Institute of System Science Automated Chatbot (ISSAC)

with **The Sunday Lunatics**

Project Report

Brought to you by:

THE SUNDAY LUNATICS

Tang Meng
Tan Jun Khiang
Tan Wei Lian
Leong Jun Hun, Darryl

Executive Summary / Paper Abstract

ISSAC is the implementation of an Automated Chatbot to provide information about the Institute of System Science (ISS) of the National University of Singapore. ISSAC is able to answer a wide range of ISS related queries, covering general information, Staff, Graduate Programmes, Executive Education/Stackable, Info Sessions and Seminar and Conferences. General Information such as what ISS does, mission, core values and details about its Management Board is provided by ISSAC. Users can also enquire about Graduate and Executive Education Programme course description, fees, application deadlines, key takeaways. Follow Up Intents were created to prompt users for their nationality when enquiring about Graduate Programme fees to enable a more natural conversation flowFor staff, users can as about staff positions, email addresses and the courses they teach.

The key challenge was to identify key information contained with the website and extracting the relevant information from the website. Information Extraction (IE) techniques were applied to the website to summarize natural language text into a structured set of facts.

On the backend, Google's Dialogflow was used.A custom python application was used to develop a backend fulfilment engine, accompanied by the Knowledge Base capability of Dialogflow. On the frontend, a Flask application was used to simulate a chatbot interface.

Business Problem Background

The ISS website contains a plethora of information - from the root domain (https://www.iss.nus.edu.sg/), there are links to more than 150 links. The main challenge was to identify useful information and automatically extract it such that information could be extracted efficiently and accurately.

Project Objectives & Success Measurements

The objectives of the project are identified as:

- To build a chatbot system for ISS website
- To answer enquiries related to ISS programs, courses, and its general information

For ISSAC to be deemed successful, we have defined the following scenarios and desired outcomes:

| Scenario	Desired Outcome
For a user who is interested in finding out more about ISS	ISSAC is able to provide information on what ISS does, Mission and Core Values, Address, Staff and Student Count
For a user who is interested in Executive Education Course or Masters Programmes	ISSAC is able to provide information on the Description, Fees, Application Deadline (for Masters Programmes)
For a user who is interested in finding out about upcoming information session	ISSAC is able to provide details such as date, timing and venue for upcoming information sessions
For a user who is interested to know about a specific faculty member	ISSAC is able to provide details such as designation, email addresses and the courses taught by the faculty member.
For a user interested in contacting ISS	ISSAC provides email address and contact number to call

Project Solution

We identified the need to address both structured and unstructured data from the ISS website. Structured data would be data relating to course details, such as course fees, application

deadline as well as faculty information. Example scenarios which use these data are detailed in the success measurement section of this report.

Since the critical functionality of the chatbot is dependent on the structured data, we identified it as the key focus of ISSAC.

Process Flow
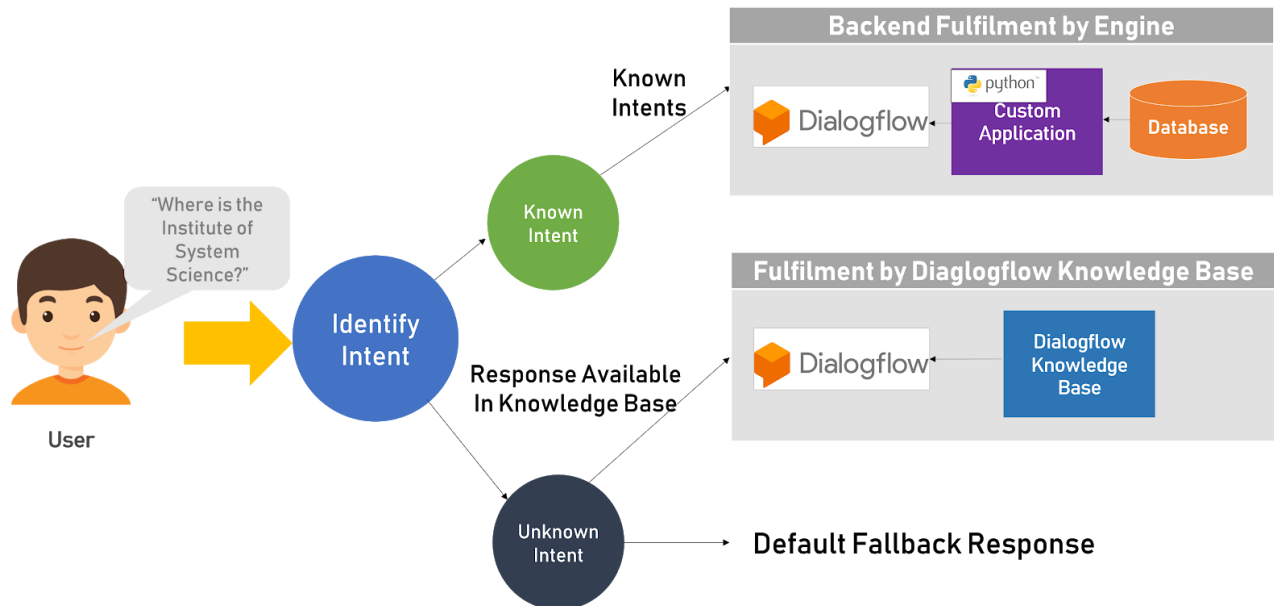
A typical user flow of ISSAC is as follows:
1. A user asks a question
2. ISSAC attempts to identify Intent:
   a. If the Intent is identified, an API call is made to the backend fulfillment engine, together with the slots
   b. If the Intent cannot be identified:
      i. ISSAC looks to the knowledge bases provided by DialogFlow
      ii. If no response can be identified, ISSAC returns the Default Fallback Response

Information Extraction Approach

Rationalist (Rule-based) and Empirical (Corpus-based) approaches were considered for Information Extraction. Due to the more precise results achieved from a rule-based system and the lack of volume of data for a corpus based approach, a Rationalist approach was chosen.

4

The primary method of information extraction was similar to the tabular form [1] , where a set of values usually in the neighborhood of several keywords was extracted. The output structure of this method resembles a table.

For the ISS website, information was extracted based on the div names of html tags. The python package, BeautifulSoup was used for this purpose. BeautifulSoup created a parse tree for parsed pages, which we used to extract data from HTML. Samples of the Information Extraction script can be found in Appendix B.

The following Domain Model was developed to captured the key information to be captured from the ISS website. The key entities identified were Staff, Graduate Programmes, Executive Education/Stackable, Info Sessions and Seminar and Conferences.
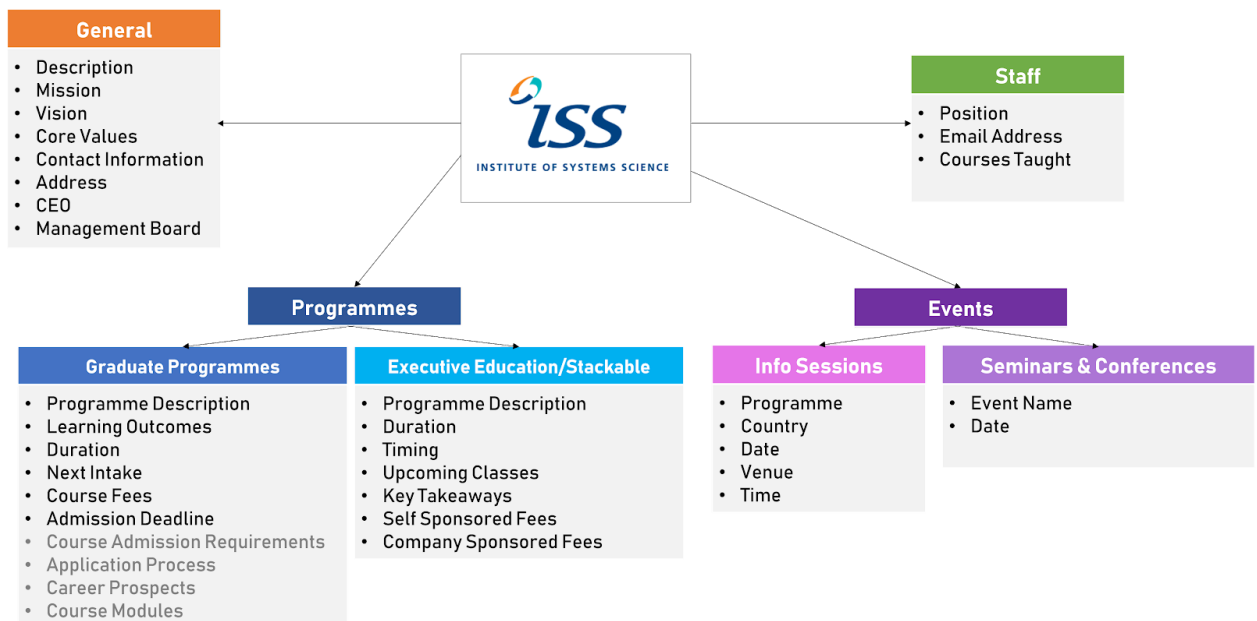
Intents and Sample Utterances were developed. These Intents were mapped into the following categories.

However, only the Intents related to Graduate Programs were implemented.

Category	Intents
General	● GetGeneralInformation ● GetMission ● GetVision

---

[1] "Information Extraction From University Web Pages - UKnowledge."
https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1219&context=gradschool_theses. Accessed 29 May. 2019.

	● GetCoreValues ● GetLocation ● GetAvailableProgrammes ● GetCEO ● GetDirector ● GetFacultyCount ● GetStudentCount ● GetProgrammeCount ● GetFoundingYear ● GetContactDetails ● GetEmailAddress ● GetPhoneNumber ● GetEnquiryLineOperatingHours ● GetOfficeOperatingHours ● GetManagementBoardChairman ● GetManagementBoardMembers ● GetBestLecturer ● GetWorstLecturer ● GetBestProgramme
Graduate Programmes	● getCourseFee ● getCourseDescription ● getCourseLearningOutcomes ● getCourseDuration ● getCourseNextIntake ● getCourseAdmissionRequirements ● getCourseApplicationProcess ● getCourseCareerProspects ● getCourseModules ● getApplicationDeadline
Executive Education Programme	● getCourseFee ● getEEDuration ● getEETiming ● getEEDescription ● getEETakeAway ● getEEUpcomingClasses
Info Sessions	● getNextInfoSession
Teaching Staff	● GetPosition ● GetEmailAddress ● GetCourseCount ● GetCoursesTaught

Structured Data Representation

Three DialogFlow Entities were created to ensure high accuracy of response and facilitate follow up intent.

- Graduate-programme captures the graduate programmes offered by ISS and possible synonyms.
- Application-group captures the possible nationality groups which interested applicants can fall under.
- Graduate-programme-ptft captures whether the nature of study, whether it is part time or full time.

graduate-programme

Reference Value	Synonyms
graddip-sa	Graduate Diploma, Graduate Diploma in System Analysis, System Analysis
mtech-dl	Master Of Technology In Digital Leadership, mtech-dl
mtech-ebac	Enterprise Business Analytics, MTech Ebac, MTech Enterprise BA, Master Of Technology In Enterprise Business Analytics
mtech-is	Intelligent Systems, MTech IS, MTech Intelligent Systems, Master of Technology in Intelligent Systems, Masters of Intelligent Systems
mtech-swe	Master Of Technology In Software Engineering, mtech-swe

application-group

Reference Value	Synonyms
non-sg	Foreigner, Foreigners, non Singaporean, non-Singaporean, non-Singaporeans, non-sg
sg	Singapore PR, Singaporean, Singaporean PR, Singaporeans, Singaporeans PR

graduate-programme-ptft

Reference Value	Synonyms
ft	ft, full time, full-time, fulltime
pt	part time, part-time, pt

Follow Up Intent: Graduate Programme Fees

We leveraged on DialogFlow's follow up intent capabilities to enable a more natural conversation flow for Graduate Programme Fees.When a user asks about Programme Fees for a specific course, ISSAC will ask the user whether he/she is a Singaporean and whether the he/she intends to take the programme on a full time or part time basis.

## Unstructured Data Representation

The Knowledge functionality within DialogFlow was used for Unstructured data representation. It utilizes knowledge connectors which complement defined intents as discussed in the Structured Data Representation section. The knowledge connectors parse documents (for example, FAQs or articles) to find automated responses[2]. The following list of FAQ web pages and extracted information are uploaded into the DialogFlow Knowledge database:

- ISS General Info FAQs

---

[2] "Knowledge connectors | Dialogflow Documentation | Google Cloud."
https://cloud.google.com/dialogflow-enterprise/docs/knowledge-connectors. Accessed 30 May. 2019.

- Programme-related FAQs
- Graduate Admission FAQs
- NUS Alumni FAQs
- Professional Conversion Programme (PCP) and General FAQs
- SkillsFuture Credits FAQ
- SkillsFuture Study Award FAQs

Project Implementation
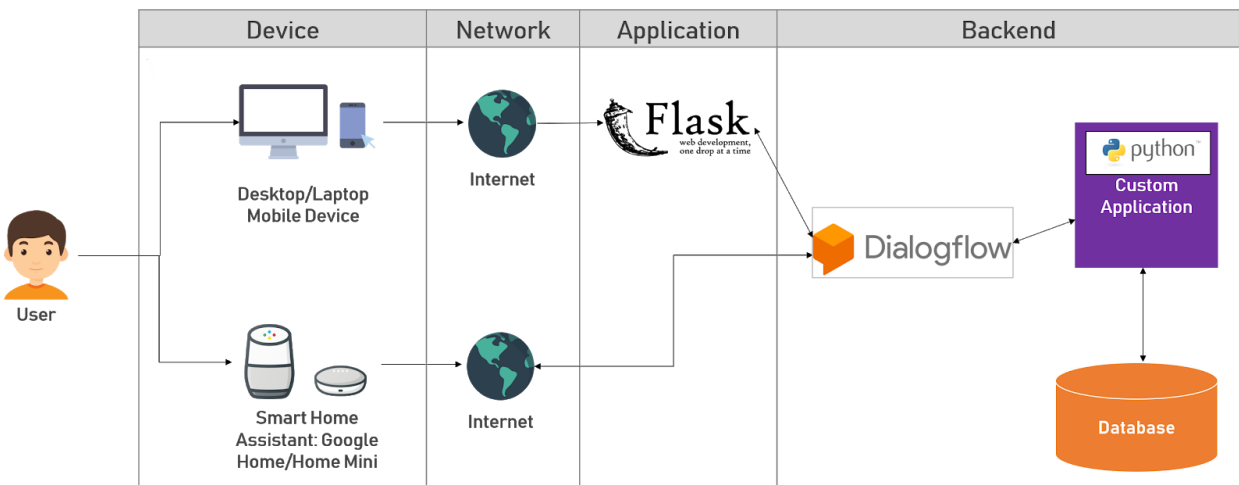
- Dialogflow: human–computer interaction technologies based on natural language conversations
- Custom Application: A webhook application configured as a fulfillment in Dialogflow
- Flask: Frontend web application framework with material design.
- Database

Project Performance & Validation

From the identified scenarios, the testing was performed. The following table summarizes the results of testing:

Scenario	Desired Outcome	Achieved?
For a user who is interested in finding out more about ISS	ISSAC is able to provide information on what ISS does, Mission and Core Values, Address, Staff and Student Count	Yes (See screenshot below)

For a user who is interested in Executive Education Course or Masters Programmes	ISSAC is able to provide information on the Description, Fees, Application Deadline (for Masters Programmes)	Yes (See screenshot below)
For a user who is interested in finding out about upcoming information session(s)	ISSAC is able to provide details such as date, timing and venue for upcoming information sessions	Yes (See screenshot below)
For a user who is interested to know about a specific faculty member	ISSAC is able to provide details such as designation, email addresses and the courses taught by the faculty member.	Yes (See screenshot below)
For a user interested in contacting ISS	ISSAC provides email address and contact number to call	Yes (See screenshot below)

Scenario: For a user who is interested in finding out more about ISS
- What does ISS do?
- What is ISS mission?
- What are ISS Core Values?
- Where is ISS?
- How many Students are there in ISS?

Scenario: For a user who is interested in Executive Education Course or Masters Programmes
- What is the Masters in Intelligent Systems?
- How much is the Masters in Intelligent Systems?
- When is the application deadline for the Masters in Intelligent Systems?
-

What is the Masters in Intelligent Systems?

The NUS Master of Technology in Intelligent Systems programme is targeted at working professionals who wish to be able to design and build systems that utilise Artificial Intelligence and other Smart Systems techniques. Application areas are wide and diverse, and include robotics, autonomous vehicles, intelligent sensing systems, Internet of Things, Smart City applications and Industry 4.0 applications, as well as applications within business and commerce.

How much is the Masters in Intelligent Systems?

The Full Time Fees for Master Of Technology In Intelligent Systems are S$55,319.00 to S$59,171.00 for International Students and S$20,939.90 to S$22,095.50 for Singaporeans. The Part Time Fees for Master Of Technology In Intelligent Systems are S$20,939.90 to S$22,095.50 for International Students and S$55,319.00 to S$59,171.00 for Singaporeans.

When is the application deadline for the Masters in Intelligent Systems?

For Full-time - Overseas applicants Master Of Technology In Intelligent Systems the deadline for application is 28 February 2019,For Full-time - Local applicants the deadline for application is 31 March 2019

Scenario: For a user who is interested in finding out about upcoming information session
- When is the next information session for the Masters in Intelligent Systems?

When is the next information session for the Masters in Intelligent Systems?

The next information session for the Master of Technology in Intelligent Systems Information Session is in Singapore on 8 June 2019 at 01:30 PM to 03:00 PM at National Library Building.

Scenario: For a user who is interested to know about a specific faculty member
- Who is Fan Zhen Zhen?
- What is Fan Zhen Zhen's email?

12

- How many courses does Fan Zhen Zhen teach?
- What does Fan Zhen Zhen teach?
- Who is Wang Ao Bo?

Who is Fan Zhen Zhen?

FAN Zhen Zhen is a Senior Lecturer & Consultant, Artificial Intelligence Practice

What is Fan Zhen Zhen's email?

FAN Zhen Zhen's Email is issfz@nus.edu.sg

How many courses does Fan Zhen Zhen teach?

FAN Zhen Zhen teaches 3 courses.

What does Fan Zhen Zhen teach?

FAN Zhen Zhen teaches NICF- Reasoning Systems (SF) ,NICF- Text Processing using Machine Learning (SF) ,NICF- Text Analytics (SF)

Who is Wang Ao Bo?

WANG Aobo is a Lecturer & Consultant, Artificial Intelligence Practice

Scenario: For a user interested in contacting ISS
- What is ISS phone number?
- What is ISS email address?

What is ISS phone number?

For Registration of Courses or Training Enquiries, please call 6516 2020. For Graduate Programme Enquiries, Application & Admission, please call 6601 3161. For Current Graduate Programme Students & Administrative Matters, please call 6516 1654.

What is ISS email address?

For general enquiries, please send an email isstraining@nus.edu.sg

Additional Scenarios:

Management Board

- Who is the chairman of the ISS Management Board?
- Who is on the ISS Management Board?

Project Conclusions: Findings & Recommendation

By adopting an automated information extraction approach, we were able to efficiently and accurately extract key information from the ISS website into tabular format, converting unstructured data into structured form. By leveraging on the knowledge base functionality of Dialogflow, we were also able to capture less critical information (e.g. recent news, events) Through the validation we performed, ISSAC has demonstrated its usefulness in providing information to users in a natural conversational like manner.

The inherent limitation of the Information Extraction approach used in ISSAC is its lack of extensibility. Currently, the scripts used to extract information from the ISS website cannot be used to extract data from similar websites of other universities.

Future enhancements to the project could involve more detailed Information Extraction using Natural Language Processing techniques. By performing type tagging, normalization and pre-parsing of the text, more information could be extracted from the website.

References

"Information Extraction From University Web Pages - UKnowledge."
https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1219&context=gradschool_theses. Accessed 29 May. 2019.

"Knowledge connectors | Dialogflow Documentation | Google Cloud."
https://cloud.google.com/dialogflow-enterprise/docs/knowledge-connectors. Accessed 30 May. 2019.

Appendix

ISS	- Vision, Mission, Core Values - History - Achievements - Career Services - Community - Industrial Partners - Alumni - Students - Staff - Student Life - Newsletter - Location - Contact Details - Careers in ISS
Programme Types - Executive Education - Graduate Programmes - Stackable Certificates	- Discipline / Specialization - Training and Career Roadmap - Eligibility - Course Syllabus and Objectives - Course Schedule and Fees - Financial Assistance and Funding - Application Process and Procedures - Preview Talks and Information Sessions - Reviews by Organizations and Students
Centre of Excellence	- Vision, Mission, Core Values - History - Capabilities and Capacity

Appendix B

For the extraction of details from Executive Education Courses.(e.g. https://www.iss.nus.edu.sg/executive-education/course/detail/nicf--design-secure-mobile-architecture-ssdla-sf/cybersecurity)

Upcoming Classes:

Upcoming Classes

Class 1	20 Nov 2019 to 22 Nov 2019 (Full Time)	+
Class 2	08 Jan 2020 to 10 Jan 2020 (Full Time)	+

```
#upcomingClasses
  try:
    allModules= soup.find('h2', text ='Upcoming Classes').findNext('div', {'class':
'accordion_header clearfix'})
    sessions = allModules.findAll('span', {'class': 'accordion_header-date'})
    records = []
    for session in sessions:
      sessionDate = session.text.replace('\n', '').replace('\r', '')
      sessionDateStripped = sessionDate.strip()
      records.append(sessionDateStripped)
    ee_upcoming_classes = records
  except:
    ee_upcoming_classes = 'There are currently no upcoming classes scheduled.'
```

Self Sponsored Fees:

Self-sponsored

	International Participants	S'poreans and PRs (aged 21 and above)	SkillsFuture Mid-Career Enhanced Subsidy[1] (S'poreans aged 40 and above)	Workfare Training Support[2] (S'poreans aged 35 and above, and earn ≤ $2,000 per month)
Full course fee	S$2700	S$2700	S$2700	S$2700
SSG grant	-	(S$1890)	(S$1890)	(S$1890)
Nett course fee	S$2700	S$810	S$810	S$810
7% GST on nett course fee	S$189	S$56.70	S$56.70	S$56.70
Total nett course fee payable, including GST	S$2889	S$866.70	S$866.70	S$866.70
Less additional funding if eligible under various schemes	-	-	(S$540)	(S$675)
Total nett course fee payable, including GST, after additional funding from the various funding schemes	**S$2889**	**S$866.70**	**S$326.70**	**S$191.70**

```
#Self Sponsored
try:
    all_ss_fees= soup.find('h4', text ='Self-sponsored').findNext('table')
    last_row = all_ss_fees("tr")[-1]
    #International Participants
    ss_international = last_row("td")[1].text.replace('\n', '')
    #Singaporeans and PRs
    ss_singaporean = last_row("td")[2].text.replace('\n', '')
    #SkillsFuture Mid-Career Enhanced Subsidy
    ss_skillsfuture = last_row("td")[3].text.replace('\n', '')
    #Workfare Training Support
    ss_workfare = last_row("td")[4].text.replace('\n', '')
except:
    ss_international = 'tbd'
```

```
ss_singaporean = 'tbd'
ss_skillsfuture = 'tbd'
ss_workfare = 'tbd'
```