New Honours Project Archive

Transition of the Honours Project Archive to the SimpleDL Toolkit

Richard Paterson
Commerce
University of Cape Town
Western Cape South Africa
+27721234197
ptrric011@myuct.ac.za

Simangaliso Mncwango
Computer Science Honours
Western Cape, South Africa
+27836956857
mncsim006@myuct.ac.za

CCS Concepts

• Digital Libraries • Data Management

Keywords

Digital Libraries, Data Management, Simple DL, Archiving

1. INTRODUCTION

The Computer Science department at the University of Cape Town has stored and hosted a collection of honours projects since 2003. These archives display the heritage of the Computer Science department and allow visitors to see the development of technology and formation of knowledge that has happened over the past 20 years. This archive is accessible to all members of staff, students (current and past), and even general members of the public who are interested in accessing such a rich collection of works. While this archive allows such access, the archive's main software solution was created at the same time the archive itself began. And over the past 20 years this software solution has largely remained the same. This software is still completely functional, however, new and more advanced solutions are available now that could improve upon the current archive.

One such solution is using the Simple DL toolkit. This project plans to utilize the Simple DL toolkit to create a new Honours Archive that has the potential to provide a better archive solution that would be more scalable, simpler to use, more resistant to poor networks and also allowing for a more modern and user-friendly interface in which to interact with.

2. Related work

2.1 User Centered Design in Digital Archives

When designing digital archives, consideration needs to be given to the skill level and experience of the users. In many archives there are a large variety of users who will be using it so more simple and commonly used features should be displayed clearly and more advanced features for experienced users can be made with less ease of use [1]. Users will also have different goals for using the system, which would motivate finding user requirements for the project and including them to ensure user satisfaction [1].

While much effort is needed to enhance usability, human centered design (HCI) should not be the only consideration. Care needs to be taken to not detract from functionality in favor of simplicity and ease of use. This design methodology is activity centered design (ACD) and is often used when functionality of the software is more important than the user experience [2]. ACD requires a longer learning curve, but it works better in complex systems [2]. This could be used for more advanced querying in the new Honours Archive, for experienced users to utilize more advanced functionality where functionality is more important than simplicity.

2.2 Simple DL

The Simple DL toolkit is a digital library software toolkit that is designed to work in low resource environments [3]. It is able to be run offline or on a local network and has shifted functions such as searching from the server to a JavaScript script run on the client. This reduces the load placed on the server and also reduces the minimum requirements to run this toolkit. Simple DL is also able to customize the outputted HTML files that display the archive's contents, typically through the use of CSS. Simple DL also uses XML and Excel to store metadata for the objects instead of using a database.

2.3 Bleek and Lloyd Collection

An example of a digital library that was developed without a database is the Bleek and Lloyd Collection. This collection is comprised of a large collection of "paper-based artefacts" that record the language and culture of the |xam speakers [4]. The software system that this collection was built on was designed according to the following principles: that software system mediation was to be avoided, that a network connection may be unstable, and that it would be preferable to create static representations through the pre-processing of data. Suleman [2007] demonstrated how XML, XSLT and XHTML could be used to develop a useful static, and portable digital library. XML-centric solutions are advantageous for libraries that are expected to be used for long-term data preservation purposes [4]. Linked XHTML pages were used from the XML source data and were stylized using XSLT stylesheets [4]. These hyperlinked pages utilized an Ajax-based search system that allowed the collection to be browsed and items accessed completely within the browser. Thus, making this approach useful for projects that do not necessitate a server-side search engine.

3. Problem Statement

Currently, the software used for the honours project archive is very outdated and there is a large potential to use new and more modern tools and techniques. There is a lack of functionality, the only function was to add a new archive and view existing archives. This caused issues such as users were unable to modify submissions.

Therefore, we are proposing to improve the digital honours archive by developing a web application with advanced features and a better user interface that provides a more pleasant user experience. The new honours archive will include:

- New and modern features

- Using new technologies
- Meeting users' needs that did not exist 20 years ago
- Provide a better user interface

3.1 Aims

The new honours project archive tool aims to improve the current honours archive website by focusing on developing the two specific areas below:

- **Presentation component** - focus on creating a web application to present from the honours archive. Taking the object from the archive and put it into a user interface so that the users can visualise and interact with it, developed by Simangaliso Mncwango
- **Submission component** - focus on the uploading of the projects and the overall submission workflow on the site. While also adding in new features, developed by Richard Paterson

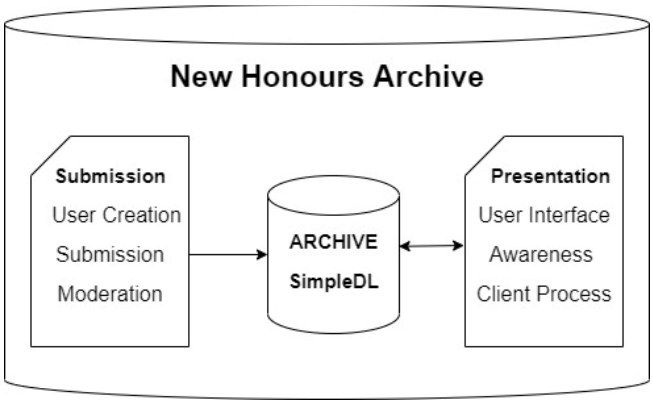

Figure 1

Figure 1: shows the two components of the New Honours Archive

3.2 Software Development Details

The objective of this project is to improve the current honours project archive in the department of computer science at the University of Cape Town. This new honours projects archive website will be developed with advanced tools and provide features that will amplify its usefulness and provide an interactive user interface. The new honours archive will offer an easier and clear transaction flow within web pages without the use of many pages to reduce page load time. Furthermore, the digital archiving tool software will provide the storage of projects and allow easy access of honours projects. As described from the previous section, the project is divided into two parts: Presentation and Submission.

The presentation side of the project includes the following features:

- User interface
- View previous projects
- Download projects
- Categorization of projects
- Searching of projects
- Server and client processing on the repository
- Interoperability of the repository

The Submission side of the project includes the following features:

- Create user
- Modify user (change password/email address)
- Forgot Password
- Login User
- Delete User
- Grouping of users in a project
- Submit an Archive
- Add metadata
- Show a provisional view of archive
- Moderation of archives
- Delete archive
- Modify archive

The users of the new honours archive are:

- *Honours Students* – all students who are doing honours under the Department of Computer Science
- *Computer Science Staff* – lecturers, supervisors, course convener and other computer science staff under the Department of Computer Science
- *Other interested parties* – anyone who is interested in viewing the honours projects. This can be students from other faculties and levels, and any individual who is keen to know more about honours projects.

4. Procedures and Methods

4.1 Development Plan

For the overall development of the project an iterative design process, Agile development, will be used. This will allow us to focus on key software elements and reduce the risk of not providing a complete project at the end of the duration.

Before beginning development of the project, user requirements will be gathered. From these requirements, a Use Case diagram will be created to ensure all of the required functionality is captured.

We will then create development artefacts such as UML diagrams and Class diagrams. This will ensure that the system design is better able to be communicated for any future developers who are likely to use this software if it gets adapted for the UCT honours archive.

Along with the design artefacts, a Lo-Fi prototype will be generated. This will show the basic layout of the HTML webpages and the links between pages. This prototype will help with the development of the actual webpage and, as it can be modified easily because it is a Lo-Fi prototype; many iterations of the design can be created.

4.2 Presentation

This section will present the processes for developing the Presentation component of the digital archive tool, which is the front-end design. I will explain explicitly the design features, method used and tools that are used to develop the overall presentation side. We will also discuss some of the challenges that can be encountered during the development process.

4.2.1 Implementation Strategy

The important feature of the presentation side of the archiving tool will be the ability to access, view and download honours projects

from their categorised groups. Projects will be categorised in terms of Human Computer Interaction, Artificial intelligence, Computational and year uploaded for better accessibility of the projects and layout. Another feature will be the ability to provide a pleasant user interface that will provide an easier user experience.

Projects will be ranked in terms of relevancy. This will improve the searching function on the repository. As the repository scales, we will need a good searching function or capability to ensure that users are able to get results faster in the relevancy of their queries.

4.2.2 Website Development

The digital archive tool will be developed using HTML, CSS, JavaScript and CGI scripts. These web programming languages were chosen based on personal experience. To provide a responsive web archiving tool and for styling purposes, we decided to use a front-end tool kit, Bootstrap Framework. This framework will accommodate the digital archive tool to be accessed in various kinds of gadgets, computers, tablets, and cell phones. The simpleDL toolkit will be used for data storage and for improved features of the archiving tool [3].

4.2.3 Expected Challenges

In spite of my prior experience in website development, I should expect challenges. Some of the challenges include:

- Inexperience in developing digital archive tool – this is my first time to develop a digital archive tool with an advanced tool kit, therefore I will spend some time to understand the tool kit.
- Working with CGI scripts - I do not have prior experience in working with CGI scripts. To understand it, I will learn it through watching video tutorials on YouTube and asking help from my colleagues.

4.2.4 Evaluation

The new honours web application needs to be evaluated to rate its functional performance and to find whether it met users' requirement or not. Participants will be recruited from the honours students and lecturers who are the key users of the system. We aim to recruit a total of 30 participants to test our software web application. The 30 participants will include both lecturers and students. All Important codes of conduct for research project such as the signing of the consent form will be observed.

The final web application will be tested and evaluated based on the following:

- Comparison of the current web application with the new honours web application presentation
- Usability – How usable the software it is, or does it make sense to the users?

The participants will be given two links; the first link will be directing the user to the current honours software and the second link will be directing the user to the new honours web application. Participants will be given tasks with instructions to perform on both software. After the completion of the tasks, participants will be asked to answer a list of questions based on their experience with the two software. Their feedback will be used for qualitative analysis in the later stage.

The Card Sort usability evaluation approach will be adopted for evaluating the usability of the new honours web application. This approach was chosen based on the recommendations we got from the dry runs presentation. The card sort method will help us to design and evaluate not only the information architecture of the web application, but also include features that make sense to the users. More detailed information and techniques chosen pertaining the card sort approach will be finalized close the software evaluation phase.

4.3 Submission

4.3.1 Implementation Strategy

4.3.1.1 User Management

For every honours project, a group will be created. In this group all of the members of a group will be added. These users will all have their UCT email address as the unique identifier and a password. These login details will be stored on the server.

To create a user, an administrator will create each honours project group and specify the users in the group. An email will then be sent to each of the users with a default password generated; emails will be sent using a python script linked with an email address. In this email, a link will be sent to each user so that they can input their details such as their name and will allow them to choose a new password.

When a user attempts to login, an HTML form will be used to collect the data and a python Common Gateway Interface (CGI) script will be run on the server. This script will verify the login and create a cookie saving the valid login result. Once a user is logged in, they will have the ability to add to or modify files in their honours archive. They will also have the ability to change their details that they have, such as their saved name in the event of a mistake being made, change their password or deleting their account from the system along with any information stored. All modifications will be handled using CGI scripts.

4.3.1.2 Submission of Archives

A HTML page will be available to handle modification of the archive. By choosing an option made available, using links, such as uploading new files, changing meta-data, deleting old files, etc, the user will be redirected to a new HTML page to handle the specific task.

Each HTML page will handle different functionality; however, any modifications will run a python CGI script to modify the data on the server.

Once all changes are made, the changes will be placed in a temporary location containing the old files and all of the modifications done. A display of the new expected archive will be rendered so that a user can ensure that it is functioning as expected. Once they are happy, they can confirm their modification and the files from the temporary location will be moved to a pending file location.

In the pending files folder, all of the archives are stored waiting for an administrator to approve. This ensures that no content will be shown on the website until approval has been granted. When approval has been granted, a script will run to interact with the Simple DL toolkit to append the file to the archive and the file will be stored there and removed from the pending list.

4.3.1.3 Website Design

This website will be created using HTML, CSS and JavaScript. For the design of the website, a CSS framework (Bootstrap) will be used. This will ensure there is consistency across the website as much of the CSS will be using the Bootstrap framework and also

that the website is responsive and able to change depending on the screen size.
Bootstrap is designed to be mobile first so the New Honours Archive will also be able to be viewed on mobile devices, tablets, laptops or desktop. This will improve accessibility to the website, especially for the general populace.
A prototype of the website will also be created; however, this will only be done after the user requirements have been gathered as the users will also be able to give input on how they would like the design to look.

4.3.2 Expected Challenges

Some expected challenges will involve using the python CGI scripts; this is because I have had no exposure to this technology before this project. I will watch guides and tutorials on how to use them as well as reading up on the documentation.
Another challenge will be the handling of user privileges. Some users need to have admin rights whereas some users will only be allowed to upload to a specific project. Privileges may also need to be removed from users once the project deadline has been reached.

4.3.3 Evaluation

To evaluate the success of the project, testing will be done to ensure the correct behaviour of the submissions and user handling. This can be done through manual integrated testing of the system as a whole. Once the system is fully operational and ethical clearance has been given, we can start with user testing. Users will be given a set of tasks to complete and relevant accounts with privileges. Users will then attempt to use the system and will score the system on different criteria out of 10. The users will also be asked to provide feedback comments for qualitative evaluation.
Once the scores have been gathered, a statistical process will be used to analyse how well the system performed. Closer to the time, more research will be done into the best standards for user testing along with standard statistical analytics tools to measure and comment on the results.
Other testing will be done in the form of stress testing with large files being uploaded and multiple uploads simultaneously to ensure that the system can handle more extreme use.

5. Ethical, professional and legal issues

This project involves collaboration with different stakeholders, and it is necessary to comply with the ethical code of conduct for academic research work to ensure that ethical procedures are not violated. We anticipate applying for the ethical clearance to the University of Cape Town before the start of the project. With the possession of the ethical clearance approval, this will permit us to do user requirements and help during the user testing period for the prototype. We anticipate working with the UCT honours students and computer science staff for user requirement gathering and testing.

6. Anticipated outcomes

6.1 System

The expected outcome for this honours archive tool is to have an improved version of the current honours archive site. Advanced and improved features will be found in this tool to enhance and maximize its functionality for users. These features include a pleasant and a clear user interface that will allow better navigation throughout the system. Honours students as users will be able to create their accounts before uploading and submitting projects with metadata. Easy retrieval of projects through searching will be implemented. The archiving site will be responsive to run on cell phones, tablets, and computers.

7. Project Plan

7.1 Risks

Risk Matrix can be found in appendix 1.

7.2 Timeline

Gantt Chart can be found in appendix 2.

7.3 Resources Required

In order to achieve the final outcome of the project, we will need computers for system development, and we ought to utilize our personal computers. Furthermore, our development of the archiving tool will be supplemented using the SimpleDL toolkit for advanced system features, Python CGI for scripts and Bootstrap Framework for styling for and the overall attractiveness of the website archiving tool.

7.4 Deliverables

The precise deliverables of the project are as follows:

- Literature review
- Project proposal
- Ethical clearance approval
- Project proposal
- Software feasibility demonstration
- Two academic papers
- System source code and documentation
- Project poster and website

7.5 Work Allocation

The presentation site will be created by Simangaliso, and the submission site will be handled by Richard. Both of us will collaborate and will be involved with designing the look and feel for the website. Each of us will complete our own reports, with user testing; the final poster and website design will be done collaboratively with smaller tasks being distributed closer to the time at which they are done.
Each of us has clear non-interdependent tasks for the creation of the new Honours Archive with the exception of the design and homepage.

8. REFERENCES

[1] Blandford, A. et al. 2004. Usability of Digital Libraries (Editorial). *International Journal on Digital Libraries*. (2004)

[2] Norman, D.A. 2005. Human-Centered Design Considered Harmful. *Interactions*. 12, 4 (Jul. 2005), 14–19. DOI:https://doi.org/10.1145/1070960.1070976.

[3] Suleman, H. 2021. Simple DL: A Toolkit to Create Simple Digital Libraries. *Towards Open and Trustworthy Digital Societies* (Cham, 2021), 325–333.

[4] Suleman, H and Phiri, L. 2012. In Search of Simplicity: Redesigning the Digital Bleek and Lloyd. *DESIDOC Journal of Library & Information Technology*. 32, (May 2012), 306–312. DOI:https://doi.org/10.14429/djlit.32.4.2524.

[5] Suleman, H. 2007. An African Perspective on Digital Preservation. *Multimedia Information Extraction and Digital Heritage Preservation*. 295–306.

9. Appendix

9.1 Appendix 1

Risk	Probability	Impact	Mitigation	Monitoring	Management
One of us may drop out of the project	1	8	Try to keep abreast with work to ensure we are coping	Regular meeting to check up on each other	Project parts are largely decoupled so if one of us drops out the other can create a finished project
One of us may fall behind in the project	3	5	Try to keep abreast with work to ensure we are coping	Regular meeting to check up on each other	Ensure there are minimal dependencies on each other's work
Failure of one of our computers/ laptops with the project on	2	4	Replacing faulty components and keeping software up to date	Regularly running health checks on computers	Use Github to store work online and push regularly
Lack of participants for user testing	5	6	Try to encourage users to test the system through small pay-outs	Keep track of how many users tested the software	Ensure a minimum number of users are able to test the software

9.2 Appendix 2