

OSSES: An Online System for Studies on Evaluation of Systems

Catherine Mulwa, Mary Sharp, Vincent Wade
Knowledge and Data Engineering Group
School of Computer Science and Statistics
Trinity College, Dublin
mulwac@scss.tcd.ie, Mary.Sharp@scss.tcd.ie, Vincent.wade@scss.tcd.ie

Abstract

In recent years, due to exponential increase in the number of internet users, finding the appropriate information is difficult. Web crawlers represent a significant component in Web search engines. This contribution describes a Web based distributed system designed as a 3-tier architecture consisting of: the presentation layer, the business logic layer and the data persistence layer. The main educational benefit of our system is to provide a reference tool that has an interactive database to encourage evaluations of systems that fulfill certain methodological requirements. The synopsis of studies collected can be used as a basis of a searchable online database that provides an overview of the state-of-the-art to the scientific community and encourages other scientists to evaluate their own system. It will also help students identify pitfalls in the planning process as well as in the analysis of collected data and also identify omissions in the state-of-the-art in future. The collaborative nature of our tool enables sharing information among research students providing them a larger view of the state-of-the-art. The architecture includes RSS Feed Management, Paper Subscription, Smart URL Analysis and Document Downloading. The RSS Feed Management allows a user to manage a set of Web feed formats that will publish most recent papers. As soon as a paper is published via RSS Feed, the paper subscription module automatically creates metadata. Upon receiving the document link, the Document Downloading module copies the document to a local repository. The system has been tested and evaluated showing good performance.

Keywords: adaptive systems, 3-tier architecture, Object Oriented, RSS Feed, and Search engine, S.M.A.R.T URL and Web Crawler.

1. Introduction

Due to the rapid increase of Web linked information on the WWW, it has become difficult for the search engines to find exact appropriate information. The use of such large scale search engines, such as Google are very common for surfing the World Wide Web. The capability of these search engines is impressive. Search engines have five components: a crawling module, an indexing module, a page ranking module, a search module, and a page repository (Olston & Pandey, 2008). The crawling module is responsible for the process of downloading documents from the internet. This process is done by Web crawlers which start with a set of seed URLs, download Web pages, and extra links from the downloaded pages for further download. This paper describes the implementation of an online system for studies on the evaluation of systems (OSSES) from 1990 to date. The system is composed of two crawlers (RSS Feed and Document Crawler) and an online database. The Object-oriented approach to software engineering was adopted which involved six development activities; requirement elicitation, analysis, system design, implementation, testing and evaluations (Bruegge & Dutoit, 2009).

Our goals were to: i) design and implement an online system capable of crawling and automatically retrieving the most recently published studies on evaluations of systems that fulfill certain methodological requirements, ii) encourage new researchers from different diversities to perform research on evaluations of systems and iii) develop an online database which will be used to identify gaps in the state of the art and act as a guide for planning new evaluations.

The main educational benefit of our system is to provide a reference tool that has an interactive database to encourage evaluations of systems that fulfill certain methodological requirements. The synopsis of studies collected can be used as a basis of a searchable online database that provides an overview of the state of the art to the scientific community and encourages other scientists to evaluate their own system. It will also help students identify pitfalls in the planning process as well as in the analysis of collected data and also identify omissions in the state of the art in future. For people outside the community the system will serve to teach them the usefulness or insufficiency of certain types of systems. The system can also be used to inform student researchers of currently used methods and criteria as well as omissions and problems in current evaluations. Our contribution supports education because it provides new knowledge to researchers towards a larger view of the state of the art in their research field. The collaborative nature of the architecture enables sharing information among similar users.

Currently the system is being used by the author to collect and to analyze studies on user evaluations of adaptive systems especially adaptive systems which combine adaptive hypermedia and information retrieval

techniques. The online database will help to identify pitfalls in the planning process of evaluations as well as in the analysis of the collected data. It is important that evaluators avoid well-known pitfalls and that writer's of future evaluation reports increase their empirical value, by reporting the methodology used and results in such a fashion that replication of the study is possible. We are convinced that the quality of evaluations will benefit and that, indirectly, the user will be served in the process. The system can also be used by different diversities of researchers to automatically retrieve published papers in PDF format as soon as they are published and store them in a local repository. For example if a paper has been published and is available in Google Scholar, instead of searching through Google Scholar manually, the paper will be retrieved automatically and stored in a local repository. The system searches the first ten pages in Google and from each page; it retrieves the first four PDF documents and also eliminates duplicates. This is very significant because due to the large amount of available information, the Google search engine generates a set of documents dynamically, which results in a different result for every new search query, in our case, the query is an RSS feed. The system can also be used to serve as a reference for researchers in the field of system evaluation and assurance; (i.e. evaluation of adaptive systems, especially adaptive systems which combine adaptive hypermedia and information retrieval techniques). By providing a searchable categorized set of studies, interested people get suggestions of experimental designs, criteria and other experimental issues.

The remainder of this paper is structured as follows: Section 2 reviews related work. Section 3 describes the OSSES system implementation. It also introduces the system methodology, architecture, the crawling process, database design and system functions. Section 4 presents the performance testing and evaluation. Section 5 discusses the advantages and limitations of the OSSES. Finally Section 6 concludes the paper and recommends some future work.

2. Related Work

Evaluation is an important tool of software quality assurance. A typical software quality program involves: i) establishment, implementation, and control of requirements, ii) establishment and control of methodology and procedures, and iii) software quality evaluation (Farooq, 2008). The software quality evaluation component is aimed at evaluating products (both in process and at completion), and methodologies (for appropriateness and technical adequacies). Evaluation of all systems is important. It is important to not only evaluate but also to ensure that the evaluation uses the correct method (Brusilovsky, 2004). Evaluation is defined as the process of examining the product, system components, or design, to determine its usability, functionality and acceptability (S Weibelzahl, 2003), which is measured in terms of a number of criteria essential for any software development project. Why should evaluation take place and who are the stakeholders? It is carried out by designing an evaluation plan which consists of: specifying goals, decomposing goals into evaluation questions, setting criteria for deciding questions, identifying data required to answer questions, selecting methods for collecting data and analyzing data, conducting experiments either in the experimental lab or in the field; collecting data that can be qualitative and/or quantitative, or subjective and/or objective; analyzing and interpreting the data; and reporting and drawing conclusions based upon the established hypothesis. System evaluation places an emphasis on the comparison of the presented system with established criteria proposed by other researchers or other related systems.

A **Web crawler** is a program that automatically traverses the Web's hyperlink structure and retrieves some information for the user. The behaviour of a Web crawler is the outcome of a combination of policies (Girardi, Ricca, & Tonella, 2006): i) A **selection policy** that states which pages to download. If the set S is implemented as a stack, the algorithm visits the Web site in depth-first visit. ii) A **politeness policy** that states when to check for changes to the Web pages. The use of Web crawlers is useful for a number of tasks, but comes with a price for the general community. The costs of using Web Crawlers include network resources and servers overload. A partial solution to these problems is the robots exclusion file which indicates which Web sites (or portions) should not be accessed by crawlers iii) A **parallelization policy** that states how to coordinate processes or threads in the downloading phase. Crawlers are used by search engines to make a local copy of the Web. This copy is then indexed for further fulfillment of user queries and iv) A re-visit policy that states when to check for changes to the Web.

There is limited research on user evaluations of adaptive systems. An adaptive system tailors its output, using implicit inferences based on interaction with the user. Currently, the system contains studies and evaluated systems on user evaluation of adaptive systems particularly adaptive systems which combine adaptive hypermedia and information retrieval techniques. For each of these studies, the relevant information is accessed, processed and recorded in the database: this information include system name, function, application area, evaluation method, criteria used, purpose of the system, and data type analysis. The evaluation of adaptive systems is not easy, and several researchers have pointed out potential pitfalls when evaluating these systems. Examples of pitfalls mentioned in (Tintarev & Masthoff, 2009) and (S Weibelzahl, 2005) include: i) Difficult in attributing cause, ii) Insignificant results due to variance between participants, iii) Difficult in defining the

effectiveness of adaptation, iv) Allocation of insufficient resources, v) Too much emphasis on summative rather than formative evaluation and vi) Most importantly measures for Adaptivity success have not been investigated systematically up to now. A few researchers have compiled a synopsis of studies published on the evaluation of adaptive systems; Weibelzahl compiled a synopsis of 43 studies published in the UMUIAI journal and in the proceedings of User Modeling Conferences (1997 -1999), and considered the most important works in the area of adaptive systems. He accepted that about a quarter (11/43) of the studies examined, had either only a single user, hypothetical users or that the sample size was not reported. Actually, only 14 out of 43 studies were high quality in terms of sample size and statistical analysis. He also reported that the most frequent measures include accuracy, precision and recall, domain knowledge and duration of the interaction. Last year Van Velsen conducted a literature review study identifying 63 studies on user-centered evaluations of adaptive and adaptable systems (Van Der Geest T, Van Velsen, Klaassen, & Steehouder, 2008). Of the systems mentioned in these studies; 23 systems were adaptive systems they fail to identify which studies were only on user-centred centered evaluations of adaptive systems. During the process of a literature review it was discovered that although a few researchers have researched the evaluation of adaptive systems, very little research has been conducted on the user evaluation of adaptive hypermedia systems (AHS) and there is a lack of studies in this field. The past 16 years have seen notable progress in the field of AHS. Many systems were created and some of them brought new, interesting, and unusual ideas. Unfortunately a few were almost forgotten since their documentation was difficult to find. AHS are systems composed of three main elements: user models, domain models and adaptation models. The user model is based on the user characteristics. It contains for instance their prior knowledge among other things. The domain model contains for instance the information concerning the subject area that the system is built around and finally the adaptation model which describes how the navigation will be carried out and also how the material will be presented.

A few researchers have implemented online databases with similar functions to the OSSES, but some of these databases are out of date(Weibelzahl & Weber, 2001) making it difficult to obtain clear and up to date metadata for the evaluation of adaptive systems. This system is a valuable tool to new PhD students since it will help reduce cost and time spent while conducting literature reviews.

3. OSSES System Implementation

3.1 Our Methodology

The Rational Unified Process (RUP) Methodology was used. By adopting the RUP method the developer was able to manage the Object-oriented software development approach since it is extensible and has features such as: iterative development, requirement management, and component-based architecture vision, visual modeling of the system, quality management and change control management.

3.1.1 System Architecture

The OSSES is designed as typical 3-tier architecture (see Figure 1) which consists of: **i) the presentation layer** which is the topmost level of the application which displays information related to services such as browsing. It communicates with other tiers by outputting results to the browser/client tier and all other tiers in the network. **ii) The business logic layer** which is pulled out from the presentation tier and, has its own layer, it controls an application's functionality by performing detailed processing and **iii) the data persistence layer** this tier keeps data neutral and independent from application servers or business logic. Giving data its own tier improves scalability and performance. The presentation layer is implemented using Java Server faces (JSF) technology. By leveraging the Apache MyFaces library, two modules are implemented namely; the Administration Interface module and the end User Interface module. The second layer is the business logic layer which consists of four modules namely; RSS feed management, subscription to newly published paper, Self-Monitoring, Analysis, and Reporting Technology (S.M.A.R.T) analysis of paper URL and document downloading. The third layer is the data persistence layer which consists of a MySQL backend Database Management Systems (DBMS) and the mapping from Java classes to relational database tables.

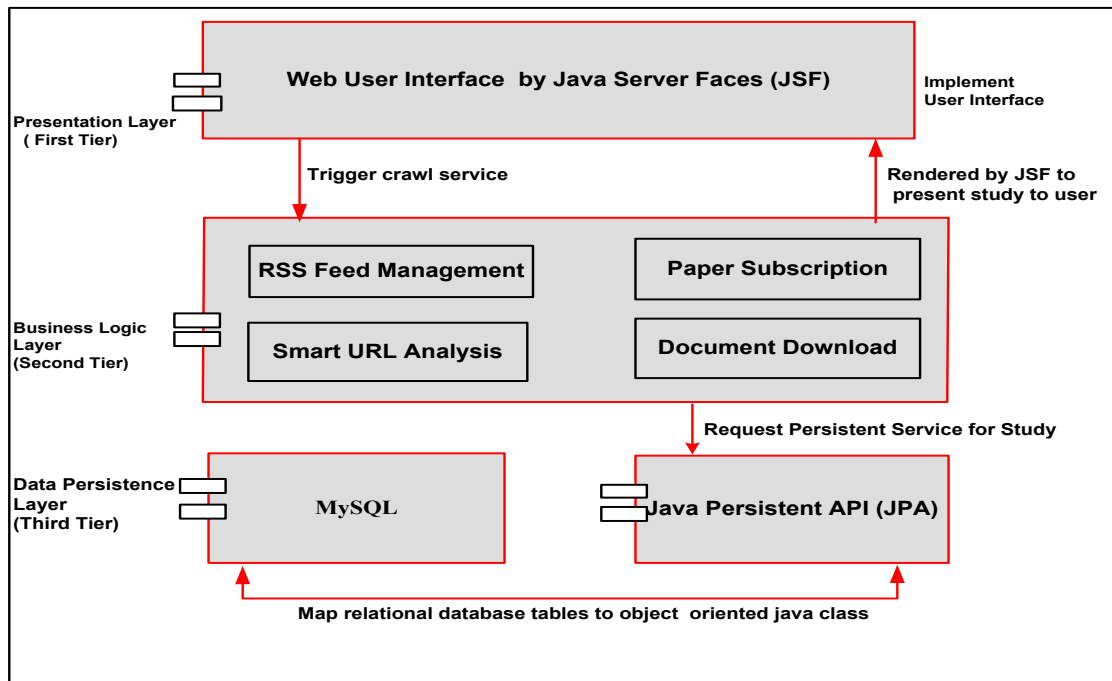


Figure 1 OSSES Architecture

The core of the system is the business logic layer in the middle (see Figure.1), the RSS Feed Management allows users i.e., system administrators, to manage a set of Web feed formats that will publish most recent papers to the user. As soon as a paper is published via RSS feed, the Paper Subscription module will create Meta data for that particular paper, such as title, author, published date, and most importantly the URL for the hard copied paper. Ideally, the URL is supposed to download a published paper document directly. But in the real world, the URL can point to indirect location, such as a Web page containing a link to the actual document or a search result pointing to the document. To solve the indirect URL problem, the S.M.A.R.T URL Analysis module will testify if the URL is directly pointing to document. If not, it will do a programic search in Google Scholar by entering the title of the paper and return the first downloadable link to Document Downloading module. Upon receiving the document link, the Document Downloading module will automatically retrieve the document using the Client URL (cURL) tool and store it to a local repository.

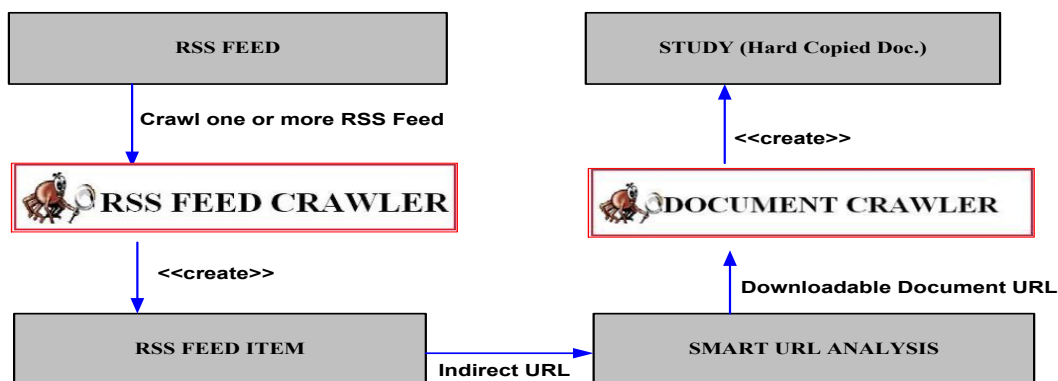


Figure 2 Process Involved in Crawling to Retrieve Published Paper

3.1.2 RSS Feed Process of Crawling

The process for the flow of data involved in retrieving a published paper from RSS is shown in Figure 2. The user (i.e., System Administrator) triggers the crawling process; an RSS Feed is retrieved by the RSS FEED CRAWLER. The RSS FEED CRAWLER sends a request to get the most recently published papers, and then creates one or more RSS FEED ITEM(S). An RSS FEED ITEM contains meta-data about the published papers such as (title, author, published date time and a URL to the paper document) which is generated automatically. Subsequently, the URL to the paper document is passed to 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 copy on the local document repository.

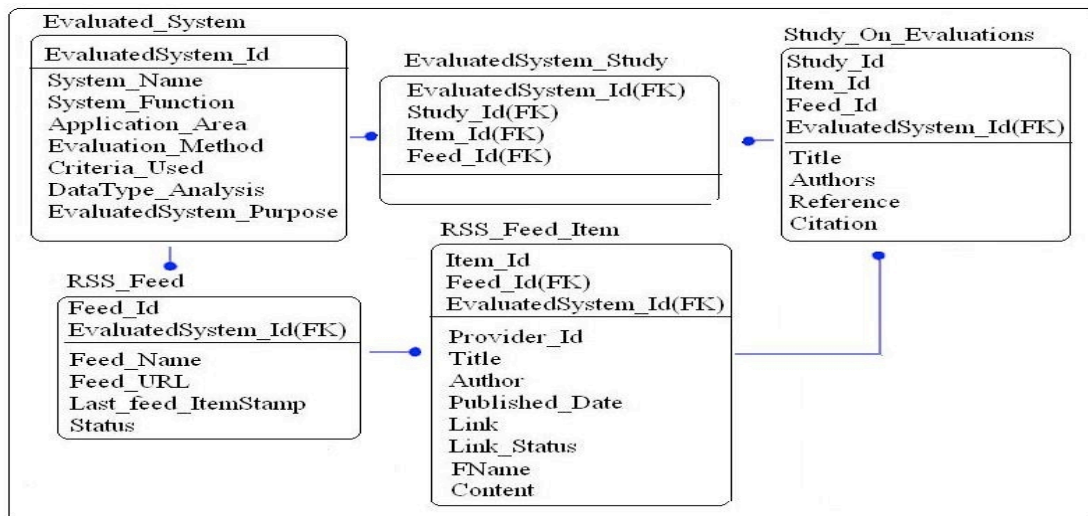


Figure 3: OSSES Database Design

3.1.3 OSSES Database Design

For each published study, the system automatically retrieves the published study from the Web. Each study has a title, author, published date, link content, citation and reference. Each evaluated system mentioned in the retrieved studies is described in terms of: system name, the functions it fulfils, the purpose, application area, evaluation methods and criteria used and data type analysis. If the Web crawler retrieves a study that is not relevant; the system administrator should delete it from the system database (see figure 4). In order to log into the Administrator User Interface, the user is required to obtain authentication from the system administrator. As a system administrator, the user can perform all the tasks (select study or system, view transaction, search for studies or system, modify system operations, delete or edit transactions, add system and study details) see Figure 4 and 5.

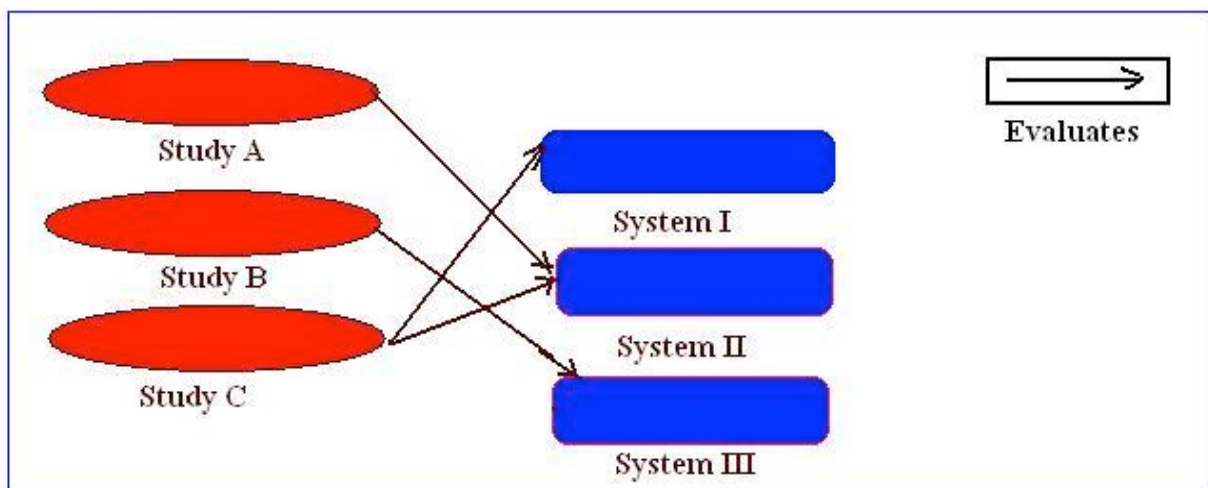
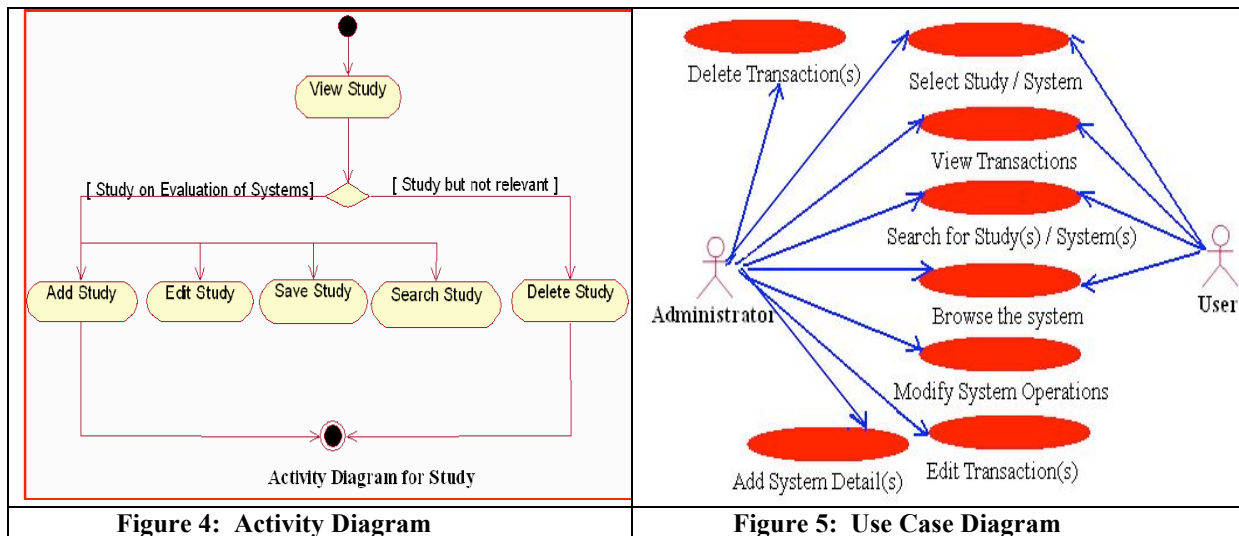


Figure 3.1 Relations of Evaluation Studies and Evaluated Systems

The relationship between evaluation studies and evaluated systems is demonstrated in figure 3.1 for example while one study might evaluate one or more systems (e.g., Study C), a system might be evaluated by one or more studies (e.g., System II).

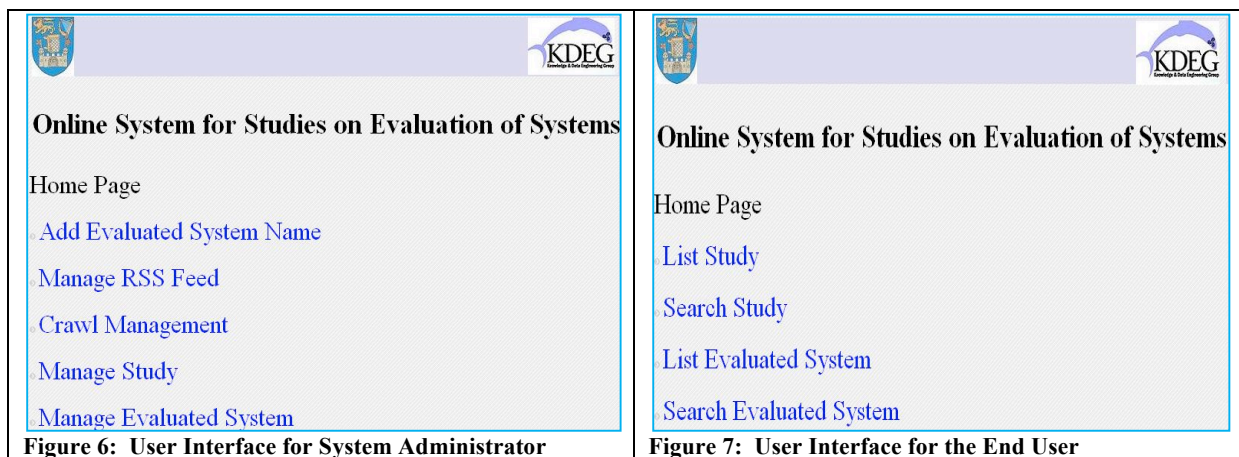
3.1.4 Software Components

The system is implemented by integrating current technologies and software's: (**Eclipse** - platform, **Apache-Openjpa** - To store and Retrieve data from database, **Apache-Tomcat** - server, **Client URL** (cUrl) tool-Downloading Tool for PDF, **Myfaces-core** - Java Server Faces (JSF) used to display data in the Web, **MySQL-win32** - MySQL database server, **MySQL-connector-java** - Connector for java to communicate to mySql, **Json** - To parse RSS Feed and **Self-Monitoring, Analysis, and Reporting Technology (S.M.A.R.T)** URL analysis – a system for monitoring computer hard disks to detect and report on various indicators of reliability) which has led to an increase in performance and a better user interface.



3.2 The OSSES System Functions

The system has two User Interfaces: System administrator and end user as shown in Figure 6 and 7. Authentication is required for the Administrator Interface.



The system administrator has full control of all the system transactions (see figure 4, 5, 6 and 7). Other users can only view study(s), search for study(s), view evaluated system(s) details and also search for evaluated system(s) (see figure 5 and 7). Following is brief description of the system functions and activities the user performs.

A. Administration Functions

- 1) **Add Evaluated System Name:** This function allows the system administrator to Quickly add, edit, and view Evaluated System details (e.g, System Name, Functions, Application Area, Evaluation Method, Criteria, System Purpose).

Figure 8: Adding RSS Feed into the System

Figure 9: Attributes of an Evaluated System

- 2) **Manage RSS Feed:** Once a system name has been added, the RSS feed management function allows the administrator to: add feed name and the URL link. It also provides a drop down list of Systems which were added in function 1. (see figure 8). Once the feed has been added the user is able to see the Feed title and the URL link (see figure 10)

FEED	URL	ACTION
Communications of the ACM: Information Systems	http://cacm.acm.org/browse-by-subject/information-systems.rss	REMOVE

Figure 10: Crawled Feed and the URL link

- 3) **Crawl Management:** The crawl management function performs two tasks (first the system administrator has to crawl the studies and then crawl documents). This task can only be performed when function 2 has been completed (see figure 11).

Figure 11: The Process of Crawling

Figure 12: Searching for a System

- 4) **Study Management:** This function allows the system administrator to view all the crawled studies from function 3 (e.g., title, authors) and perform actions such as editing the study details (e.g., title, authors,

reference, citations) and also provides a Quick Link to the drop down list of all the Evaluated Systems (see figure 13).

Figure 13: Attributes of Published Paper

- 5) **Evaluated System Management:** This function allows the administrator to enter details of the system (name, function, application area, evaluation method, criteria and purpose), perform actions such as editing the system details and also view which study is linked to that particular system and finally provides details of that study (see figure 9).

B. Usage Functions

- 6) **List of Studies (Papers):** The user can view the list of all studies, search a specific study, perform actions (e.g, Details of the study such as title, author, reference, citation, systems linked to that study, system details such as system name, application area, functions).
- 7) **Search study:** Users can search for studies and view study details (e.g, title, authors, references and citations)
- 8) **List of Evaluated System:** This function provides a list of all the evaluated systems described in the published studies. It also allows users to perform actions such as: i) view details of the evaluated system and , ii) the studies (title and authors) describing that particular system.

SYSTEM	FUNCTION	APPLICATION AREA	EVALUATION METHOD	CRITERIA USED	PURPOSE
ISIS-TUTOR					Is rather old system, the first version appeared around 1992. It is also one of the few non-web-based AHES. It uses ISIS print-formatting language to display the output. Only adaptive link annotation on the basis learned, ready to be learned and not ready to be learned is supported. It belongs to the pioneers of AHES

Figure 14: Retrieved System after Search Query

- 9) **Search Evaluated System:** The user is capable of searching for existing systems by using the following search terms(i.e, system name, function, application area, evaluation methods, criteria used, data type analysis, evaluated system purpose). For example if a user is searching for a system named “ISIS-TUTOR” (see figure 12) and it exists in the database, it is displayed with all other details relating to that system (see figure 14).

4. OSSES Testing and Evaluation

The validation of the OSSES system prototype was subdivided into two distinct tasks; functional verification and efficiency evaluation. Functional verification was used in order to verify all the functional requirements and the efficiency evaluation to ensure that user satisfaction were met. We used different software testing elements: **Methods and techniques** which included; Information Retrieval techniques, interviews, expert reviews and log file production during downloading the crawler produces a log file containing some information on the PDF documents of the downloaded studies, **Process, Empirical knowledge, Tools** the (Self-Monitoring, Analysis,

and Reporting Technology (S.M.A.R.T) which is a monitoring system for computer hard disks to detect and report on various indicators of reliability, in the hope of anticipating failures during the process of crawling.

Two different types of evaluations were conducted, **formative** and **summative** evaluations. Formative evaluation was conducted during the implementation process. A range of evaluation methods were used: i) interviews with domain experts, ii) tasks and results, iii) knowledge and data engineering research group. To evaluate the crawler, we considered features such as completeness, robustness and download limiting and the graphical User Interface. The OSSES system was evaluated internally by: i) Our research supervisors, ii) Presentation of the deployed system to a group of researchers in Knowledge and Data Engineering Group (KDEG) during our internal summer/chi workshop in college. The KDEG research group is pioneering research into the fundamental challenges and application of knowledge driven systems. The group combines innovative technology research in knowledge discovery, representation, reasoning, data management and intelligent systems engineering. The following are the questions asked by evaluators and our response:

- ✦ *Why should we use your system while we can use Google search engine? It will save the end users time and also encourage research in the area of evaluations of systems.*
- ✦ *Why use of different technologies and software's? Our response was these technologies and software's were used in order to increase the system performance and also for a better user interface.*
- ✦ *What are your future plans for this system? Our future plan is to add more functions that are specific to user evaluations of adaptive systems.*
- ✦ *How did you test the system? System was tested by functional verification and efficiency evaluations.*
- ✦ *How do you deal with retrieved studies that are not relevant? The system administrator deletes irrelevant studies, as demonstrated by the activity diagram (see figure 4).*
- ✦ *Can new functions be added that are relevant to our research? Yes. One of the research students wanted to start using the system immediately.*

In addition summative evaluation was conducted to provide information on the system's ability to perform better. In order to determine how well the system performed several **evaluation criteria** were used: Evaluation of input data (e.g., objectivity of data assessment, retest-reliability), Evaluation of Adaptation decision (e.g., retrieval accuracy, precision and recall, amount of help required, computational time; number of navigation steps, task success, user satisfaction), usability satisfaction, effectiveness, reliability, functionality, performance, time, robustness, downloading limiting and completeness.

5. Discussion

5.1 OSSES Advantages

- ✦ The system provides services that assist researchers in their Web navigation by automating the task of link traversal, creating a searchable index of the Web and fulfilling searchers' queries from the index.
- ✦ Our Web crawler's simplification of the Web experience makes the Web a more friendly and useful tool.
- ✦ This system is useful because it can provide some context for a searcher's particular query: by issuing a well-formed query, a searcher can find the breadth of information about that particular topic and can use that information to further refine his goal.

5.2 OSSES Limitations

- ✦ All the retrieved study(s) do not provide author name(s), results returned are "Unknown author".
- ✦ Once a study(s) is retrieved, the system administrator has to process the information and manually enter this information in the database.

6. Conclusion and Future Work

This research will support a PhD by encouraging new researchers from different diversities to research the evaluation of systems which fulfill certain methodological requirements. It will also serve as a reference for researchers in the different fields of evaluations of any kind of system; for example research on user evaluations of adaptive systems especially those that combine adaptive hypermedia and information retrieval techniques. The online database will help to identify gaps and pitfalls in the planning process of evaluations as well as in the analysis of collected data. It is crucial that evaluators evade well-known pitfalls and that writers of future evaluation reports increase their empirical value, by reporting the used methodology and results in such a fashion that replication of the study is possible. A user who wants to use a Web crawler has two choices: building it from scratch or downloading one from the internet. The second option has some drawbacks such as the user deciding which one to choose? Which is the best for the task at hand? Which is most complete? Which is the most robust? We provide a fully functional Web crawler which is capable of automatically retrieving recent published studies. We are convinced that the quality of evaluations will benefit and that, indirectly, the user will be served in the process.

In future extra functions will be added which are specific to research on user evaluations of adaptive systems. To find an algorithm which will be used to automatically extract references of the studies. Currently this task is performed manually.

Acknowledgements

This research is supported by Science Foundation Ireland (SFI) as part of the Center for Next Generation Localisation (www.cngl.ie).

References

- Bruegge, B., & Dutoit, A. (2009). *Object oriented software engineering*: Prentice Hall.
- Brusilovsky, P., Karagiannidis, P., and Sampson, C. (2004). Layered Evaluations of Adaptive Learning Systems. *14*, 402-421.
- Farooq, A., Dumke, Reiner R. (2008). *Evaluation Approaches in Software Testing*: Faculty of Computer Science, University of Magdeburg.
- Girardi, C., Ricca, F., & Tonella, P. (2006). Web crawlers compared. *International Journal of Web Information Systems*, 2(2), 85-94.
<http://kdeg.cs.tcd.ie/>
<http://kdeg-vm-10.cs.tcd.ie:8080/OSSES/faces/user/>
<http://kdeg-vm-10.cs.tcd.ie:8080/OSSES/faces/admin/>
- Olston, C., & Pandey, S. (2008). *Recrawl scheduling based on information longevity*.
- Tintarev, N., & Masthoff, J. (2009). Evaluating Recommender Explanations: Problems Experienced and Lessons Learned for the. *UMAP 2009*, 54.
- Van Der Geest T, Van Velsen, L., Klaassen, R., & Steehouder, M. (2008). User-centered evaluation of adaptive and adaptable systems: a literature review. *The Knowledge Engineering Review*, 23(03), 261-281.
- Weibelzahl, S. (2003). *Evaluation of Adaptive Systems*. University of Trier, Trier.
- Weibelzahl, S. (2005). Problems and pitfalls in evaluating adaptive systems. In: *Fourth Workshop on the Evaluation of Adaptive Systems in conjunction with UM'05*, 57-66.
- Weibelzahl, S., & Weber, G. (2001). *A database of empirical evaluations of adaptive systems*.