
Catherine Mulwa
Centre for Next Generation Localisation
Knowledge and Data Engineering Group
School of Computer Science and Statistics
Trinity College Dublin, Ireland
mulwac@scss.tcd.ie

Vincent Wade
Centre for Next Generation Localisation
Knowledge and Data Engineering Group
School of Computer Science and Statistics
Trinity College Dublin, Ireland
vincent.wade@scss.tcd.ie

Abstract—Based upon the analysed results of an evidence based study conducted over a period of four years, it is clear that it is difficult to identify, for a given evaluation objective, the range of appropriate evaluation techniques (methods, metrics, criteria and approaches to be used). This paper presents a Web-based evaluation framework which is freely available online, designed to support both novice and expert evaluators of adaptive e-learning systems when conducting evaluations. In order to do so we use a recommendation process to identify and recommend the appropriate evaluation techniques. The framework is populated using data extracted from 340 peer reviewed papers containing the evaluation details of adaptive systems. The dataset has the ability to grow over time as the framework itself provides a mechanism for published authors to add their own evaluations to the dataset; thus we believe the dataset is, and will continue to be, very valuable for supporting the design of evaluations of adaptive systems.

Keywords—Novice Evaluators, Evaluation Framework, Hybrid Recommender System, Personalised Search, Taxonomy, Webcrawler, Slicer

I. INTRODUCTION

This research is addressing the fact that it is difficult to see for what evaluation objective is the evaluation approach and the range of available evaluation techniques (approaches, methods, criteria and metric). The authors are convinced that there is a need to support both novice and expert evaluators of adaptive e-Learning systems (AES). In this paper a novice evaluator is a researcher who has no skills in evaluation of such systems and an expert evaluator is one who has 3+ years of research experience in evaluating such systems. Based upon the analysed results of an evidence based study that is conducted over a period of four years, it is clear that it is difficult to identify, for a given evaluation objective, the range of appropriate evaluation techniques. Novice evaluators of AES need more advice around what evaluation options are available to reach their evaluation goals or objectives. There are a lot of design choices between having an adaptive system and an objective of what you want to evaluate to the actual evaluation technique. This research investigates current evaluation challenges encountered by evaluators of adaptive systems and tries to address the question of: “Can we more effectively identify appropriate evaluation techniques (or a combination of techniques) for novice evaluators of such adaptive systems?” The evaluation techniques are concrete methods to carry out the validation of the system. We have looked at what people have evaluated in adaptive systems from 2000 to 2012, what evaluation techniques they have used and then we mapped these techniques to different evaluation methods, metrics and criteria. Based on evidence based study that was conducted over a period of four years, it is clear that evaluators need more advice around what evaluation options are available in order to reach their evaluation goal. Currently it is difficult to see for what evaluation objective are the available evaluation techniques and choices.

II. SUPPORTING NOVICE AND EXPERT EVALUATORS OF ADAPTIVE E-LEARNING SYSTEMS

E-Learning plays a major role in delivering educational material to the learners. Adaptive e-Learning refers to educational systems that adapt the learning content and the user interface according to the pedagogical and the didactical aspects. The aim of such systems is to provide appropriate information to the right student at the right time. AES systems are able to keep track of usage and to accommodate content automatically for each user and for the best learning result [1]. It is clear that the Evaluations of such systems are challenging, difficult and complex [2-5]. To address the challenges outlined in the introduction, we have specified, designed and developed a web-based online system for crawling evaluation studies published from 2000 to date [6]. The process of crawling is depicted in Figure 1.
As the evaluator triggers the crawling process, an RSS Feed is retrieved by the RSS FEED CRAWLER. Next, the RSS FEED CRAWLER sends a request to get the most recently published papers, and then it automatically creates one or more RSS FEED ITEM(S). An RSS FEED ITEM contains the meta-data about the published papers such as the title, author, published date-time and a URL to the paper document. Subsequently, the URL to the paper document is passed through the S.M.A.R.T URL ANALYSIS. If the URL is a downloadable document link, the analyzer will leave the URL untouched. Otherwise, the analyzer will try to ascertain the downloadable URL for the paper as discussed above. Finally, the downloadable URL is passed to the Document CRAWLER, which uses Client URL (cURL) to retrieve the document and create a hard pdf copy on the local document repository. The system is currently being used by the researchers to crawl evaluation studies of adaptive systems developed from 2000 to 2012, more specifically focusing in adaptive technology enhanced system. To date we have crawled over 400 studies. A lot of time and effort was saved by using this system. The crawled studies were then manually sliced to create an educational evaluation dataset which was then used to populate the database of the framework proposed in this paper.

III. PROPOSED WEB-BASED EVALUATION FRAMEWORK

To address the question of “whether we can more effectively identify appropriate evaluation techniques (or a combination of techniques) for novice evaluators of such adaptive systems”, we have specified, designed and developed a software framework for supporting novice and expert evaluators in evaluating adaptive systems (EFEx)\(^1\), more specifically AES systems (Refer to Fig 2).

---

1. http://46.22.130.121:8080/effeALL/home.jsp/
Now, suppose a novice evaluator wants to get recommendations on which evaluation method to use when evaluating an AES system. In order to recommend the most appropriate evaluation method, during the implementation of the recommendation service we have considered the following main factors (refer to Table 1). A glossary of these factors can be found in Table 2.

### Table 1: Factors Considered When Recommending an Evaluation Method to Use

<table>
<thead>
<tr>
<th>Factor</th>
<th>Normalized Value</th>
<th>Weight</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of publications in which the evaluation method was used</td>
<td>NP/TP</td>
<td>1</td>
<td>“Because the evaluation method has been used NP times out of TP times in the literature (2000-2012)”</td>
</tr>
<tr>
<td>The types of (venues) publications that the evaluation method has been used in (e.g. journal, conference and workshop)</td>
<td>NV/TV</td>
<td>1</td>
<td>“Because the evaluation method appeared in (Nj journals), and (Nc conferences), and (Nw workshops)”</td>
</tr>
<tr>
<td>How many adaptive e-Learning systems belonging to the same variation type (category) have been evaluated using the evaluation method</td>
<td>NSV/TSV</td>
<td>2</td>
<td>“Because, out of the Tsv systems which belonged to the “V” variation Type, Nsv of them have been evaluated using the evaluation method”</td>
</tr>
<tr>
<td>Give an extra weight to the evaluation method according to its association with the selected evaluation purpose</td>
<td>NEPV/TEPV</td>
<td>2</td>
<td>“Because, out of your TEPV selected evaluation purposes, NEPV of them are associated with the evaluation method”</td>
</tr>
</tbody>
</table>

Throughout this process the evaluator is provided with explanations as to why each factor was taken into consideration and why that evaluation method was recommended. Explanations are significant because they provide transparency, validity, trustworthiness, persuasiveness, effectiveness, efficiency, satisfaction, relevance, comprehensibility and education [7]. The recommended methods are then computed into a score, these scores are then presented to users as stars. For example the most appropriate method is ranked at the top with 5 stars and the least appropriate with only 1 star.

2. A Personalised Search System whose database is populated using 250 evaluation studies is published from 2000 (Refer Fig 4). In this case the users are provided with an automated personalised search interface, which allow them to find evaluations of 15 inner models of adaptive systems, evaluation of 106 adaptive systems developed from 2000 and over 130 studies of such systems.

<table>
<thead>
<tr>
<th>Glossary</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>P:</td>
<td>Publication (published study)</td>
</tr>
<tr>
<td>S:</td>
<td>System</td>
</tr>
<tr>
<td>V:</td>
<td>Variation Type (category)</td>
</tr>
<tr>
<td>EVP:</td>
<td>Evaluation Purpose (Goal)</td>
</tr>
<tr>
<td>N:</td>
<td>Total Number of…</td>
</tr>
<tr>
<td>T:</td>
<td>Total Number of…</td>
</tr>
<tr>
<td>NP:</td>
<td>Number of publications published that used the evaluation method</td>
</tr>
<tr>
<td>TP:</td>
<td>Total number of publications in the database</td>
</tr>
<tr>
<td>Tj:</td>
<td>Total number of Journal Papers in the publications table</td>
</tr>
<tr>
<td>Tc:</td>
<td>Total number of Conference Papers in the publication table</td>
</tr>
<tr>
<td>Tw:</td>
<td>Total number of workshop papers in the publication table</td>
</tr>
<tr>
<td>TV:</td>
<td>Total venue score of all publications (i.e. calculated as: 4<em>Tj + 2</em>Tc + 1 * Tw)</td>
</tr>
<tr>
<td>Nj:</td>
<td>Number of Journal Papers in which the evaluation method was used</td>
</tr>
<tr>
<td>Nc:</td>
<td>Number of Conference papers in which the evaluation method was used</td>
</tr>
</tbody>
</table>

### Table 2: Glossary of the Factors

| Table 2: Personalised Search System |

3. A Taxonomy of evaluation techniques which assists novice evaluators in understanding different aspects of the evaluations of adaptive systems (Refer Fig 5).
Q1 How would you rate your evaluation skills of adaptive systems
The aim of Q1 was to identity which participant was an expert.

Q3 Please select which system characteristics you wish to focus on during evaluation. (*Please select as many as you like)

Q4 Based upon the category (variation type) of adaptive system and the system characteristics that you have selected. What would be the goal(s) or purpose(s) of the evaluation being conducted (Select as many as you like)

Q5 What Kind of question(s) would you wish to answer during evaluation (Select as many as you like)

Q6 If you were to conduct the evaluation in the form that you have selected above, what would it help you to improve in the system?

Q7 Which of the following evaluation approach(s) would you recommend to be used when evaluating an adaptive system(s) belonging to the variation type (i.e. you choose in Question 3)? Please rate the recommended approach

Q8 Which of the following evaluation methods do you feel are appropriate for evaluating systems of the variation type you choose in Q3? Please use the drop-down lists below. You can recommend one or more evaluation methods by rating them from most appropriate to not appropriate.

Q9 Which of the following evaluation criteria do you feel are appropriate for evaluating systems of the variation type you choose in Q3? Please use the drop-down lists below. You can recommend one or more evaluation criteria by rating them from most appropriate to not appropriate.

Q10 Which of the following evaluation metrics do you feel are appropriate for evaluating systems of the variation type you choose in Q3? Please use the drop-down lists below. You can recommend one or more evaluation metrics by rating them from most appropriate to not appropriate.

Q11 Which of the following evaluation techniques (i.e. the techniques you recommended in Q9) would you bundle to be used together? The term “bundle” refers to appropriate combination of a method/criteria/metric that can be used together when evaluating the properties you selected in Q4-Q7.

Q12 Do you consider such a system useful?

Q13 In this hybrid recommender system, we want to provide explanations on how the recommended techniques are derived. Do you think this would be a useful feature?

Results show that, adaptive systems developed from 2000 to 2010 could be categorized into 14 different variation types. Figure 6 presents the percentage response of experts who gave recommendations on how to evaluate an AES system after answering Q2. From Q3 to Q6, before the experts could provide recommendations, we wanted to know which properties they would give recommendations on. Based on the characteristics of selected category of adaptive system, identify the ones you would like to focus on during recommendations. Q7 to Q11 were aimed at recommendations on different evaluation techniques. In Q12 and Q13, we wanted to find out if the recommender system was significant.

B. Results

This section presents the results of the 32 expert evaluators who gave recommendations on how to evaluate AES systems. Figure 7 presents a summary of recommended evaluation approaches to be used when evaluating AES systems. We compared the recommended approaches with the results produced by the recommender system. The system produced same recommendations like the ones ranked by the experts as most appropriate. The user-centered evaluation approach (UCEA) was the highest, followed by layered approach then the utility-based. When asked whether they considered the recommender system useful to the community 91 percentage responded yes.
Figure 6: Variation Types (Categories) Adaptive Systems

Figure 7: Recommended Evaluation Approaches (Q8)

Figure 8: Usefulness of Explanations of Recommended Methods, Metrics and Criteria (Q13) and Recommender System (Q12)
Figure 8 presents the response to Q13. The average rating showed that all explanations were considered significant with explanations for recommended approaches being rated most useful followed by evaluation methods, metrics and criteria respectively.

Currently we are comparing the results of the recommended evaluation methods, metrics and criteria. The final results will be compared with recommendation produced by the recommender system and published in a later date.

C. Findings

Based on the results in Figure 7, we can positively answer question 1 because the 80% of the subjective answers by expert evaluators, matched the answers automatically computed by our recommender. Also question 2 can be answered positively because as depicted in Figure 8, the majority of experts tend to see the recommender system explanations very useful and useful.

CONCLUSION AND FUTURE WORK

In conclusion we have partially tackled the question being addressed by this research “Can we more effectively identify appropriate evaluation techniques (or a combination of techniques) for novice evaluators of such adaptive systems?” Also the educational evaluation dataset introduced has the ability to grow overtime as the framework itself provides a mechanism for published authors to add their evaluation cases to the dataset; thus we believe the dataset and the evaluation framework for supporting novice evaluators of AES systems is significant and in the future could become a very valuable tool.

In future we will analyse the remaining the results of the remaining user trials which are currently underway and then deploy the framework online.

ACKNOWLEDGMENT

This research is based upon works supported by Science Foundation Ireland (Grant Number: 07/CE/11142) as part of the Centre for Next Generation Localization (www.cngl.ie). The authors are grateful for the suggestions of the reviewers for this paper.