Project Proposal - ADVICE: Virtual Student Advisor

AVHUSAHO RAMALALA, The University of Cape Town, South Africa
MICHAEL BROUGH, The University of Cape Town, South Africa
TINOTENDA MUZAMBI, The University of Cape Town, South Africa

CCS Concepts: • **Computer systems organization → Neural networks**; • **Software and its engineering → Cloud computing**; **Software design techniques**; • **Theory of computation → Logic and databases**; • **Human-centered computing → HCI design and evaluation methods**.

Additional Key Words and Phrases: datasets, neural networks, text tagging, software design, software engineering

1 PROJECT DESCRIPTION

Having a Student Advisor to help guide students through their degree is a big step on the road to success [18]. Advising students is one of the most important responsibilities a university must fulfill. The process of completing this task efficiently and effectively becomes extremely complicated when there are thousands of students involved. The University of Cape Town (UCT) is faced with the challenge of enrolling over 25000 students annually, all of whom will have to come into contact with a Student Advisor at some point for various reasons.

The University houses students from many different backgrounds, and they all have to be catered for. Sorting out queries for all these students, especially over email, is inefficient and leads to frustrated students and student advisors. This may have detrimental consequences on a student's academic career [22]. Some universities have attempted to solve this challenge by creating an automated Student Advising system that students can use as an alternative to Student Advisors for simple queries [27; 29; 32].

The aim of this project is to create a website that acts a virtual student advisor, with the aid of a chatbot, for students. A chatbot is a software tool that is capable of having a human-like conversation in natural language [6; 11; 28]. The system should cater for students from low-income areas and students who have limited access to the university campus. The website will include a chatbot that students can communicate with in real time at any time of the day. The goal is to offer pre-admission guidance on courses and transferring credits, explaining policies and procedures such as add/drop protocols, establishing prerequisites, and helping students figure out the best fit for their needs. Confidentiality and privacy are vital aspects of any credible software design and should also be accounted for during this project.

2 PROBLEM STATEMENT/AIM

2.1 Problem

Student Advisors attend to various student queries ranging from simple queries such as how long a certain degree takes to complete and how many credits a student will need to graduate if they drop or choose a course. Presently, the main form of communication between students and Student Advisors is through email or in-person meetings during office hours (although, recently in-person meetings have been discouraged due to COVID-19). The current system is not efficient enough because there is a limited amount of Student Advisors who must help a large number of students. In addition, Student Advisors usually have many other commitments which may lead to them rushing through their student advising duties so they can return to other matters [22]. An easily accessible online system would be able to cater for students from low-income areas, first-time university students, rural students and many other students who have little to no access to universities and university resources.

Students are expected to read through several handbooks to understand the procedures, policies, course requirements and many other aspects of university registration. There is no centralised, easily accessible platform where students can use to find all this information. The process of going through all the handbooks and visiting several disparate websites to find the necessary information is cumbersome. Understanding the structure of each faculty's handbook and language is an extra challenge, thus making the process more complex. First-time students could also find navigating the websites difficult, studies by Margolin et al [24] and Majid et al [23; 26] have found that several university websites lack in terms of usability and accessibility. Having a system that could search the handbooks for students and simplify the information would make the process significantly easier.

2.2 Issues/difficulties with the current work

2.2.1 Existing systems are not built for South African universities. The structure of the academic year and degree of South African universities may differ slightly in various aspects [16] compared to other universities from North America, Europe, or Asia where most of the existing systems have been implemented. Terminology used in South Africa is also different [7] so it will have to be customised for South African standards.

2.2.2 Language. Most of the existing systems have been implemented in English or the native language of the university in which they were designed in. South Africa has eleven official spoken languages [20] . The most common of these are English, Zulu and Xhosa. A system that caters for students coming from backgrounds that speak these languages would help students transition between high school and university more smoothly. However, implementing a multilingual system is beyond the scope of this project's first iteration. It is a vital consideration for future work on this project or on similar systems like this one in the future.

2.2.3 Information is spread over several UCT websites. The current issue is not that information is not available online, it is that information is not centrally available online. There is no single website

Authors' addresses: Avhusaho Ramalala, The University of Cape Town, Cape Town, South Africa; Michael Brough, The University of Cape Town, Cape Town, South Africa; Tinotenda Muzambi, The University of Cape Town, Cape Town, South Africa.

or platform where students can visit and get all the information they need without having to visit another website or read a handbook. UCT has a vast amount of information displayed on many sites such that finding something may be stressful and feels unintuitive. This is mostly a concern for prospective students who are not familiar with UCT websites or student handbooks.

2.2.4 Confidentiality and Data and privacy. It is evident from the reviewed work of existing implementations of a virtual student advisor that security was not a vital part of design. The systems give none or very little information about security and how privacy will be ensured. Our system should improve upon this and ensure security is a vital aspect of the system.

2.3 Main Requirements

2.3.1 Answer students queries about UCT procedures and processes. A large part of getting registered into university is the administration aspect of it. Navigating this step can be a daunting task especially for first time students. The website will offer information on how to fill in forms and what documents are required for particular processes. The system will offer detailed and simplified explanations of various processes and tasks students can complete.

2.3.2 Help students calculate credits. Calculating how many credits are required to qualify for a degree can become complicated and time-consuming. The task becomes more complex if a student's majors span across multiple faculties as the student must browse multiple handbooks to figure out how many credits each course is worth. Our platform will provide a simple tool that students can use to calculate credits based on their courses.

2.3.3 Assist students with understanding UCT terminology. Prospective students may not be familiar with UCT terminology such as course codes and task names (i.e., "registration" vs "enrollment"). Students will be provided with a summary of commonly used terms at UCT, this makes it easier for students to understand documents and requirements.

2.3.4 Offer summary of handbooks. The handbook reader will make it possible to have a database of information based on the handbooks. It will no longer be necessary to browse through all the handbooks in search of information, students can simply search for information through the platform.

2.3.5 Give general information about degrees and possible career paths for such degrees. Choosing the right degree for oneself can be a challenging task, especially if one does not know what they can do with a degree. The platform will provide a summary of degrees offered at UCT and the different careers one could pursue based on the degree. This will help students better understand what they want or prefer.

2.3.6 Real-time conversation chatbot. The platform will include a chatbot capable of simple real time conversation with students. The bot will be available to students 24 hours daily. Any problems that are too complex for the chatbot can be passed on to a recommended Student advisor. The chatbot will make use of the database created by the handbook reader to answer questions.

2.3.7 Security-Confidentiality and Data privacy. The system will store students' personal and confidential information so that it is personalized for each student therefore sufficient security will have to be implemented to ensure that students' data is protected. Students need to feel safe and comfortable when using any aspect of the platform . A relevant privacy policy must be provided [15] to ensure transparency so that users are aware of how data will be processed and stored. Implementing security should be part of the design process rather than an afterthought. Information should be shared over a secure connection to mitigate the risk of malicious software intercepting the messages. HTTPS connections have been found to be secure and offer a great deal of security for online connection [21]. The platform will have to abide by the regulations stated in the privacy Act [10], POPI Act, that was recently passed in South Africa.

The System should be easily accessible, have the ability to help students with course decisions and filling out actual paperwork. The platform must be data efficient because internet access has not yet been standardised in South Africa [19] . Overall, the system should be able to carry out most tasks that human student advisors usually complete. The objective of the system is not to replace human student advisors but rather offer an aid which would leave more time for them to deal with the more complex queries which require more time spent with students.

3 PROCEDURES AND METHODS

3.1 Website

For the implementation of the website, we will be using popular cutting-edge technologies to both take advantage of an established community of developers as well as make use of the latest technologies that will benefit our platform. One technology that embodies this better than most others is React.js and that will be the technology of choice on the front-end (the part of the system responsible for the visuals and what the user interacts with). We chose to go with a JavaScript framework because it allows us to harness the power of the libraries [13] that can be used along with them as well allowing for reusing components and the ease of embedding logic in the layout of the website. We will specifically be using Next.js which is a React framework which provides all the same benefits as React with additional benefits such as built in TypeScript and Sass support as well as built-in support for routing and client-side/server-side rendering [30].

We will in fact be taking advantage of the built-in support for TypeScript and have a site-wide implementation using TypeScript. TypeScript is a superset of JavaScript that introduces static typing [14]. We will also be using Sass over regular CSS for many of the same reasons.

The back-end will be implemented using Node.js to once again take advantage of the well-established community and the libraries available. Some tasks that the back-end will handle will include handling the storing of the data, handling authentication and any other background services that will be run on the platform. For the storing of data, we will use MongoDB. We opted for a NoSQL database because we anticipate that a lot of the data we will have to deal with will be unstructured and presented in JSON format. NoSQL

databases handle such data better than their SQL counterparts [25]. The back-end will be hosted within Next.js' built-in API route system as this allows for much easier deployment on Vercel.

3.2 Chatbot

Previous implementations have mostly been catered to their respective universities, as such, this project will have to customise its application for students at UCT. This will come with its own set of challenges such us the language of choice as UCT is diverse. The chatbot could be implemented to process English initially and then expanded to cater for vernacular languages and various other international languages in the future.

In terms of technologies the chatbot will be implemented with Python and JSON, this includes a chat and voice interface which will be hosted by Google. The challenging part is extracting information and semantics from the input string, matching it with a response from the database (MongoDB) and sending the response back to the user. The Python libraries that we plan to use include NLTK, Pytorch and Gensim. These will aid us in the Natural Language Processing aspect of the chatbot. NLTK will be used for text processing, Gensim for word tagging and PyTorch for semantic analysis. Students responses and previous messages will also be stored in the database.

The chatbot will also expose an API endpoint that will be implemented using either Django [3] or Flask [4]. This endpoint will be how the website interacts with the chatbot. The front-end will communicate with this back-end to pass it input from the user as well as to receive an appropriate response.

3.3 HCI & PDF Reader

3.3.1 HCI. The HCI aspect of this project involves the investigation into how humans perceive natural language processing bots and which design principles lead to the best user experience possible. The bot should try to mimic interpersonal communication as closely as possible leading to a positive user experience. A contextual inquiry into which humanising techniques are most effective for our target audience should be conducted to discover what should be considered when designing high-fidelity prototypes and the final Virtual Student Advisor tool.

3.3.2 PDF Reader. Yildiz, Kaiser and Miksch [33] formulated a method to extract information from tables in PDF documents using table decomposition and other methods on over 150 tables across several pdf documents. Strategies and methods from this paper can be adapted for implementation in the UCT Student Handbook Reader.

In order to successfully convert data extracted from the student handbooks to tables in a database, NoSQL is an adequate tool for the task and we will specifically use MongoDB. To use NoSQL effectively in a Java program, one must implement it using the NosDB library [5] which will also format the data in a JSON format such that it can be used by the web server hosting the Virtual Student Advisor.

The PDF Reader will only have to communicate with the database so no integration with the website will be necessary. Uploaded JSON data from the PDFs will be posted straight to the database. Thereafter, the website will simply pull the data directly from the database.

4 ETHICAL, PROFESSIONAL AND LEGAL ISSUES

This project does have ethical and legal implications as there will be aspects of the project that require human interaction. This will take the form of interviews with UCT students and staff prior to the development of the platform as well as some user testing once a prototype has been developed.

For this reason, our project will need ethics clearance and the document has been prepared and submitted. We will, as researchers, endeavour to keep the best interests of the participants at heart and ensure that they are well informed about their rights going into any studies/interviews with us. We aim to obtain trust by anonymising the data we collect and letting participants know that no identifying information will be collected.

In terms of legal issues, this research could potentially negatively impact UCT's public image from the perspective of participants. If a participant has a bad experience with a researcher, they could extend that feeling towards UCT as an institution. Participants could represent the University negatively on public platforms like social media websites. However, we will take utmost care when representing UCT during interviews with participants as to ensure no bad relations are formed and to uphold UCT's public image to the best of our abilities.

Security and privacy is being taken very seriously and with high regard on our platform. Our platform will be in line with all the modern security and privacy regulations. These include the POPI Act [1] for South Africa and the GDPR [2] for the European Union to name a few. User data will be anonymised and kept confidential as far as possible.

5 RELATED WORK

5.1 Website

Currently, a few systems have been created at various universities that are meant to assist students with frequently asked questions and offering student advice in general[27; 29; 32]. However, each of these implementations has been specifically designed for the respective academic institution. There is not an existing set of standards or template on how to design a virtual student advising system. Additionally, none of the existing implementations are in South Africa or Africa as a whole. Current websites do not cater for South African universities. The website will have to be implemented with South African standards and languages in mind.

5.2 Chatbot

There is a significant amount of work related to implementing a chatbot into an online system. There is not much work in implementing a chatbot in the context of a student advising system. Similar chatbot applications include the more personalized approach by Carolis et al [12] . Their implementation attempted to mimic human conversation together with the use of an animated character with human resemblance. Herry D Wijaya et al. [32] and Bhavika R. et al. [27] have also created similar applications meant to assist students in their respective academic institutions. The chatbots were meant to answer Frequently Asked Questions and offer information management for a college, respectively. Popular chatbot applications,

bereft of a virtual student advisor system, are the Artificial Linguistic Internal Computer Entity (ALICE) by Dr Richard Wallace [9] and Elizabeth (adapted from ELIZA, created by Joseph Weizenbaum in 1966 [31]). Prominent voice chatbots include Google's Google Assistant, Amazon's Alexa, Apple's Siri, and Microsoft's Cortana. These are advanced implementations which have been existed for several years now, it may be possible to integrate one of these implementations, such as Google Assistant, into the system for voice communication.

5.3 HCI and PDF Reader

Examined literature on the appearance of chatbots or websites in general show that there are specific design principles that affect users' perception of a system. Literature by Go and Sundar [17] gives insight on the importance of "humanness" in chatbots and why mimicking human conversation is vital in conversational agents There are existing libraries that can help with extracting information from PDF documents and creating a database. Apache PDFBox for java [8] is one such library. Yildiz, Kaiser and Miksch [33] have provided work on strategies and methods that can be used to extract information from tables in pdf documents like those in UCT handbooks.

6 PROJECT OUTCOMES

This section details the outcomes that we expect for the end of the project, this includes what impact we expect this project to have on the UCT registration environment and the key success factors of this project.

6.1 Expected Impact

After the successful implementation of the Virtual Student Advisor tool, we expect that numerous UCT students from the Humanities, Science, and Commerce faculties will be able to use it to their advantage. We expect that this tool will ultimately decrease the number of requests for consultations with UCT student advisors during the registration process. We also expect that this tool will be a valuable resource for student advisors to use for reference when providing registration recommendations to students. Overall, the tool will improve the UCT registration process by streamlining the student advisor process and reducing the need for the scheduling of in-person consultations with student advisors.

6.2 Expected Results

We expect that the virtual student advisor tool will be ready for use on UCT's web domain. It will have all of the basic functionality aforementioned in this document. The tool will be able to accept user input in the form of text and potentially voice and relay information in a manner that mimics human interaction. It will be able to use Natural Language Processing to appropriately service user's requests and will be user-friendly with a focus on user-experience design such that all current and prospective UCT students will be able to use the tool without assistance. The tool We expect this project to become a valuable tool in the UCT registration process which students can use to gain information about courses and degrees to better inform their choices when registering at UCT.

6.3 Key Success Factors

We will evaluate the success of this project based on the criteria outlined below:

- Successfully building and implementing front and back-end of the website that hosts the chat bot on UCT's web domain.
- Conducting a contextual inquiry into the most effective design principles of chat bots by interviewing multiple students and student advisors.
- Designing, building, and implementing the chat bot that efficiently responds to user input and queries.
- Ensuring that the implemented system is built with security principles and practices. User's personal information within the system is kept confidential at all times.
- Successfully designing and building the PDF reader for UCT Student Handbooks.

7 PROJECT PLAN

7.1 Risk Anticipation

All foreseeable risks and their mitigation strategies are illustrated in the risk assessment matrix found in Appendix A.

7.2 Timeline and Milestones

From the 24th of June 2021 (submission of project proposal), the project will take approximately 17 weeks to complete. The exact breakdown of tasks and major milestones is illustrated in the Gantt chart found in Appendix B.

7.3 Deliverables

Below is a list of all deliverables for this project:

- **09/07/2021** - Proposal Presentation
- **30/07/2021** - Revised Finalised Proposal
- **10/08/2021** - Initial Software Feasibility Demo
- **06/09/2021** - Draft of Final Paper
- **17/09/2021** - Final Project Paper
- **20/09/2021** - Final Project Code
- **04/10/2021** - Final Project Demonstration
- **11/10/2021** - Project Poster
- **18/10/2021** - Web Page

8 TEAM MEMBER WORK SPREAD

Tinotenda Muzambi will be responsible for the development and implementation of the website. Avhusaho Ramalala will be responsible for the implementation of the chatbot. Michael Brough will be responsible for the Human Computer Interaction component of the project as well as implementing the PDF reader.

REFERENCES

[1] Protection of personal information act (popi act), 2013.
[2] General data protection regulation, 2019.
[3] Django: The web framework for perfectionists with deadlines, 2021.
[4] Flask, web development one drop at a time, 2021.
[5] Introduction to java api in nosql database - nosdb, 2021.
[6] AbuShawar, B., and Atwell, E. ALICE Chatbot: Trials and Outputs. *ComputaciÃy Sistemas 19* (12 2015), 625 – 632.
[7] Alberts, M. Terminology in south africa. *Lexikos 9* (1999), 18–35.
[8] Apache. Apache pdfbox | a java pdf library., 2021.

[9] Bani, B. S., and Singh, A. P. College enquiry chatbot using alice. *International Journal of New Technology and Research 3*, 1 (2017).

[10] Botha, J., Grobler, M., Hahn, J., and Eloff, M. A high-level comparison between the south african protection of personal information act and international data protection laws. In *ICMLG2017 5th International Conference on Management Leadership and Governance* (2017), p. 57.

[11] Chung, M., Ko, E., Joung, H., and Kim, S. J. Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research 117* (2020), 587–595.

[12] De Carolis, B., Pizzutilo, S., Cozzolongo, G., Drozda, P., and Muci, F. Supporting students with a personal advisor. *Journal of Educational Technology & Society 9*, 4 (2006), 27–41.

[13] Elrom, E. *React and Libraries*. Apress Media LLC, 2021.

[14] Fischer, L., and Hanenberg, S. An empirical investigation of the effects of type systems and code completion on api usability using typescript and javascript in ms visual studio. *ACM SIGPLAN Notices 51* (10 2015), 154–167.

[15] Følstad, A., Nordheim, C. B., and Bjørkli, C. A. What makes users trust a chatbot for customer service? an exploratory interview study. In *Internet Science* (Cham, 2018), S. S. Bodrunova, Ed., Springer International Publishing, pp. 194–208.

[16] Fourie, M. Institutional transformation at south african universities: Implications for academic staff. *Higher Education 38* (2016), 275–290.

[17] Go, E., and Sundar, S. S. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior 97* (2019), 304–316.

[18] Gurantz, O., Pender, M., Mabel, Z., Larson, C., and Bettinger, E. Virtual advising for high-achieving high school students. *Economics of Education Review 75* (2020), 101974.

[19] Juma, C., and Moyer, E. Broadband internet for africa. *Science 320*, 5881 (2008), 1261–1261.

[20] Kamwangamalu, N. M. A new language policy, old language practices: status planning for african languages in a multilingual south africa. *South African Journal of African Languages 20*, 1 (2000), 50–60.

[21] Krombholz, K., Busse, K., Pfeffer, K., Smith, M., and von Zezschwitz, E. "if https were secure, i wouldn't need 2fa" - end user and administrator mental models of https. In *2019 IEEE Symposium on Security and Privacy (SP)* (2019), pp. 246–263.

[22] Krupp, M. The student-advisor disconnect. *Women in Higher Education 23*, 11 (2014), 18–18.

[23] Majid, I., and Yanduri, V. Analysis of university websites - a study. 11–15.

[24] Margolin, J., Miller, S. R., and Rosenbaum, J. E. The community college website as virtual advisor: A usability study. *Community College Review 41*, 1 (2013), 44–62.

[25] Nasution, M. Nosql concepts and characteristics.

[26] Peker, S., Kucukozer-Cavdar, S., and Cagiltay, K. Exploring the relationship between web presence and web usability for universities: A case study from turkey. *Emerald Group Publishing Limited 50*, 2 (2016), 157–174.

[27] Ranoliya, B. R., Raghuwanshi, N., and Singh, S. Chatbot for university related faqs. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2017), pp. 1525–1530.

[28] Shawar, B., and Atwell, E. *A comparison between Alice and Elizabeth chatbot systems*. University of Leeds, School of Computing research report 2002.19, December 2002.

[29] Suvethan, N., Huzaim, K. A. M., Mathusagar, R., Gamage, M., and Imbulpitiya, A. Virtual student advisor using nlp and automatic appointment scheduler and feedback analyser.

[30] Thakkar, M. *Building React Apps with Server-Side Rendering*. Apress Media LLC, 2020.

[31] Weizenbaum, J. Eliza — a computer program for the study of natural language communication between man and machine. *Commun. ACM 26*, 1 (Jan. 1983), 23–28.

[32] Wijaya, H. D., Gunawan, W., Avrizal, R., and Arif, S. M. Designing chatbot for college information management. *IJISCS (International Journal of Information System and Computer Science) 4*, 1 (2020), 8–13.

[33] Yildiz, B., Kaiser, K., and Miksch, S. pdf2table: A method to extract table information from pdf files. In *IICAI* (2005), pp. 1773–1785.

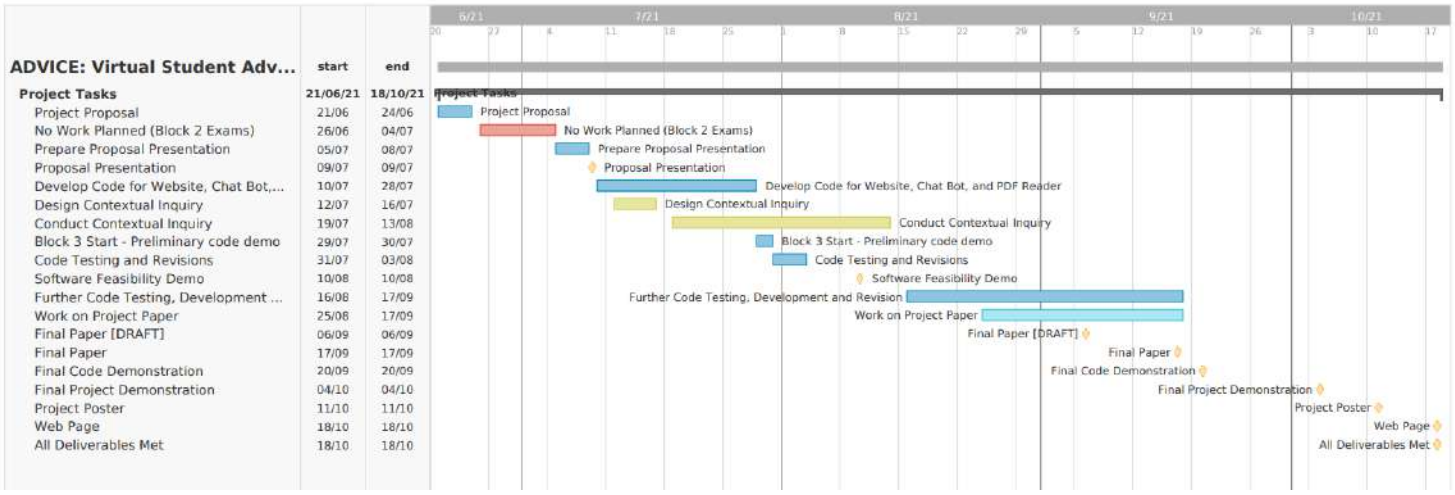
APPENDIX A

Risk Assessment Matrix

Risk	Probability	Impact	Mitigation	Management
Project members become sick, preventing them from working on the project.	Medium	Medium	All project members will take health and safety precautions and adhere to COVID-19 guidelines and protocols.	Ensure that the project is keeping ahead of schedule so that any delays can be accounted for.
Extended waiting period for ethics clearance from UCT.	Medium	Medium	Start the project as early as possible and work on code sections while waiting for clearance.	Follow up on ethics clearance if it has not yet been acquired after 2 weeks. Complete as much code as possible before conducting contextual inquiry.
A team member drops out	Low	High	Group communications kept cordial and regular. Schedule the working times to ensure that members are not overwhelmed by workload	The work would be redistributed amongst the remaining group members.
Communication issues arise due to remote work environment.	Low	Low	Engage in group meetings regularly to keep up to date with project progress.	Ensure that the project is keeping ahead of schedule so that any delays can be accounted for.
Scope Creep	High	Low	Team members should be wary of adding "gold-plating", focus on fundamental features first.	Remove the least impactful features from the project. Potentially freeze features until fundamentals are finished.
Difficulty in getting UCT students and staff to participate in the contextual inquiry.	Medium	High	Start recruiting participants as early as possible. Ask UCT staff for assistance in contacting students and other staff.	Conduct the contextual inquiry with as many participants as possible and conduct the analysis of the results with a smaller dataset.

APPENDIX B

Gantt Chart



Gantt Chart displaying project timeline and milestones