ISS AMAB (ASK-ME-ANYTHING-BOT)

KHOO WEE BENG

Table of Contents

1 EXECUTIVE SUMMARY

A Chatbot (sometimes referred to as a chatterbot) is a computer program that attempts to simulate the conversation of a human being via text or voice interactions.

Commands or inputs are received from the user and the Chatbot will respond in a satisfactory way to the user which results in the completion of the action or conversation initiated by either of them.

This project aims to provide a viable Chatbot for NUS ISS website.

2 PROBLEM OVERVIEW

There are many topics in NUS ISS website and it might be an arduous task to look for information. The main navigation has the following, *Executive Education*, *Graduate Programmes*, *Stackable Programmes*, *Centres of Excellence*, *Community*, and *About Us*. After browsing the contents, the following is assumed:

Executive Education	ISS program, course and related information
Graduate Programmes	ISS program, course and related information
Stackable Programmes	ISS program, course and related information
Centres of Excellence	ISS program, course and related information
Community	Non program related information
About Us	Non program related information

Further analysis of *Executive Education* reveals numerous *Disciplines*, which includes a list of modules available for each *Discipline*. The modules for the *Disciplines* generally adhere to information, that were categorized in sections; *Overview*, *Key Takeaways*, *Who should attend*, *What will be covered*, *Fees and Funding*, *Instructors*, *Certification*, and *Preparing for your course*.

For *Graduate Programmes*, information was categorized in a slightly different fashion, namely; *Overview*, *Modules*, *Projects & Internships*, *Timetable & Exams*, *Fees & Loans*, *Admission & Application*, and *Career Pathways*. There were six *Graduate Programmes*, a small number, as compared to the hundreds of modules for *Disciplines*.

There are four *Stackable Programmes* links, however, they were un-alike in terms of information sections. Similarly, *Centres of Excellence* links had different categorization for their pages

3 CHATBOT APPLICATION

3.1 INTRODUCTION

There will be two parts to the implementation of a ChatBot. First, will be to gather information in a fast and usable manner such that it can be usable for ChatBot. Next, would be to explore the capabilities of DialogFlow by implementing the options available (within DialogFlow).

3.2 TECHNICAL SOLUTIONS

3.2.1 WEBSCRAPING

Python was used to extract information from NUS ISS website. There were no issues extracting the raw HTML from the website, however, crawling through the URLs starting from www.iss.nus.edu.sg returns URLs that was not relevant to the project. Hence, the URLs were "clicked-out" and collected in a list to let the Python program know which pages to crawl. *Executive Education* URLs was collected to run on the initial Python program.

Initially, BeautifulSoup Python library was used to strip HTML tags and to identify the sections in the pages. The process was not optimal; specific HTML tags, or names has to be identified in order to have the correct information extracted, additionally, the same codes could not be used on another URL.

Regular expressions were explored as an alternative to BeautifulSoup, which had a better returned result. Using the first URL for *Discipline* Module (https://www.iss.nus.edu.sg/executive-education/course/detail/machine-reasoning/artificial-intelligence), various sections was identified by detecting HTML comment tags or sub header tags, as **Begin** and **End** of the extraction. Applying the same codes to the next URL, the returned results was promising. After testing a few URLs, it was noticed that there were minor differences in the sub headers. The following are the results of testing and running from the collected Discipline URLs:

	Begin	**End**
Content	<!--start content-->	<!--end content-->
CourseTitle	<h1>	</h1>
Overview	Overview <!-- overview -->	<!-- classes --> <!-- menu \(links\) -->
Key Takeaway	Key Takeaways Key Takeaways	Who Should Attend Who Should Attend
Who Should Attend	Who Should Attend	Pre-requisites Prerequisites Pre-requisites Pre-requisites Prerequisites Prerequisites

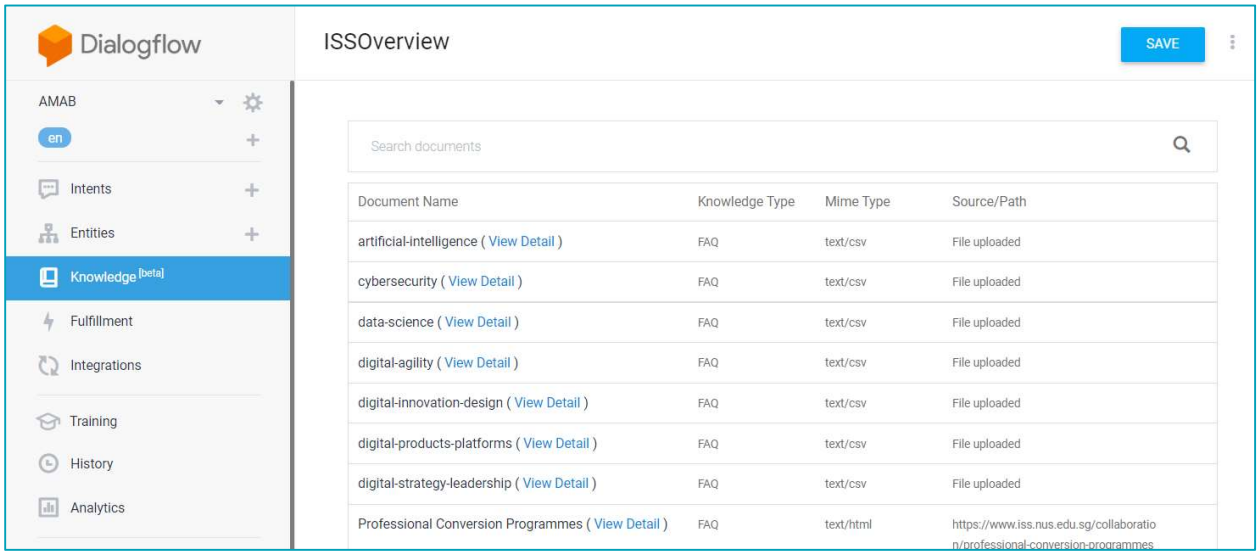
		Pre-requisites Pre-requisites Pre-requisites
Pre-requisites	Pre-requisites Prerequisites Pre-requisites Pre-requisites Prerequisites Prerequisites Pre-requisites Pre-requisites Pre-requisites	What Will Be Covered
What Will Be Covered	What Will Be Covered	Fees and Funding Fees & Funding Fees and Funding Fees & Funding
Fees and Funding	Fees and Funding Fees & Funding Fees and Funding Fees & Funding	Certification Certificate Certificate Exams & Certificate
Certification	Certification	Preparing for Your Course

Questions to the extracted information was subsequently added such that it would relate to the response(answer).

E.g. Can you give me an overview of "courseTitle" ?

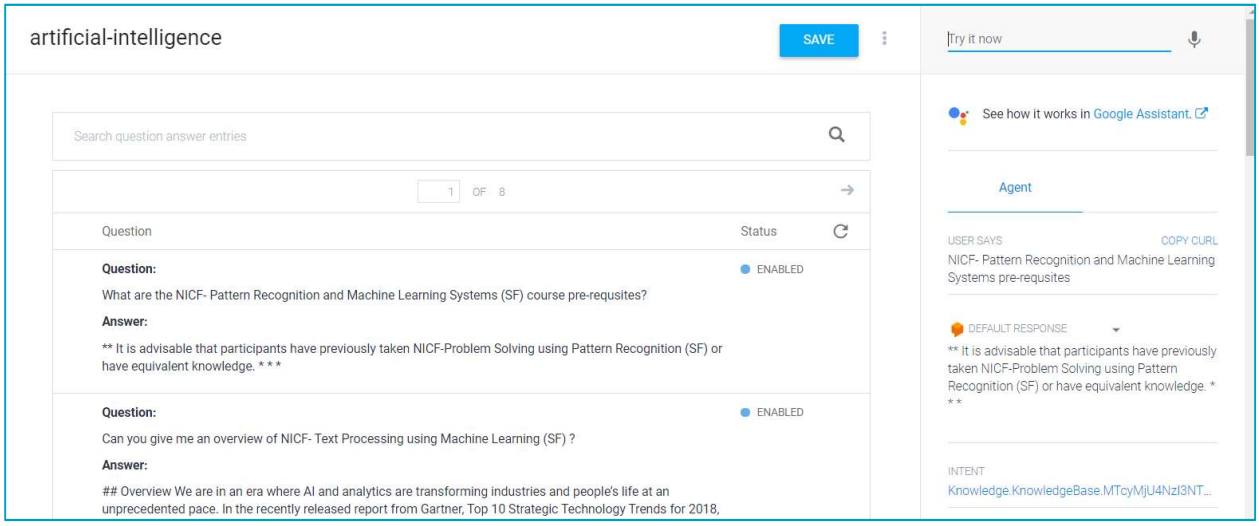
The results were also formatted as comma separated values.

Hence, the Knowledge Base was constructed and imported into DialogFlow for testing.

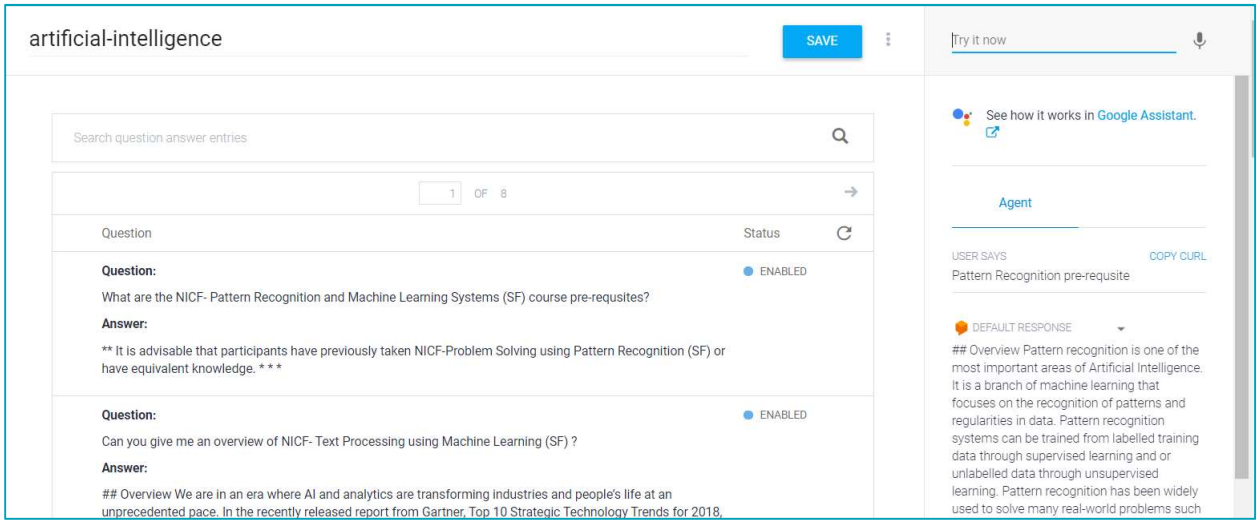


3.2.2 DIALOGFLOW – KNOWLEDGE BASE

The results from testing of knowledge base shows that the questions asked has to be rather explicit and close to the actual question.



Rephrasing questions in a simplified form may not return the correct response.



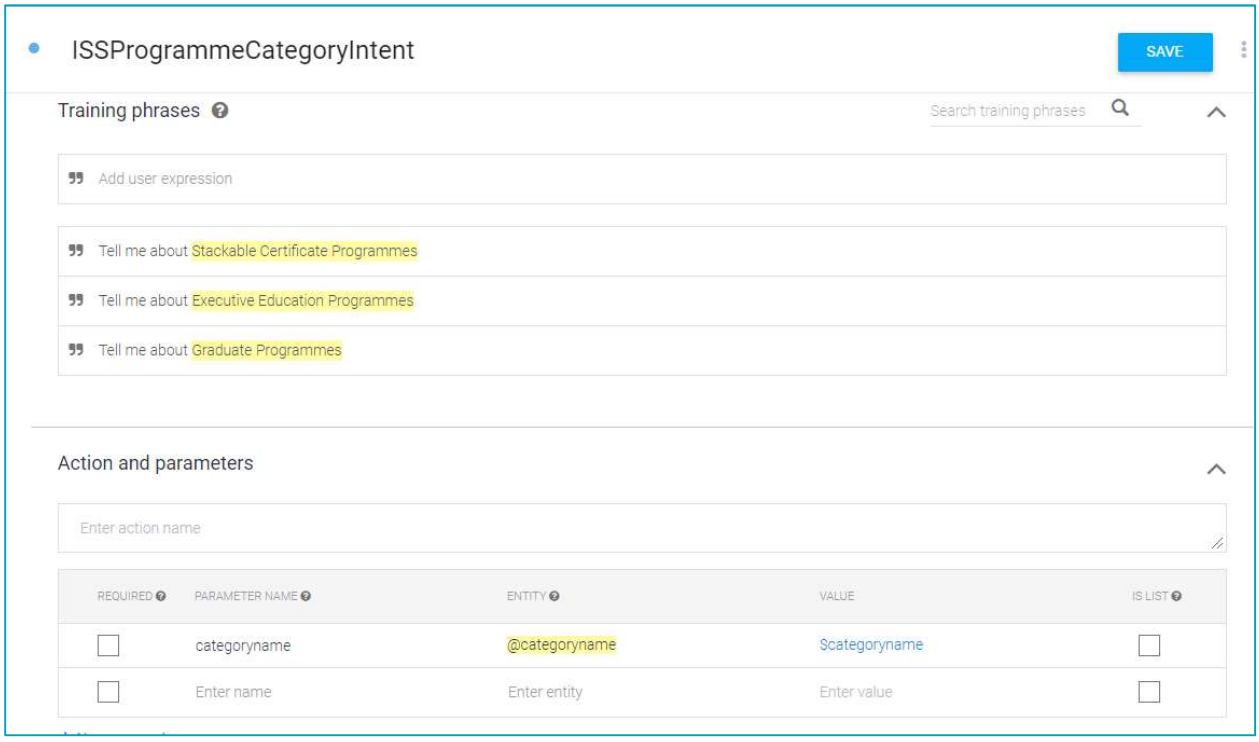
It was also noticed that there were similar terms available as modules in the collected Knowledge Base FAQ. Hence, responses for FAQ may be rather poor.

3.2.3 DIALOGFLOW – INTENT WITH FOLLOWUP

Exploration on Intent with followup was done using the *Graduate Programmes*. As there were six pages, the various sections were collected manually from the webpages and entered into an excel. Entities (*see Appendices 6.1 Entities*) were identified and created based on *Graduate Programmes*, and the fulfillment script was done on the Inline Editor.

Intents was created:

Training phrases was added so that the entities can be detected directly.



Webhook has to be enabled for all created intents.

A welcome function is created to prompt users to select the topics available; which can be detected by the entities identified in the respective intents.

```
function welcome(agent) {
    agent.add(`Hello! I am ISS Ask-Me-Anything-BOT (AMAB).  How may I help you?`);
    agent.add(new Card({
        title: `NUS ISS`,
        text: `Founded in 1981, the Institute of Systems Science (Abbreviation: ISS; Chinese: 新加坡国立大学系统科学院) at the National University of Singapore provides graduate education, executive education, consultancy and research services. Its objectives are to develop infocomm leaders, and to drive business and organisation innovation.`,
        buttonText: 'NUS ISS Wikipedia Page',
        buttonUrl: amabISSUrl
      })
    );
    agent.add(`To answer enquiries related to ISS programs, courses, and related information`);
    agent.add(new Suggestion(`Tell me about Graduate Programmes`));
    agent.add(new Suggestion(`Tell me about Executive Education Programmes`));
    agent.add(new Suggestion(`Tell me about Stackable Certificate Programmes`));
  }
```

```
function getISSProgrammeCategoryIntent(agent) {
    const catname = agent.parameters.categoryname;

    if (catname === `Graduate Programmes`) {
      agent.add(new Suggestion(`Graduate Diploma in Systems Analysis`));
      agent.add(new Suggestion(`Master of Technology in Digital Leadership`));
      agent.add(new Suggestion(`Master of Technology in Enterprise Business Analytics`));
      agent.add(new Suggestion(`Master of Technology in Intelligent Systems`));
      agent.add(new Suggestion(`Master of Technology in Software Engineering`));
    } else if (catname === `Executive Education Programmes`) {
      agent.add(`Please select one of the discipline.`);
    } else if (catname === `Stackable Certificate Programmes`) {
      agent.add(`Click on the links below to know more`);

    }
  }
```

The next intent was created as a follow-up intent, and the categoryname can be identified from the context using

const catname = agent.parameters.categoryname;

Subsequent intents were created, and the previously selected values was extracted from the context. A map was created so that the eventual intent could do a lookup for the correct response.

var map = {"Graduate Diploma in Systems Analysis:Overview:Purpose": "The Graduate Diploma in Systems Analysis programme (GDipSA) is designed for non-IT graduates intending to craft a new career path in the IT industry. IT graduates and professionals who wish to advance their careers in their current field and recognise the need to equip themselves with the latest IT knowledge and skills to stay relevant may apply as well."}

Graduate Diploma in Systems Analysis <- the first selected value

Overview <- second selected value

Purpose <- third selected value

3.2.4 CHATBOT WEB INTEGRATION

The NUS ISS website was crawled as a whole to create the raw HTML page for the ChatBot to "sit" in. A ChatBot web integration layer was used to visualize the ChatBot on the HTML. Like most ChatBot, clicking on the icon *(Figure 3a)* will reveal the ChatBot window (*Figure 3b)*; user will be able to type question and receive response from this window.

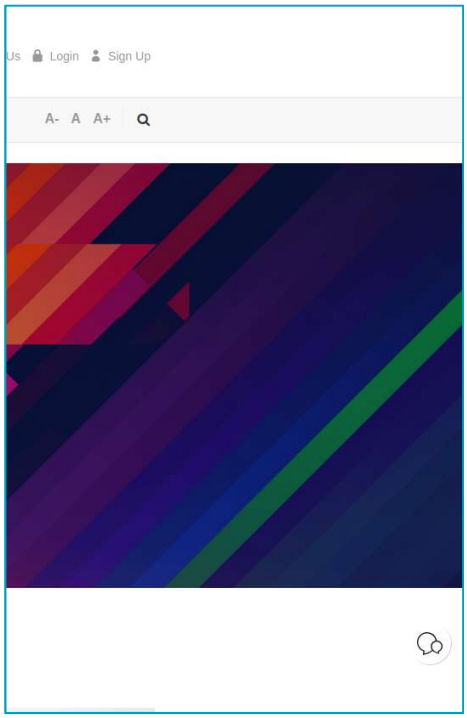


Figure 3a

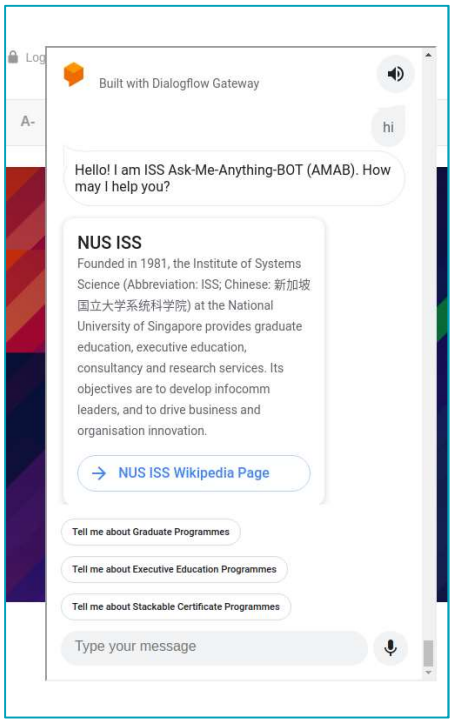


Figure 3b

The particular ChatBot web integration layer was selected as it could display rich text (i.e. buttons and cards) from DialogFlow's responses.

Further testing reveals that the combination of both Intent based and FAQ could not provide accurate results. Hence, in order to make the ChatBot effective, all extracted FAQs was converted to Entities, and Intents (with follow-up intents). The FAQs can still yield correct response when the questions asked is spot-on.

4 CONCLUSION & REFERENCES

4.1 LIMITATIONS

There are too many variations of HTML pages from the URL sources. Much time will be needed to explore the text extraction criteria in order for the information to be comprehensive. What that was extracted was around 70-80% complete, based on Discipline modules. There was insufficient time to explore the other top-level menu items. For the 20-30% errors, no mapping was created, and a friendly error message will be displayed.

agent.add(`Sorry, but i can't find the answer.`);

To portray a complete picture, links was provided for those categories that was yet to be explored/extracted.

4.2 IMPROVEMENTS AND FUTURE ENHANCEMENTS

Possible improvements to the data model would be to gain insights on prospective students and also the institution, namely, NUS ISS itself. The website could be organized in a way such that students can easily find the information they are looking for. A content management system (CMS) can be setup to catalog and manage the content/information provided in NUS ISS website. The same content management can also provide the data needed for the ChatBot to consume, so that; students can find what they need, and the institution (NUS ISS) can push the relevant message it wants to portray to the students. With the content management system providing data, there would not need a webscraping mechanism to extract information from the website.

If webscraping is the only way to go, than much time will be needed to explore and optimize the extraction of the categories/modules in the website. The following would be my recommendations:

1) Catalog the possible intents prospective students may be looking for
2) Classify the webpages (and sections of the webpages) according to the identified intents
3) Create webscraping scripts to extract webpage information, based on identified intent
4) Clean the information collected (some pages tend to be rather descriptive, short answers would be preferred in a ChatBot due to limited viewing area)
5) Create an application with database to store and retrieve the information collected
6) Create Dialogflow entities and intents
7) Link up application to Dialogflow's fulfillment
8) Create a ChatBot server to assess Dialogflow and to visualize the results based on Dialogflow's responses

4.3 CREATION OF DIALOGFLOW

This section is only necessary if it is required to create a new DialogFlow.

Import as ZIP AMAB.zip

Fulfillment Inline Editor

Use firebaseFulfillment.zip

Create Knowledge Base

Upload individual files as knowledge document

(Knowledge Type = FAQ, Mime Type = text/csv)

From scrapy-knowledge based faq.zip

Follow instructions from https://github.com/MishUshakov/dialogflow-web-v2 and https://dialogflow.cloud.ushakov.co/ to link DialogFlow to the gateway. Google Cloud account will be required.

Download/clone https://github.com/MishUshakov/dialogflow-web-v2

Follow instructions to link to your Google Cloud account.

5 BIBLIOGRAPHY

1) Dialogflow
https://dialogflow.com/

2) Google Cloud: Cloud Computing Services
https://cloud.google.com/

3) Web App for Dialogflow V2. 100/100 PWA. Actions on Google, Rich Components
https://github.com/MishUshakov/dialogflow-web-v2

4) Dialogflow Gateway
https://dialogflow.cloud.ushakov.co/

5) Stripping HTML from text and markdown for readability
http://mikelev.in/2014/01/stripping-html-text-markdown-readability/

6 APPENDICES

6.1 ENTITIES

categoryname	programmename	level1	level2
Graduate Programmes	Graduate Diploma in Systems Analysis	Overview	Purpose
Graduate Programmes	Graduate Diploma in Systems Analysis	Overview	Next Intake Time
Graduate Programmes	Graduate Diploma in Systems Analysis	Overview	Duration
Graduate Programmes	Graduate Diploma in Systems Analysis	Overview	Application Deadline
Graduate Programmes	Graduate Diploma in Systems Analysis	Overview	Technical learning outcomes
Graduate Programmes	Graduate Diploma in Systems Analysis	Overview	Non-technical learning outcomes
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Methodology
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Technology
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Databases
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Programming – C# Suite
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Programming - Java Suite
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Programming - Python Suite
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Full stack Solution Development
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Internet Application Development
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Project Management
Graduate Programmes	Graduate Diploma in Systems Analysis	Modules	Mobile Solution Development
Graduate Programmes	Graduate Diploma in Systems Analysis	Project & Internship	Summary
Graduate Programmes	Graduate Diploma in Systems Analysis	Project & Internship	Objectives
Graduate Programmes	Graduate Diploma in Systems Analysis	Project & Internship	Learning Outcomes
Graduate Programmes	Graduate Diploma in Systems Analysis	Fee & Loans	
Graduate Programmes	Graduate Diploma in Systems Analysis	Admission & Application	
Graduate Programmes	Graduate Diploma in Systems Analysis	Career Path	

Graduate Programmes	Master of Technology in Enterprise Business Analytics	Overview	Purpose
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Overview	Next Intake Time
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Overview	Duration
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Overview	Application Deadline
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Overview	Learning outcomes
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Modules	Analytics Project Management and Delivery
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Modules	Core Analytics Techniques
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Modules	Customer Analytics
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Modules	Big Data Processing
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Modules	Practical Language Processing
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Modules	Advanced Predictive Modelling Techniques
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Project & Internship	Summary
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Project & Internship	Objectives
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Project & Internship	Learning Outcomes
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Fee & Loans	
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Admission & Application	
Graduate Programmes	Master of Technology in Enterprise Business Analytics	Career Path	
Graduate Programmes	Master of Technology in Digital Leadership	Overview	Purpose
Graduate Programmes	Master of Technology in Digital Leadership	Overview	Next Intake Time
Graduate Programmes	Master of Technology in Digital Leadership	Overview	Duration
Graduate Programmes	Master of Technology in Digital Leadership	Overview	Application Deadline
Graduate Programmes	Master of Technology in Digital Leadership	Overview	Learning outcomes
Graduate Programmes	Master of Technology in Digital Leadership	Modules	Practice of Digital Business
Graduate Programmes	Master of Technology in Digital Leadership	Modules	Digital Transformation
Graduate Programmes	Master of Technology in Digital Leadership	Modules	Digital Leadership & People
Graduate Programmes	Master of Technology in Digital Leadership	Modules	Digital Leadership Capstone
Graduate Programmes	Master of Technology in Digital Leadership	Project & Internship	Summary
Graduate Programmes	Master of Technology in Digital Leadership	Project & Internship	Objectives

Graduate Programmes	Master of Technology in Digital Leadership	Project & Internship	Learning Outcomes
Graduate Programmes	Master of Technology in Digital Leadership	Fee & Loans	
Graduate Programmes	Master of Technology in Digital Leadership	Admission & Application	
Graduate Programmes	Master of Technology in Digital Leadership	Career Path	
Graduate Programmes	Master of Technology in Intelligent Systems	Overview	Purpose
Graduate Programmes	Master of Technology in Intelligent Systems	Overview	Next Intake Time
Graduate Programmes	Master of Technology in Intelligent Systems	Overview	Duration
Graduate Programmes	Master of Technology in Intelligent Systems	Overview	Application Deadline
Graduate Programmes	Master of Technology in Intelligent Systems	Overview	Learning outcomes
Graduate Programmes	Master of Technology in Intelligent Systems	Modules	Intelligent Reasoning Systems
Graduate Programmes	Master of Technology in Intelligent Systems	Modules	Pattern Recognition Systems
Graduate Programmes	Master of Technology in Intelligent Systems	Modules	Intelligent Robotic Systems
Graduate Programmes	Master of Technology in Intelligent Systems	Modules	Intelligent Sensing Systems
Graduate Programmes	Master of Technology in Intelligent Systems	Modules	Intelligent Software Agents
Graduate Programmes	Master of Technology in Intelligent Systems	Modules	Practical Language Processing
Graduate Programmes	Master of Technology in Intelligent Systems	Project & Internship	Summary
Graduate Programmes	Master of Technology in Intelligent Systems	Project & Internship	Objectives
Graduate Programmes	Master of Technology in Intelligent Systems	Project & Internship	Learning Outcomes
Graduate Programmes	Master of Technology in Intelligent Systems	Fee & Loans	
Graduate Programmes	Master of Technology in Intelligent Systems	Admission & Application	
Graduate Programmes	Master of Technology in Intelligent Systems	Career Path	
Graduate Programmes	Master of Technology in Software Engineering	Overview	Purpose
Graduate Programmes	Master of Technology in Software Engineering	Overview	Next Intake Time
Graduate Programmes	Master of Technology in Software Engineering	Overview	Duration
Graduate Programmes	Master of Technology in Software Engineering	Overview	Application Deadline
Graduate Programmes	Master of Technology in Software Engineering	Overview	Learning outcomes
Graduate Programmes	Master of Technology in Software Engineering	Modules	Architecting Scalable Systems
Graduate Programmes	Master of Technology in Software Engineering	Modules	Architecting Smart Systems

Graduate Programmes	Master of Technology in Software Engineering	Modules	Designing and Managing Products and Platforms
Graduate Programmes	Master of Technology in Software Engineering	Modules	Engineering Big Data
Graduate Programmes	Master of Technology in Software Engineering	Modules	Securing Ubiquitous Systems
Graduate Programmes	Master of Technology in Software Engineering	Project & Internship	Summary
Graduate Programmes	Master of Technology in Software Engineering	Project & Internship	Objectives
Graduate Programmes	Master of Technology in Software Engineering	Project & Internship	Learning Outcomes
Graduate Programmes	Master of Technology in Software Engineering	Fee & Loans	
Graduate Programmes	Master of Technology in Software Engineering	Admission & Application	
Graduate Programmes	Master of Technology in Software Engineering	Career Path	
Executive Education Programmes	Artificial Intelligence		
Executive Education Programmes	Cybersecurity		
Executive Education Programmes	Data Science		
Executive Education Programmes	Digital Agility		
Executive Education Programmes	Digital Innovation & Design		
Executive Education Programmes	Digital Strategy & Leadership		
Executive Education Programmes	Digital Products & Platforms		
Executive Education Programmes	Professional Conversion Programmes		
Executive Education Programmes	SkillsFuture Series		
Executive Education Programmes	Smart Health Leadership		
Executive Education Programmes	Software Systems		
Executive Education Programmes	StackUp - Startup Tech Talent Development		
Stackable Certificate Programmes	Data Science		
Stackable Certificate Programmes	Digital Solutions Development		

Stackable Certificate Programmes	Artificial Intelligence		
Stackable Certificate Programmes	Smart Systems & Platforms		

6.2 FILES & DESCRIPTION

The following describes the files that are submitted along with this project

1) ChatBot.xlsx
   Spreadsheet of the collection information, entities, question & answers, URLs and the various disciple/module/section mapping.

2) firebaseFulfillment.zip
   Fulfillment script in DialogFlow.

3) convert2.py
   Python script used to web scrap.

4) scrapy-knowledge based faq.zip
   Initial web scrap results, collected as FAQ.

5) AMAB.zip
   DialogFlow export as zip.