

A Framework for Supporting the Choice of Usability Evaluation Methods for Interactive Adaptive Systems

Amira Dhouib, Mahmoud Neji
Miracle Laboratory
Faculty of Economics and
Management Sciences
University of Sfax, B.P. 1088,
Sfax 3000 Tunisia
E-mail : amiradhoub@live.fr,
mahmoud.neji@fsegs.rnu.tn

Abdelwaheb Trabelsi
LOGIQ
Faculty of Sciences Sfax
University of Sfax, B.P. 1088,
Sfax 3000 Tunisia
E-mail: abdelwaheb@gmail.com

Christophe Kolski
LAMIH-UMR CNRS 8201,
University of Valenciennes and
Hainaut-Cambr sis, Valenciennes,
France
E-mail: Christophe.Kolski@univ-
valenciennes.fr

Abstract—The evaluation of interactive adaptive systems is considered as a difficult endeavour due to the complex nature of these systems. Issues arise in the need to identify, for a given evaluation situation the set of suitable usability evaluation methods to be used by the evaluators. This paper describes a framework for supporting the choice of appropriate methods for the usability evaluation of interactive adaptive systems. It is based on the comparison of usability evaluation methods using relevant criteria. This framework is then validated through a case study; it concerns an adaptive hypermedia system in transportation field.

Keywords—Evaluation; Interactive adaptive system; Evaluation method; Adaptive hypermedia system; Usability

I. INTRODUCTION

Evaluation has a crucial importance in the area of interactive adaptive systems (IAS) [1]. Moreover, it is necessary that correct evaluation methods are used since an incorrect evaluation method can lead to incomplete or wrong conclusions [2,3]. However, many difficulties can exist in the evaluation of such systems [4,5,6]. The difficulties can be caused by the need to distinguish between the adaptive features and the general usability of the IAS [1].

A wide variety of evaluation techniques and methods dedicated to IAS are described in the literature [7,8,9]. The diversity of these methods involves the difficulty of choosing the most appropriate ones usable for evaluate the usability of a given IAS in specific situation. The choice problem concerning evaluation methods is a complex and challenging issue and it depends on different features and criteria [10]. Considerable effort is required to understand the applicability of each method in a particular situation [11]. Such shortcomings establish motivations for the proposition of a framework supporting the choice of the suitable usability evaluation methods (UEM) for interactive adaptive systems.

The outline of this paper is organized as follows. The second section presents a state of art about the evaluation

methods and frameworks for interactive adaptive systems. This section covers an overview of the existing frameworks and their limitations. Thus, it provides the motivation for the presented work. The third section introduces the framework for supporting the choice of suitable methods that evaluate the usability of IAS. The latter is validated in the fourth section which includes a case study of an adaptive hypermedia system in transportation field. The paper is concluded by a summary and prospects.

II. THE EVALUATION OF INTERACTIVE ADAPTIVE SYSTEMS

Interactive adaptive systems refer to “interactive software systems that automatically adapt to properties and behaviors of individual users.” [12]. Evaluation of these systems is important. It is important to not only evaluate but also to ensure that the evaluation uses the correct method(s). In fact, evaluation of IAS is based on various approaches, techniques and methods [1,7].

In the two followings sub-sections, we present the most important methods and frameworks used to evaluate interactive adaptive systems.

A. Overview of Evaluation Methods for IAS

The evaluation of IAS is a fundamental stage in their development and it is based on various approaches, metrics and methods. In this section, we present an overview of methods that have been proposed to evaluate IAS.

Comparative evaluation: In which an adaptive instance of system is compared with a non-adaptive one [13]. However, it is not easy to make such comparison. It highly depends on how the non-adaptive version was obtained. According to [14], comparative evaluation do not evaluate an exact aspect of adaptation, hence it becomes very hard to point out the causes of the “success” or “failure” of the adaptation. Specially, it is very hard to identify why one aspect of adaptation can be applied to reach the goal.

- Empirical evaluation: The use of the empirical approach helps evaluators to estimate the effectiveness, the efficiency, and the usability of a system that applies artificial intelligent techniques in real-world scenarios [15]. Empirical evaluation (or controlled experiment) refers to the appraisal of a theory by observation in experiments. This approach can help to discover failures in interactive systems that would remain uncovered otherwise.
- Layered evaluation: This evaluation method does not treat evaluation as a “monolithic” process but assesses the success of adaptation by decomposing it into different layers, and evaluating them separately [14,16]. The different layers reflect the various stages/aspects of the adaptation. The layered evaluation helps evaluators in identifying the exact cause of the adaptation failure or any other error. It provides a series of advantages over previous attempts, such as facilitation of improvements, generalization of evaluation results, and re-use of successful practices.
- Utility-based evaluation: The evaluation can be seen as a utility function that maps selected, measurable criteria with respect to the performance of the adaptive system to a quantitative representation of user satisfaction [17].
- Heuristic evaluation: This method can be used to evaluate an entire system in-depth and to help evaluators by improving the identification of potential usability problems [18]. The most popular heuristics in usability testing are Nielsen’s heuristics [19] consisting of ten broad guidelines based on a factor analysis of common usability problems. In [20], Magoulas and his colleagues proposed an integration of Nielsen’s heuristics with layered evaluation for adaptive learning environments. For each level of adaptation, they modified a subset of the usability heuristics and added more detailed criteria for these heuristics.
- User-centered evaluation: For interactive adaptive systems especially user-centered evaluation methods are recommended [9] because users are both the main source of information and the main target of the application. These evaluation methods help evaluators to detect the real problems encountered by end users at the time of the execution of their task with the system [21]. Examples of user-centered evaluation methods include: interviews, focus group, discussion groups, user observation, verbal protocol, think aloud protocols, parallel design, wizard of oz simulation, usability testing, and contextual design [7,9].

B. Evaluation Frameworks for IAS

The existing evaluation frameworks are wished to find system deficits and failures, show whether adaptivity in

adaptive system is successful, and justify the efforts spent on making systems adaptive. However, important as it is, the research area is still open, there are a few evaluation frameworks reported in the adaptive systems literature. In this section, we present some of these evaluation frameworks.

The evaluation framework, suggested by Gupta and Grover [22], treats the evaluation as an integral part of the development process of adaptive hypermedia systems and takes the environment in which adaptive system is accessed, and the type of adaptation used into consideration while evaluating individual modules of adaptive hypermedia systems. The framework is an extension of the layered evaluation; it consists of four orthogonal dimensions: environment, adaptation, development process, evaluation modules. These dimensions are orthogonal to each other i.e. all the evaluation modules should address all the components of environment and adaptation during each phase of development process.

Another framework, called AnAmeter, was presented by Tarpin-Bernard *et al.* [23]; it provides the first steps towards the evaluation of the quality of a system’s adaptation. Evaluators can characterize the different types of adaptations a system features and use the resulting table to determine exactly what needs to be assessed. AnAmeter provides also an overall score for the adaptation degree of the system yet also provides sub-scores (local and semi-global). It quantifies the “degree” of adaptation in a system, and in that it is supported by a web-based tool that enables evaluators to interactively manage the tabular description of the system at hand.

Mulwa and Wade [24] propose a Web-based framework, called EFEx, which recommends appropriate methods for the evaluation of adaptive e-learning systems. A combination of different factors was considered in the recommendation process of the above mentioned framework: e.g. the number of publications in which the evaluation methods was used; the category of publications (journal, conference or workshop) in which the evaluation methods have been used; and the number of adaptive e-learning systems belonging to the same category that have been evaluated using the considered evaluation methods. The framework also allows the search of existing evaluation techniques and studies for adaptive systems using data extracted from 340 peer reviewed papers.

Many shortcomings in the previously presented framework can be noted. They are primarily related to the problem of determining which publications must be gathered and examined to find out in which type the evaluation method was used. Another limitation lies in how to filter the collected publications. Another challenge concerns the limited application domain of this framework. In fact, the latter supports only the evaluation of adaptive e-learning systems. To address the challenges outlined above, we have specified and developed a framework supporting the choice of appropriate usability evaluation method(s) for interactive adaptive systems.

The next section introduces the proposed framework, its general architecture and its contributions.

III. THE PROPOSED FRAMEWORK

A major problem which novice and even expert evaluators are confronted with is the choice of suitable UEM for adaptive systems in a specific situation. The aim of the proposed framework is then to propose and to identify the method(s) to use in priority to evaluate the usability of IAS. In other words, it helps evaluators to decide which usability evaluation methods are the most appropriate, and to have explanations as to why each method was identified.

The end user of the proposed framework can be an unskilled evaluator (IAS software developer, student in human-computer interaction (HCI) or an expert evaluator) and it is possible to apply the proposed framework on the various types of IAS. In [24], fourteen different types of IAS are identified, such as: adaptive hypermedia system, adaptive recommender system, adaptive e-learning system, adaptive educational hypermedia system, and adaptive information retrieval system, etc.

The next section discusses the general architecture of the proposed framework.

A. General Architecture

The framework is proposed to assist the evaluator in the identification of applicable evaluation method(s) for IAS in a given evaluation situation. Fig.1 illustrates the general architecture of the proposed framework, which revolves around three major phases including: (1) configuration phase, (2) usability evaluation methods selection, and (3) results generation.

Each phase will be presented in detail in the next sub-sections.

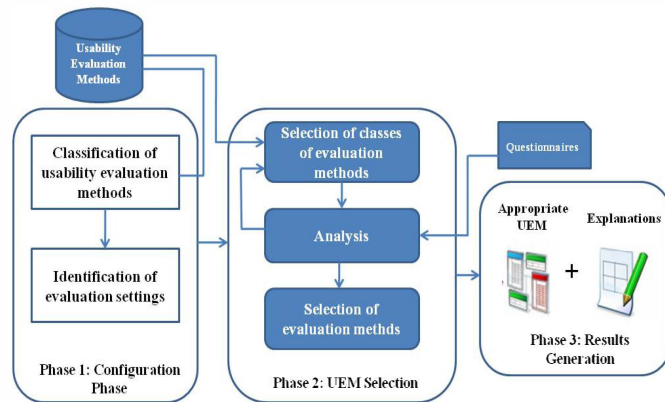


Figure 1. General architecture of the proposed framework

1) *Phase 1: Configuration phase*: This module captures the various input data that help to choose the appropriate usability evaluation method(s): the classified evaluation methods and the evaluation settings.

a) *Classification of evaluation methods for IAS*: First, an expert evaluator selects the methods that can be applied for the evaluation of the usability of IAS. Then he/she classifies them according to a set of features (dimensions).

In this paper, we distinguish essentially five main features:

- The evaluated quality factor: Which highlight the factor the methods are most suited to evaluate (i.e. usability).
- The category of IAS: Type on which the evaluation method(s) may be applied (e.g. adaptive hypermedia system, adaptive recommender system, adaptive e-learning system, and adaptive information retrieval system).
- The adaptive system development phases: Stage of the development life cycle that each method is best suited for. According to [7], we distinguish three phases: requirement phase, preliminary evaluation phase and final phase.
- The stakeholders: Participant(s) who is/are involved in the evaluation process (e.g. users, expert evaluators, designers).
- The adaptation layers of adaptive system: The layer which needs to be considered for the evaluation of IAS. According to [16], we distinguish two layers: interaction assessment layer (i.e. test the validity of the conclusions drawn by the system concerning the characteristics of the user-computer interaction), and adaptation decision layer (i.e. evaluate the validity and meaningfulness of the rules applied for the adaptation).

b) *Identification of evaluation settings*: In this level, the end user of the proposed framework (expert or novice evaluator) selects features which concern the conditions for the evaluation and the performance of evaluators; some of them include:

- User performance: Age, gender, and experience with system.
- Availability of temporal, human, and financial resources.
- Category of the adaptive system: The type of the IAS to evaluate.
- Expertise of evaluators: Evaluator can be novice or expert.
- Location: Geographical location of the accessing device that can be used to adapt the content of IAS.
- Development phase: Stage in the development process of IAS.
- Device used: PC, tablet, Smartphone, etc.

2) *Phase 2: Usability evaluation methods selection*: UEM selection is a three-step phase based on the information gathered in phase 1:

a) *Selection of classes of evaluation methods*: The proposed framework retrieves all the evaluation methods related to the evaluation settings selected in phase 1. All the retrieved evaluation methods are grouped into classes of

methods. Each class of evaluation methods is classified according to the evaluation settings selected in phase 1 (e.g. class 1 includes evaluation methods that are influenced by the available financial resources. This means that it may include for instance all the evaluation methods requiring a limited budget).

b) Analysis phase: In this level, the end user has to attribute his/her level of satisfaction with the proposed class of methods through a questionnaire. In fact, he/she has to use the class of evaluation methods and allocates to each one a score according to his/her level of satisfaction.

In the case of no satisfaction, the framework proposes a list of actions (e.g. increase the available budget, reduce the number of evaluators, etc.), which enable the end user to obtain new classes of evaluation methods and so on until satisfaction. After satisfaction with the proposed classes of evaluation methods, he/she validates the provided classes of evaluation methods and moves on to the next step.

c) Selection of evaluation methods from classes: In this level, the evaluation methods are provided from the different classes identified in sub-phase 2.1. They are selected according to their ability to meet the satisfaction of the evaluator. Then, a comparison between the different selected usability evaluation methods is made. They are compared upon a combination of different criteria. In the literature, there are various criteria that can be considered for comparative evaluation methods [25].

To identify the appropriate evaluation method(s) more effectively, during this phase we have used generic criteria; they refer to the evaluation methods and not to the target system. We identify as most important the following:

- **Validity:** Is identified in [25] as the ratio of the number of real usability problems with respect to the total number of findings for each application of usability evaluation method.

$$\text{Validity} = \frac{\text{number of real problems found}}{\text{number of issues identified as problems}}$$

- **Thoroughness:** Refers to the ratio of the number of usability problems found by the application of an evaluation method with respect to the total number of usability problems that exist in the target system [26].

$$\text{Thoroughness} = \frac{\text{number of real problems found}}{\text{number of real problems that exist in system}}$$

- **Effectiveness:** Can be identified as the product of thoroughness and validity [25].

$$\text{Effectiveness} = \text{Thoroughness} * \text{Validity}$$

3) Phase 3: Results generation: In this phase, the final ranking of the evaluation methods is generated and displayed from the highest score to the lowest one that represents the suitability of usability evaluation methods for a particular issue.

The proposed framework suggests as evaluation result a list of appropriate usability evaluation methods and explanations which can help end user to better understand why these methods are proposed. The results presented to the end user are classified according to the appropriate evaluation methods.

B. Contribution of the Proposed Framework

The contribution of this paper, along these lines, is the introduction of a framework which supports the choice of suitable usability evaluation methods, from the perspective of IAS oriented evaluation.

The proposed framework is novel compared to existing ones. By using this framework, it is possible to overcome the existing limitations encountered while evaluating a given IAS. It helps evaluators to identify applicable usability evaluation methods for a given set of evaluation settings and features. Then, it helps to reduce the time spent and the cost incurred in the evaluation of IAS. Indeed, the framework can be applied to various types of interactive adaptive systems with no limitation of application domain. Furthermore, the framework proposes a flexible database, which makes it easy to extend by adding new UEM.

IV. APPLICATION TO ADAPTIVE HYPERMEDIA SYSTEM USED IN TRANSPORTATION FIELD

The proposed framework is validated by its application to an adaptive hypermedia system in the field of transportation.

A. Presentation of the Application

The designed adaptive hypermedia system is intended to assist users in their information seeking tasks by offering personalized information about the vehicles' departures times and type. This system takes into account the geographic context depending on the user's location (e.g., at home, in the stations...). Fig. 2 presents the index page of the adaptive hypermedia system which shows the different services offered to the travelers.

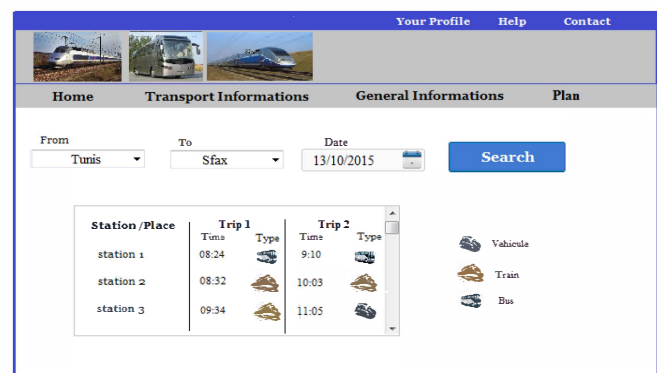


Figure 2. A screen shot of the adaptive hypermedia system

B. Results and Discussion

Before proceeding to the different steps of the proposed framework, an expert evaluator has to select and to present the features of each usability evaluation method for IAS. Fourteen classified evaluation methods, commonly used for

the investigation of usability in IAS are stored in the database. Examples of these UEM are: focus group, logging use, usability test, questionnaires, and think aloud protocol.

The end user (novice or expert evaluator) has to fulfill a questionnaire which includes some general information, e.g. full name and age, and then he/she has to select the different evaluation settings concerning the conditions for the evaluation (Fig.3). Some of these dimensions correspond to:

- Evaluation phase of IAS (in this case: the final phase).
- Type of IAS (i.e. adaptive hypermedia system).
- Available financial and temporal resources (i.e. medium budget and time).
- Human resources (i.e. expert evaluator and users).
- Level of experience with the system (i.e. high).

Once the different evaluation settings are selected, the proposed framework retrieves the different usability evaluation methods from the database. All the retrieved usability evaluation methods are grouped into classes of methods according to the evaluation settings. Then, the evaluator allocates to every class of evaluation methods a score according to his/her level of satisfaction. After that, the UEM are selected from different classes and they are compared using the criteria for the comparative UEM. As discussed in related work there are various criteria that can be considered for comparative evaluations. We make use of the criteria list presented in Section 3.1. When all the results for each method are collated, they are added up in a list. The highest level refers to the most appropriate UEM.

Finally, the framework displays the list of appropriate UEM with explanations after the validation of the generated results.

Figure 3. Questionnaires to fulfill by the evaluator in order to select the evaluation settings.

The results proposed by the framework can be different, according to the evaluation settings (the category of the adaptive system to consider, the development phase of the interactive adaptive systems...).

C. Validation of the Evaluation Results

Once the list of appropriate evaluation methods is generated, the next phase consists in evaluating the performance and usefulness of the final results generated by the proposed framework. Therefore, we propose to compare the final results recommended with the most adequate evaluation methods proposed by expert evaluators. In other words, the experts gave the list of usability evaluation methods to be used for each adaptive system with the different evaluation settings considered in the above section. Then, this list will be compared with the results generated by the framework. Twenty expert evaluators are participated in the experiment. They have +2 years of research experience in evaluating interactive system. In this study, the usability test method was the highest, followed by the focus group and the think aloud protocol for the evaluation of the usability of the given adaptive hypermedia system.

In the experiment, the aim of the expert evaluators was to complete the following question: “Which evaluation method(s) would you recommend to use when evaluating such adaptive system?” Then, we compared the methods recommended by the expert evaluators with the results produced by the proposed framework. The latter produced same evaluation methods like the ones ranked by the experts as most appropriate. Fig. 4 presents a summary of the recommended evaluation methods proposed by the expert evaluators and to be used for the evaluation of such system.

Which evaluation methods would you recommend to use when evaluating the given IAS?

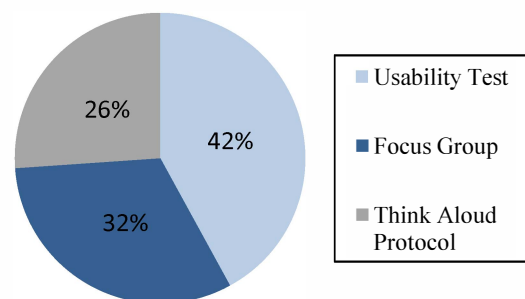


Figure 4. The recommended evaluation methods by expert evaluators

V. CONCLUSION AND FUTURE WORK

The evaluation of interactive adaptive systems is the object of numerous studies. However, the usability evaluation of such systems remains a difficult task to perform for novice evaluators, and even for experts in IAS evaluation. It is difficult to assist automatically the choice of usability evaluation methods for IAS to be used in each situation. In this paper, we introduced briefly an overview of the current methods and frameworks that contribute to the

evaluation of usability of IAS. Then, we proposed a framework supporting the choice of usability evaluation methods for IAS. The originality of this research lies in helping the evaluators to decide which methods are the most appropriate for the evaluation of the usability of interactive adaptive systems. Then, we presented an adaptive hypermedia system in the field of transportation. It allowed us to validate the use of the framework proposed to support the choice of the appropriate usability evaluation methods for interactive adaptive systems.

Since we have tested the proposed framework only within an adaptive hypermedia system in the field of transportation, it would be also interesting to generalize the framework with other fields of application and other types of IAS such as: adaptive information retrieval system and adaptive e-learning systems.

REFERENCES

- [1] C. Mulwa, S. Lawless, M. Sharp and V. Wade, "The evaluation of adaptive and user-adaptive systems: a review," *International Journal of Knowledge and Web Intelligence*, 2011, Vol. 2, No.2/3 pp. 138-156.
- [2] C. Mulwa, S. Lawless, M. Sharp, I. Arnedillo-Sanchez, and V. Wade, "Adaptive educational hypermedia systems in technology enhanced learning: a literature review," *Proceedings of the 2010 ACM Conference on Information Technology Education*, 7-9 October 2010, Midland, USA.
- [3] M. De Jong, and P. Schellens, "Reader-focused text evaluation: an overview of goals and methods," *Journal of Business and Technical Communication*, Vol. 11, pp.402-432, 1997.
- [4] F. Del Missier, and F. Ricci, "Understanding recommender systems: experimental evaluation challenges," *Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems*, held at the 9th International Conference, pp. 31-40, 2003.
- [5] J. Masthoff, "The evaluation of adaptive systems," in Patel, N. (Ed.): *Adaptive Evolutionary Information Systems*, Idea Group Publishing, London, pp.329-347, 2002.
- [6] T. Lavie, J. Meyer, K. Beugler, and J. Coughlin, "The evaluation of in-vehicle adaptive systems", *User Modeling: Work on the EAS*, pp. 9-18, 2005.
- [7] C. Gena, "Methods and techniques for the evaluation of user-adaptive systems," *the knowledge engineering review*, vol 20:1, pp. 1-37, 2005.
- [8] C. Gena, and S. Weibelzahl, "Usability engineering for the adaptive web," vol. 4321, *LNCS*, pp. 720-762, 2007.
- [9] L. Van Velsen, T. Van der Geest, R. Klaassen, and M. Steehouder, "User-centered evaluation of adaptive and adaptable systems: a literature review," *The knowledge engineering review*, vol. 23, pp. 261-281, 2008.
- [10] A. Nendjo Ella, C. Kolski, F. Wawak, C. Jacques, and P. Yim, "An approach of computer-aided choice of user interface evaluation criteria and methods," *Computer-Aided Design of User Interfaces II*, Kluwer Academic Publishers, Dordrecht, pp. 319-328, 1999.
- [11] X. Ferre, N. Bevan, and T.A. Escobar, "UCD method selection with usability planner," In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10)*. ACM, 2010.
- [12] A. Jameson, "Systems that adapt to their users: an integrative perspective," Saarbrücken: Saarland University, 2001.
- [13] K. Höök, "Steps to take before intelligent user interfaces become real," *Interacting with Computers*, 12(4):409-426, 2000.
- [14] P. Brusilovsky, P. Karagiannidis, and C. Sampson, "Layered evaluations of adaptive learning systems," *International Journal of Continuing Engineering Education and Lifelong Learning*, Vol. 14, Nos. 4/5, pp.402-421, 2004.
- [15] S. Weibelzahl, and G. Weber, "Advantages, opportunities and limits of empirical evaluations: Evaluating adaptive systems," *Künstliche Intelligenz*, vol. 3(2), pp. 17-20, 2002.
- [16] C. Karagiannidis, D. Sampson, and P. Brusilovsky, "Layered evaluation of adaptive and personalized educational applications and services," *10th International Conference on Artificial Intelligence in Education, Workshop on Assessment Methods in Web-based Learning Environments and Adaptive Hypermedia*, San Antonio, Texas, US, 2001.
- [17] E. Herder, "Utility-based evaluation of adaptive systems," *9th International Conference on User Modeling (UM2003)*, Second Workshop on Empirical Evaluation of Adaptive Systems, Pittsburg, US, pp. 25- 30, 2003.
- [18] J. Nielsen, and R.Molich, "Heuristic evaluation of user interfaces," *CHI '90*, Seattle, Washington, pp. 249-256, 1990.
- [19] J. Nielsen, "Heuristic evaluation," In: Nielsen, J., Mack, R.L. (eds.) *Usability Inspection Methods*, pp. 25-64. John Wiley & Sons, New York, 1994.
- [20] G.D. Magoulas, S.Y. Chen, and K.A Papanikolaou. "Integrating layered and heuristic evaluation for adaptive learning environments," In: *2nd Workshop on Empirical Evaluation of Adaptive Systems at UM2003*, Johnstown, PA, pp. 5-14, 2003.
- [21] S. Lawless, C. Mulwa, and A. O'Connor, "A proposal for the evaluation of adaptive personalised information retrieval," *The 2nd International Workshop on Contextual Information Access, Seeking and Retrieval Evaluation*, Milton Keynes, UK, p. 4, 2010.
- [22] A. Gupta, and P. Grover, "Proposed evaluation framework for adaptive hypermedia systems," *3rd Workshop on Empirical Evaluation of Adaptive Systems*, 23-26 August 2004, Eindhoven University of Technology, The Netherlands.
- [23] F.Tarpin-Bernard, I. Marfisi-Schottman, and H.Habieb-Mammar, "AnAmeter: the first steps to evaluating adaptation," *6th Workshop on User-Centred Design and Evaluation of Adaptive Systems at UMAP2009*, Trento, Italy: CEUR, pp. 11-20, 2009.
- [24] C. Mulwa and V. Wade, "A web-based evaluation framework for supporting novice and expert evaluators of adaptive elearning systems," *International Conference on E-Technologies and Business on the Web (EBW2013)*, Thailand, 2013.
- [25] H.R. Hartson, T.S. Andre., and R.C. Williges, "Criteria for evaluating usability evaluation methods," *International Journal of Human-Computer Interaction* 15, 145-181, 2003.
- [26] A. Sears, "Heuristic walkthroughs: finding the problems without the noise," *International Journal of Human-Computer Interaction* 9(3), 213-234 , 1997.