Research Environment. In: S. Lawless, M. Agosti, P. Clough, & O. Conlan (Eds.), Proceedings of the First Workshop on Exploration, Navigation and Retrieval of Information in Cultural Heritage. SIGIR 2013, pp. 39-42. Dublin.