

A Database of Empirical Evaluations of Adaptive Systems

Stephan Weibelzahl

Gerhard Weber

Pedagogical University Freiburg

Kunzenweg 21

79117 Freiburg, Germany

[weibelza,weber]@ph-freiburg.de

The development of an online database for studies of empirical evaluations of adaptive systems (EASy-D) is proposed. Such a database will serve as reference for researchers in the field of adaptive systems and as guide for planning new evaluations. It aims at encouraging more evaluations that fulfill certain methodological requirements. The structure of the database records is discussed and the functionality of the web interface is described.

1 Current Evaluations of Adaptive Systems

Empirical evaluations of adaptive systems are rare, e.g., only about one out of four articles published in *User Modeling and User Adapted Interaction* (UMUAI) reports significant empirical evaluations (Chin, 2001). Moreover, it is still controversial which experimental designs and criteria are useful.

Researchers who want to evaluate their system empirically have to get an overview about current evaluations to plan their own study. In other fields, there are compendia and guidelines that provide such an overview. However, composing some simple guidelines for evaluation planning is not adequate for adaptive system evaluation at the moment if we consider the thin base of results that is currently available. Thus, a more promising approach is to improve the access to relevant studies by establishing an online database of empirical evaluations of adaptive systems. This paper describes the aims of such a database, what the record structure looks like, and how the system will be maintained and successively extended.

2 Developing a Database of Empirical Evaluations

We developed a searchable online database for studies on empirical evaluations of adaptive systems which is available at <http://art7.ph-freiburg.de/easy-d>.

2.1 Aims

The main goal of this database is to encourage empirical evaluations of adaptive systems. By providing a searchable categorized set of studies, interested people get suggestions of experimental designs, criteria, and experimental issues.

Name: the system's full name

Function of system: rough category to describe the system. These categories include

- tailor information presentation
- recommend products or other objects
- help user to find information
- support learning
- give help
- adapt interface
- take over routine tasks
- support collaboration
- other function

Task: description of the task user and system perform together. If possible the domain is reported as well.

Adaptation rate: micro or macro adaptation (Cronbach, 1967); basically a question of whether a long-term or a short-term user model is used

Purpose of adaptation: why is adaptation used; which goals or advantages have been the reason for using adaptation

Method of adaptation: in which way does adaptation take place (e.g., selection of appropriate level of difficulty of tasks, change layout of display)

Figure 1: Description of the system categorization

Such a database will help to identify pitfalls in the planning process as well as in the analysis of collected data. Moreover, it will identify lacks in the state of art, e.g., a certain category of systems might appear to be not evaluated at all.

For people outside the community the database will serve as reference for the usefulness (or insufficiency) of adaptive systems in general, of certain types of systems, or of a specific system as it describes the current state of the art.

2.2 Record Structure

From the user's point of view, there are two kind of entries: experimental studies and systems. While each study evaluates one or more systems, a system which is categorized by a specific function and an adaptation mechanism might be evaluated in several studies.

2.2.1 Systems

Each evaluated system is described in terms of its name, the function it fulfills, the task that it performs, and a brief description of the adaptation mechanism. This way of characterizing an adaptive system and most of the categories are adopted from Jameson (1999). See Figure 1 for a detailed description of the categories.

The purpose and the method of adaptation are important to help the reader in understanding what the system does. For finding related systems the functioning and the task are probably more important.

References: references where the information about the system was drawn from

Evaluated system: name of the evaluated system(s)

Evaluation layer: according to the framework proposed by Weibelzahl (2001) a study can be assigned to one or more of the following evaluation layers

- evaluation of input data
- evaluation of inference
- evaluation of adaptation decision
- evaluation of interaction

Method of evaluation: a short description of the evaluation, using one of the following categories

- without running system
 - results of previous research
 - early exploratory studies
 - knowledge acquisition from experts
- studies with a system
 - controlled evaluations with users
 - controlled evaluations with hypothetical (i.e., simulated) users
 - experience with real world use

Data type: brief description of the kind of analyzed data (e.g., observed behavior, questionnaire, interview, log-files, etc.)

Criteria: which were the main criteria, and which measures were used, if possible measures (e.g., elapsed time) are grouped in reference to the abstract criterion (e.g., user satisfaction)

Criteria categories: one or more of the following categories apply if at least one of the criteria belong to it

- efficiency
- effectiveness
- usability

Figure 2: Description of the study categorization

2.2.2 Studies

For each study the database provides a citation, a reference to the evaluated system, and a detailed description of the evaluation design (see Figure 2). For experimental studies statistical data and methods of analysis are reported as far as available (see Figure 3). The categories of *evaluation method* were also adopted from Jameson (1999).

The studies are categorized in reference to an evaluation framework introduced by Weibelzahl (2001). This framework proposes to evaluate four different information processing steps in a so-called layered evaluation (Karagiannidis & Sampson, 2000). Recently, two alternative frameworks have been proposed (Brusilovsky, Karagiannidis, & Sampson, 2001; Paramythis, Totter, & Stephanidis, 2001) which could serve as additional categorization.

2.3 Online Interface

Users may search for either a system or a study. For finding a related system it is most important to search for a specific function. Searching for a specific name, and full text search are supported as well.

N: number of subjects, sample size
k: number of groups or conditions
randomization: is the assignment of subjects to groups randomized or quasi-experimental
statistical analysis: which statistical methods are used (e.g., ANOVA, MANOVA, correlation)

Figure 3: Additional information for experimental studies

The presentation of results includes the complete information that is available about this system, as well as a link to studies that evaluated this system.

When searching for a related study the user might either fill in a method of evaluation, specify the evaluation layer, or limit the search to a certain data type or criterion. In principle other search criteria (e.g., sample size) would be easy to implement, but appear to be not very useful.

In addition, there is a glossary that explains the categories and entries, and a form for authors to submit a new study. The submission procedure is explained in the next section.

3 Implementation and Maintenance

EASy-D is based on MySQL and PHP. Currently the database contains 24 studies most of them from *UMUAI* and *User Modeling Conferences*. Of course this small number of records would not require a complete database and should be seen as a starting point only. However, we hope that other authors are interested in making studies (either their own studies or papers that are of importance) available in EASy-D.

New records are submitted with an online form by categorizing and describing a study and—as long as it is not available in the database—the evaluated system. However, submissions are reviewed before being published to avoid abuse and to keep entries consistent in terms of language and format.

Certainly, submission requires a published study. However, we strongly encourage authors to submit both successful and unsuccessful studies, because especially from the later ones others might learn a lot for planning their own study. Peer reviewed studies are preferred but not required as long as the publication is available for the community.

4 Summary

Compared to a usual literature search users of EASy-D may search for system functions and evaluation methods very easy. Moreover, the studies are presented in a standardized way which gives a quick overview of the study. Another advantage is that related studies that evaluate the same system or a system with the same function are identify quickly.

We hope that this database will become the central contact point for researchers who are planning empirical evaluations of their adaptive systems and invite everybody to enhance EASy-D by submitting studies or giving feedback.

References

Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2001). The benefits of layered evaluation of adaptive applications and services. In S. Weibelzahl, D. N. Chin, & G. Weber (Eds.), *Empir-*

ical Evaluation of Adaptive Systems. Proceedings of workshop at the Eighth International Conference on User Modeling, UM2001 (pp. 1–8). Freiburg.

- Chin, D. N. (2001). Empirical evaluation of user models and user-adapted systems. *User Modeling and User-Adapted Interaction*, *11*(1-2), 181–194.
- Cronbach, L. (1967). How can instruction be adapted to individual differences. In R. Gagné (Ed.), *Learning and individual differences*. Ohio: Columbus.
- Jameson, A. (1999). *User-adaptive systems: An integrative overview*. (Tutorial presented at the Seventh International Conference on User Modeling, Banff, Canada, June 20th 1999)
- Karagiannidis, C., & Sampson, D. G. (2000). Layered evaluation of adaptive applications and services. In P. Brusilovsky & C. S. Oliviero Stock (Eds.), *Proceedings of International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH2000, Trento, Italy* (p. 343-346). Berlin: Springer.
- Paramythis, A., Totter, A., & Stephanidis, C. (2001). A modular approach to the evaluation of adaptive user interfaces. In S. Weibelzahl, D. N. Chin, & G. Weber (Eds.), *Empirical Evaluation of Adaptive Systems. Proceedings of workshop at the Eighth International Conference on User Modeling, UM2001* (pp. 9–24). Freiburg.
- Weibelzahl, S. (2001). Evaluation of adaptive systems. In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva (Eds.), *User Modeling: Proceedings of the Eighth International Conference, UM2001* (pp. 292–294). Berlin: Springer.