UCEAS: User-centred Evaluations of Adaptive Systems

Catherine Mulwa, Séamus Lawless, Mary Sharp, Vincent Wade

Knowledge and Data Engineering Group
School of Computer Science and Statistics
Trinity College, Dublin
{mulwac, seamus.lawless, mary.sharp, vincent.wade}@cs.tcd.ie

Abstract. *Adaptive Hypermedia (AH) research is concerned with the dynamic composition and personalisation of hypermedia documents in order to provide more context sensitive retrieval and reuse of digital content. The evaluation of Adaptive Hypermedia Systems (AHS) is difficult due to the complexity of such systems [1]. Several Problems and pitfalls are encountered when evaluating these systems [2-6]. Very little research has been carried out to address this problem. This PhD work proposes a user-centred approach to the evaluation of the adaptive mechanism of AHS. The proposed approach will be validated using personalised systems developed by the Centre for Next Generation Localisation (CNGL). The framework developed by this research will help to standardise current approaches, offer hints regarding the identification of failures and misconceptions of the adaptive mechanism. It will be applicable to all adaptive systems with no limitations of domain or inference mechanism. A review of approaches, methodologies and techniques adopted by existing systems was conducted and the results analysed. An architectural design of the framework has been designed and currently implementation work is going on.*

Keywords: Information Retrieval, Personalisation, Adaptive Hypermedia, Evaluation

1   Introduction

The research field of Adaptive Hypermedia (AH) has been growing rapidly during the past 15 years and this has resulted in terms, models, methodologies, and a plethora of new Adaptive Hypermedia systems (AHS). Recently, research has been undertaken exploring how to enhance and combine key aspects of AH research with information retrieval (IR) techniques to provide advanced annotation, slicing, retrieval and composition of multilingual digital content drawn from corporate documents repositories as well as open corpus sources [7-8]. The evaluation of these systems is important. It is essential to ensure that the evaluation uses the correct methods and techniques [9-10]. Existing approaches and methods such as the layered approach, empirical approach, utility approach and heuristic research still encounter inherent problems [21]. This work concentrates on introducing new ideas into the evaluation of the adaptive mechanism in AHS, particularly AHS which incorporate IR techniques for the personalised retrieval of content.

This research will tackle the question of:
*"What are the affordances of user-centred evaluation techniques for end user evaluation of adaptive systems, in particular adaptive systems which combine adaptive hypermedia and information retrieval techniques?"*The research introduced in this paper addresses the following challenges: i) to investigate, analyze and identify the affordances of user-centred evaluations (UCE) techniques for end-user evaluation of adaptive systems specifically adaptive

systems which combine adaptive hypermedia and information retrieval techniques.-, ii) to design an architectural model for UCE of AHS using a hybrid approach of UCE and the layered approach from the studies analysed as part of challenge 1, and iii) to design a generic and reusable framework applicable to all adaptive systems with no limitations of domain or inference mechanism. It will help to standardise current approaches and offer hints regarding the identification of failures and misconceptions of the adaptive mechanism and also how to evaluate the mechanism.

One major contribution of this research are: the introduction of a hybrid evaluation methodology for interactive adaptive systems which combine IR, AH and adaptive web techniques and technologies [10]. The complex functionality of such systems coupled with the variety of potential users makes the evaluation tricky, expensive and time consuming. This evaluation methodology requires both component-level scientific evaluation and user-based evaluation. A minor contribution the provision of an interactive and collaborative user interface; the collaborative nature of the architecture enables the sharing of information among similar users,

Following is a brief overview of the *research classification* and *characteristics*: (i) *Research Paradigm:* Engineering: (Observe existing solutions –>propose better solutions –> build or develop –> measure and analyse – Repeat), ii) *Research topic:* Systems software: Software lifecycle/engineering & methods/techniques, system/software: measurement/metrics, (iii) *Research Approach:* Quantitative (Review of literature, Evaluative: evaluative critical, design based user-centered). (iv) *Research method:* Literature Review/ analysis, Data analysis, user trials/experiments (v) *Reference Discipline:* Computer Science & science (vi) *Analysis Level:* Technical.

The rest of the paper, which reflects mid-stages of this Ph.D. research, presents a brief overview of the state-of-art of UCE of adaptive systems and a brief description of the proposed framework, architectural design and proposed evaluation methodology for the framework.

2   State of the Art

In the following section, a number of areas will be discussed: the problems and pitfalls faced by evaluators; different evaluation approaches; methodologies and techniques; variables; and metrics; adopted by existing systems. Fifty six publications were selected as a representative set of UCE evaluation studies.

The evaluation of AHS is a difficult task due to the complexity of such systems, as shown by many studies[1]. It is of crucial importance that the adaptive features of the system can be easily distinguished from the general usability of the designed tool. Issues arise in the selection of applicable criteria for the evaluation of adaptivity. Many metrics can be used to measure performance, for example: knowledge gain (AEHS), amount of requested materials, duration of interaction, number of navigation steps, task success, usability (e.g., effectiveness, efficiency and user satisfaction). The evaluation of adaptive systems is not easy and several researchers have pointed out potential pitfalls when evaluating adaptive systems [2-6]: e.g. generalisation of problem; allocation of resources; specification of control conditions; sampling; definition of criteria; asking for adaptivity effects; reporting results; difficulty in attributing cause; difficult in finding significant results due to variance; difficulty in defining the effectiveness of adaptation; difficulty in finding resources due to allocation of insufficient resources or not

enough resources left; too much emphasis on summative rather than formative evaluation; and evaluation results are reported incomplete or anecdotally

To tackle the above inherent usability problems, several researchers have applied the following *approaches* in evaluating the adaptive mechanism of AHS: (i) The *empirical approach:* Empirical evaluations, also known as controlled experiments, refer to the appraisal of a theory by observation in experiments. These evaluations help to estimate the effectiveness, efficiency and usability of a system and may uncover certain types of errors in the system that would remain otherwise undiscovered [11]; (ii) The l*ayered approach:* This approach [13][14] separates the 'interaction assessment' and the 'adaptation decision'. Evaluating AHS on a layer by layer basis has been recommended as a more comprehensive approach [14][15]; (iii) The *utility-based approach* [17] offers a perspective on how to reintegrate the different layers; and iv) The *heuristic approach:* The use of heuristics ensures that the entire system can be evaluated in-depth and specific problems can be discovered at an early design stage before releasing a running prototype of a system [19]. This approach can help evaluators by improving the detection and diagnosis of potential usability problems, v) *User-centered approach*: This approach can serve three goals; verifying the quality of an adaptive system, detecting problems in the system functionality or interface, and supporting adaptivity decisions. Potential benefits are savings in terms of time and cost, ensuring the completeness of system functionality, minimizing required repair efforts, and improving user satisfaction.

The following adaptive variables, methods of UCE, metrics for evaluating adaptivity and evaluation criteria were selected from the UCE approach studies, to be used in validating the research question, objectives and the developed framework for UCE of adaptive systems:

A total of 21 adaptive variables that can prompt adaptivity were identified, these variables make UCEAS systems a variable tool for developers in technology-enhanced learning environments (TELE) [12-13] (i.e., appreciation, background and hyperspace, environment, individual traits, intention to use, groups of users, knowledge of domain, personal data, perceived usefulness, preferences, trust and privacy, usability, usage data, user skills and capabilities, user cognitive workload, user experience, user goals, user interests and user behaviour). *Methods for UCE* [11-12, 14]*: i.e., (* Interviews, questionnaires, focus group, discussion groups, user observation, the systematic observation, verbal protocol and think aloud protocols, expert review, parallel design, cognitive walkthroughs, Wizard of 0z simulation, scenario-based design, usability testing, contextual design, cultural probes, creative brainstorming sessions, task analysis, qualitative (Ethnograph), quantitative (grounded theory)). *Metrics for evaluating adaptivity* [15-16]*:* i.e., Architectural metrics structural metrics, Interaction metrics, Personalisation metrics and documentation metrics. *Evaluation Criteria:* (e.g., aesthetic, consistency, self-evidence, naturalness of metaphors, predictability, richness, completeness, motivation, hypertext structure, autonomy, competence and flexibility).

The next section introduces our current work; framework for UCE of adaptive systems, the architectural design, implementation and proposed evaluation methodology
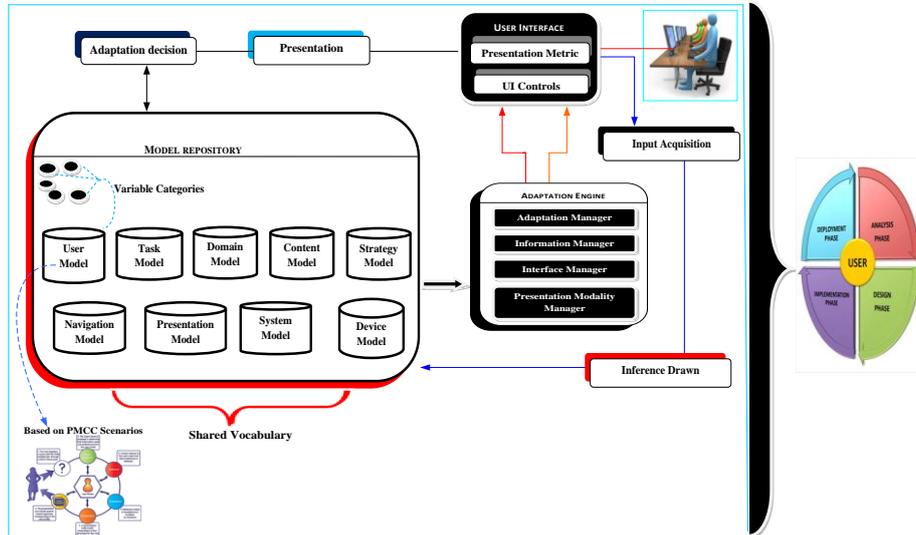
Figure 1: Architectural design for Frameworks of UCEAS: Module 1

3   Framework for UCE of Adaptive Systems

A framework for UCE evaluation of adaptive systems was specified designed and currently is being implemented. The framework is generic and reusable and it will be applicable to all adaptive systems with no limitations of domain or inference mechanism. It is divided into four modules; i) Module 1: Review of how different models of adaptive systems have been evaluated, ii) Module 2: Review of UCE of adaptive systems, iii) Module 3: Quantifying for adaptivity adaptability and iv) Module 4: Recommendation on how to evaluate adaptive systems. Following is a brief overview of the architectural design of the four modules:

3.2 Architectural design for the Framework

Module 1: Adaptive systems are composed of different models i.e., user, domain, task, content, strategy, navigation, representation etc This module presents information on how existing models for adaptive and adaptive hypermedia systems have been evaluated. Currently no research is going on in this area, thus the significance of this module. Figure 1 presents an architectural design that uses a hybrid approach of UCE and Layered.

Module 2: This module has been developed, validated and evaluated [22]. Figure 2 presents the architectural design. It consists of 3 layers: presentation, business logic and data persistence. The architecture includes RSS Feed Management, Paper Subscription, SMART URL analysis and Document Downloading. The educational benefit of this module is in the provision of an interactive reference tool to encourage the evaluation of adaptive systems. The studies collected can be used as a basis of a searchable online database that provides an overview of the state-of-the-art to a scientific community and encourages scientists to evaluate their own systems. The collaborative nature of the module facilitates the sharing of information among research students.
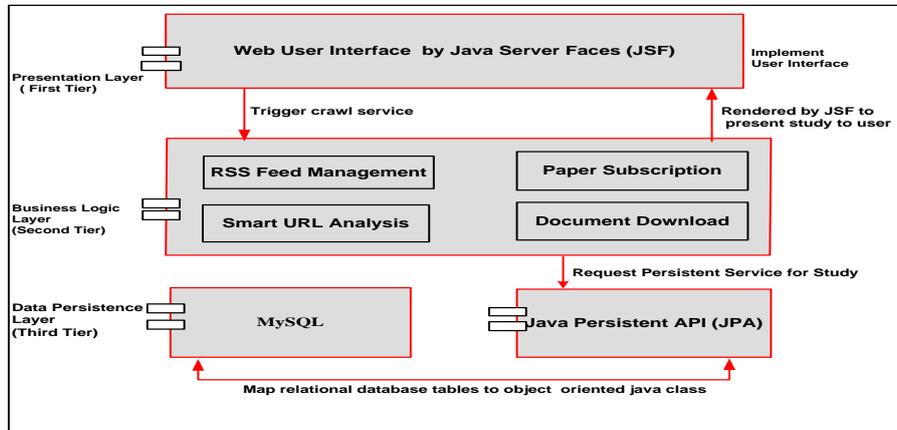
Figure 2: Architectural design for Frameworks' Module 2

ii) *Module 3:Quantifying for Adaptivity and adaptability:* This module provides examples of systems that adapt automatically and the ones that allow the users to set adaptation. It also presents a matrix used in testing for adaptivity. Seven dimensions are identified refer to Figure 3 below:
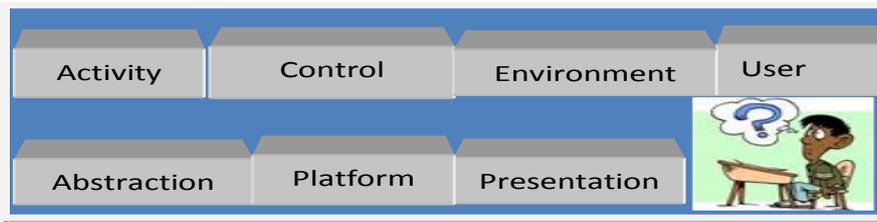
Figure 3: Module 3 User Interface

iii) Module 4: Recommendations for Evaluation of Adaptive Systems. This section provides suggestions to developers, evaluators and to researchers on how to plan and evaluate their AHS. e.g., what evaluation methods, techniques, metrics, criteria and data types to use. This involves filling a form and submitting it into the frameworks database. Figure 4 below presents the user interface for this module:

Figure 4: User Interface for Module 4

3.2 Framework Implementation and Evaluation Methodology

Currently implementation work is going on for module 1, 3 and 4. The proposed evaluation methodology consists of a hybrid of UCE and Layered approaches. The framework will be evaluated based on: i) query formation; (ii) evaluating retrieval effectiveness; (iii) for user-centric evaluation i.e. evaluating adaptivity effectiveness, efficiency and user satisfaction and iv) measuring framework efficiency.

3.3 Scenario: How the Framework Works

Let's assume "I am a researcher who has developed an adaptive system that I want to evaluate. I come to this Framework looking for information on: i) How other adaptive systems have been previously evaluated, what evaluation methods and measurement criteria have been used, ii) How to test for adaptivity and adaptability, iii) recommendation on how to evaluate my system, and iv) how different models e.g., user model, domain model, content model etc have been evaluated". The framework provides an interactive technology enhanced user interface, where researchers can interact and are provided with the above information. This framework will help standardize current evaluation approaches and offer hints regarding the identification of failures and misconceptions of the adaptive mechanism and it will serve as a reference tool for researchers in the different fields of any kind of adaptive system.

4   Discussion and Conclusion

In order to produce effective results, evaluation should occur throughout the entire design cycle and provide feedback for design modification [6]. The proposed solution is novel compared to existing approaches. User-centered evaluation (UCE) can serve three goals: verifying the quality of an AHS, detecting problems in the system functionality or interface, and supporting adaptivity decisions. Earlier evaluation studies compared adaptive versions of the system with the non-adaptive versions [24]. A major criticism of this was that the non-adaptive versions usually, when implemented using adaptive systems with the adaptivity switched off, were not "optimal". It is the contention of this paper that contextual, and specifically personalised, approaches to IR could benefit from the evaluation experience of the Adaptive Hypermedia community.

References

1.   Gena, C. *Evaluation methodologies and user involvement in user modeling and adaptive systems*. in *User Modeling and User-Adapted Interaction*. 2003. Torino, Italy: Proceedings Of Simposium on Human-Computer Interaction.
2.   Weibelzahl, S. *Problems and pitfalls in evaluating adaptive systems*. 2005: Citeseer.
3.   Tintarev, N. and J. Masthoff, *Evaluating Recommender Explanations: Problems Experienced and Lessons Learned for the Evaluation of Adaptive Systems*, in *User Modeling, Adaptation and Personalization*. 2009: Trento, Italy. p. 10.
4.   Masthoff, J. *The evaluation of adaptive systems*. 2003: IGI Publishing.
5.   Raibulet, C. and L. Masciadri. *Evaluation of Dynamic Adaptivity Through Metrics: an Achievable Target?*
6.   Gena, C. and S. Weibelzahl, *Usability engineering for the adaptive web.* The Adaptive Web, 2007: p. 720-762.
7.   Jones, G.J.F. and V. Wade, *Integrated Content Presentation for Multilingual and Multimedia Information Access.* 2006. **40**: p. 31-39.
8.   Steichen, B., et al., *Dynamic Hypertext Generation for Reusing Open Corpus Content*, in *Hypertext 2009*, I.t.P.o.t.t.A.C.o.H.a. Hypermedia, Editor. 2009: Torino, Italy.
9.   Brusilovsky, P., Karagiannidis, P., and Sampson, C., *Layered Evaluations of Adaptive Learning Systems.* 2004. **14**: p. 402-421.
10.   Mulwa, C., et al., *A Proposal for the Evaluation of Simulated Interactive Information Retrieval in Customer Support*, in *SIGIR Workshop on the Automated Evaluation of Interactive Information Retrieval*. 2010, ACM: Geneva, Switzerland. p. 2.
11.   Gena, C., *Methods and techniques for the evaluation of user-adaptive systems, The Knowledge Engineering Review.* 2005. **20**(1): p. 1-37.

12. Van Velsen, L., et al., *User-centered evaluation of adaptive and adaptable systems: a literature review.* The Knowledge Engineering Review, 2008. **23**(03): p. 261-281.
13. Mulwa, C., et al., *Adaptive Educational Hypermedia Systems in Technology Enhanced Learning: A Literature Review*, in *Association for Computing Machinery's Special Interest Group for Information Technology Education*, Sheridan, Editor. 2010: H Hotel 111 W Main St. Midland,MI 48640. p. 10.
14. Díaz, A., A. García, and P. Gervás, *User-centred versus system-centred evaluation of a personalization system.* Information Processing & Management, 2008. **44**(3): p. 1293-1307.
15. Masciadri, L. and C. Raibulet. *Frameworks for the Development of Adaptive Systems: Evaluation of Their Adaptability Feature Through Software Metrics*. 2009: IEEE.
16. Mulwa, C., et al. *A Proposal for the Evaluation of Adaptive Information Retrieval Systems using Simulated Interaction*. in *Workshop on the Simulation of Interaction in Automated Evaluation of Interactive Information Retrieval, SimInt 2010, at the 33rd Annual ACM SIGIR Conference*. 2010. Geneva, Switzerland.
17. Mulwa C., et al., *Adaptive Educational Hypermedia Systems in Technology Enhanced Learning: A Literature Review*, in *Association for Computing Machinery's Special Interest Group for Information Technology Education*, Sheridan, Editor. 2010, ACM Proceedings: Central Michigan University at Midland MI USA p. 10.
18. Jameson, A., *Adaptive Interfaces and Agents. The Human-Computer Interaction Hand-book*. 2003, New Jersey: Lawrence Erlbaum Associates. 316-318.
19. Gena, C., *A USER-CENTERED APPROACH TO THE RETRIEVAL OF INFORMATION IN AN ADAPTIVE WEB SITE. *, in *Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer*, E. Alkhalifa, Editor. 2006.
20. Dix, A., et al., *Human-computer interaction*. Second Edition, ed. P. Hall. 1998: Prentice hall.
21. Raibulet, C. and L. Masciadri. *Evaluation of dynamic adaptivity through metrics: an achievable target?* in *Software Architecture, 2009 & European Conference on Software Architecture. WICSA/ECSA 2009. Joint Working IEEE/IFIP Conference on*. 2009.
22. Mulwa, C., et al., *OSSES: An Online System for Studies on Evaluation of Systems*, in *ED-MEDIA World Conference on Educational Multimedia, Hypermedia & Telecommunication*. 2010, Educational & Information Technology Digital Library: Toronto,Canada. . p. 10.
23. Knutov, E., P. De Bra, and M. Pechenizkiy, *AH 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques.* New Review of Hypermedia and Multimedia, 2009. **15**(1): p. 5-38.
24. Gupta, A. and P. Grover. *Proposed evaluation framework for adaptive hypermedia systems*. 2004.