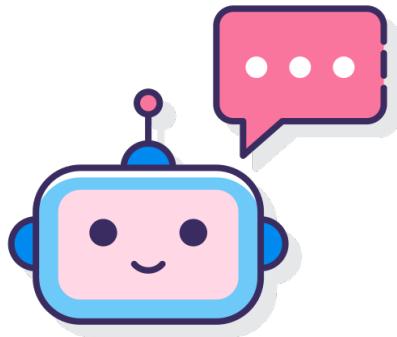


CareerWise Chabot

Master of Technology

Project Report



Team Members

Kwatt Ivy (A0269639H)
Chua Jack Yune (A0269363U)
Borromeo, Angelie Quiapo(A0270177A)

EXECUTIVE SUMMARY

Making career decision is hard, it is even harder when your career conflicts with your passion. For a fresh graduate, that will be 40 years for the rest of their life doing something without much joy and passion. CareerWise Chatbot can help them decide a career choice that matters.

To help individual with this dilemma, we design a chatbot that can deliver a tailored career path recommendations based on user input on either their interests or domain of expertise. CareerWise Chatbots can also provide additional information about educational programs, such as course descriptions, and university options and potential compensation trends.

The benefits of using CareerWise Chatbots for career path selection include convenience, accessibility, and personalised recommendations. CareerWise Chatbots can be potentially accessed through a variety of platforms, including social media, messaging apps, and websites (but for this project implementation the team have created a python flask base application which is intended to be hosted in Public Cloud Platform). This makes it easier for individuals to receive initial course recommendations on the go, without the need for lengthy research or face to face consultations.

We hope to provide a convenient and personalised way to receive career recommendations, making the career path selection process more accessible and efficient. As technology continues to evolve, it is likely that chatbots will become an

increasingly important part of education and career path selection.

Table of Contents

1	BUSINESS PROBLEM AND BACKGROUND	7
1.1	Current Solutions	7
1.2	Subject Matter Expert Interview	8
1.3	The Anonymous Survey	8
2	Project Objective	11
2.1	Target Audience	11
3	System's Features	12
3.1	System Intelligence	12
3.2	Ease of Access and Scalability	12
3.3	User Management	12
4	Knowledge Modelling	13
4.1	Knowledge Identification	13
4.2	Career Service Expert	15
4.3	O* NET Database	15
4.4	Knowledge Specification	17
4.4.1	Occupational Analysis	17
4.5	Job Descriptions Dataset	20
4.5.1	Salaries Dataset	20
4.6	Generating the Recommendation Scores	21
4.6.1	Career Interest	21
4.6.2	Remuneration	22
4.6.3	Work Experience and Education	22
4.6.4	Work Context	23
4.7	Similarity Score Aggregation	25
4.8	Additional Recommendations	27
4.8.1	Career Path Knowledge Graph	27
4.8.2	Education Programs	29
4.9	Knowledge Refinement	29
5	Solution and System Architecture	31
5.1	User Registration / Login	31
5.1.1	Technical Implementation	32
5.2	Chatbot Window	32
5.2.1	Technical Implementation	33

5.3	Question-answering session.....	34
5.3.1	Technical Implementation	35
5.4	Recommendations Results Index	37
5.4.1	Technical Implementation	38
5.5	Result Details Page	40
5.5.1	Technical Implementation	42
6	Overall System Architecture	44
7	Limitations.....	45
8	Conclusion	46
9	Improvements	47
10	Bibliography	48
11	Appendix	49

Figures

Figure 1-1 RIASEC Model taken from MySkillsFuture	7
Figure 1-2 Career journey survey	9
Figure 4-1 Example of raw data for Education, Training and Experience from O*NET	17
Figure 4-2 Example of processed O*NET data.....	18
Figure 4-3 PCA and Factor Analysis results for (from top to bottom) Values and Work Styles	19
Figure 4-4 MyCareersFuture dataset sample.....	20
Figure 4-5 Knowledge Representation	21
Figure 4-6 Steps taken in knowledge refinement	30
Figure 5-1 Screenshot of login/register page.....	31
Figure 5-2 Screenshot of chatbot window at the beginning of the session. The sidebar on the left shows the user's chat history.....	33
Figure 5-3 Screenshot of chatbot window at the end of the session. A flashing green block button appears above the input section, prompting the user to click to generate results.....	35
Figure 5-4 question and answer chatbot process flow.....	37
Figure 5-5 Recommendation results index	38
Figure 5-6 Result details page shows the occupation description and related jobs (from MyCareersFuture) and salary ranges	41
Figure 5-7 Sankey diagram shows how this occupation is related to other occupations and the path it takes from a lower level occupation (left) to a higher level occupation (right)	41
Figure 5-8 Educational Programs and Fresh Graduate Trends	42

Figure 6-1 Overall System Architecture Diagram.....	44
Figure 6-2 PostgreSQL Entity Relationship Diagram.....	44

Tables

Table 4-1 Knowledge Source and Acquisition Technique.....	14
Table 4-2 O*NET Occupational Profile Domains Dataset	16
Table 4-3 Open-ended questions and priority option	23
Table 4-4 Work context questions.....	25
Table 4-5 Data architecture for recommender system.....	26
Table 4-6 Career pathways in O*NET	29
Table 11-1 System Functionalities - Modules Mapping	50

1 BUSINESS PROBLEM AND BACKGROUND

Choosing a career path and selecting the right course of study can be a challenging and daunting task for both young adults and experienced professionals. With so many options available, it's important to take a balanced, multi-faceted approach which carefully weighs one's interests, skills, and long-term career goals. Although tools and services have been developed to help aspiring individuals search for a matching career, many of these have their limitations.

1.1 Current Solutions

An individual who intends to decide on a new career path can engage various services to help them make better decisions, such as career counselling. Psychometric tools, such as Holland's RIASEC Code personality test, are also available online as free or paid services to help individuals identify careers that match their interest and personality.

This model is proposed by American psychologist John L. Holland in the 1970s, which claims that vocational choices are determined by our interests and how we approach life situations. It groups people into categories of occupations: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E) and Conventional (C).

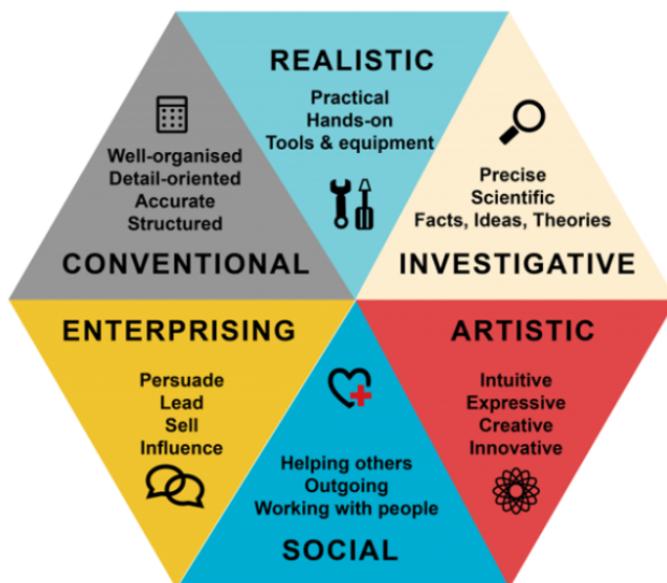


Figure 1-1 RIASEC Model taken from MySkillsFuture

This is the most widely adopted model for career classification model and is used in many career advice tools in the market like Self-Directed Search (<https://self-directed-search.com/>), O*NET Interest Profiler, and Singapore's MyCareersFuture Assessment Tool (<https://www.myskillsfuture.gov.sg/content/portal/en/assessment/landing.html>). The limitations of these tools is that they were built as extensions of paper-and-pencil models, where users have to fill in a questionnaire of at least 60 questions, and the scores would be summed up for each category to determine one's major category. Career recommendations

are given from a static list of jobs written by an occupational psychologist, providing broad recommendations based on the six categories.

1.2 Subject Matter Expert Interview

We had arranged a 2-hour long interview with a NUS Career Service Expert to get her input about our project's proposal objective, and scope of project. She had pointed several valid inputs which helped us recalibrate our intent and scope in the project. The following are some of the critical key points:

- Identify the pain points of the target audience that the system can solve.
- Collect data from individuals about what factors motivates clients' career decision making.

1.3 The Anonymous Survey

Following the advice and guidance of the Subject Matter Expert, we designed and administered an electronic survey to collect data on the research topic. We aimed to find out the challenges or pain points faced by fresh graduates and working professionals when seeking career advice when starting out on their career or making a career switch. The survey consisted of 6 multiple choice questions and 2 open ended questions. The main findings of the survey analysis are presented below.

For those students who are in the lower socio-economic group there is a limited accessibility to engage with the career practitioner or draw any guidance from different professionals due to a finite social circle. There is also some feeling of anxiety over raising simple questions, what question to ask and how career practitioners can be of support. We propose that by providing a tool as an initial point of assessment, would help the career path seekers to determine the best possible match based on their interests or field of expertise/skills.

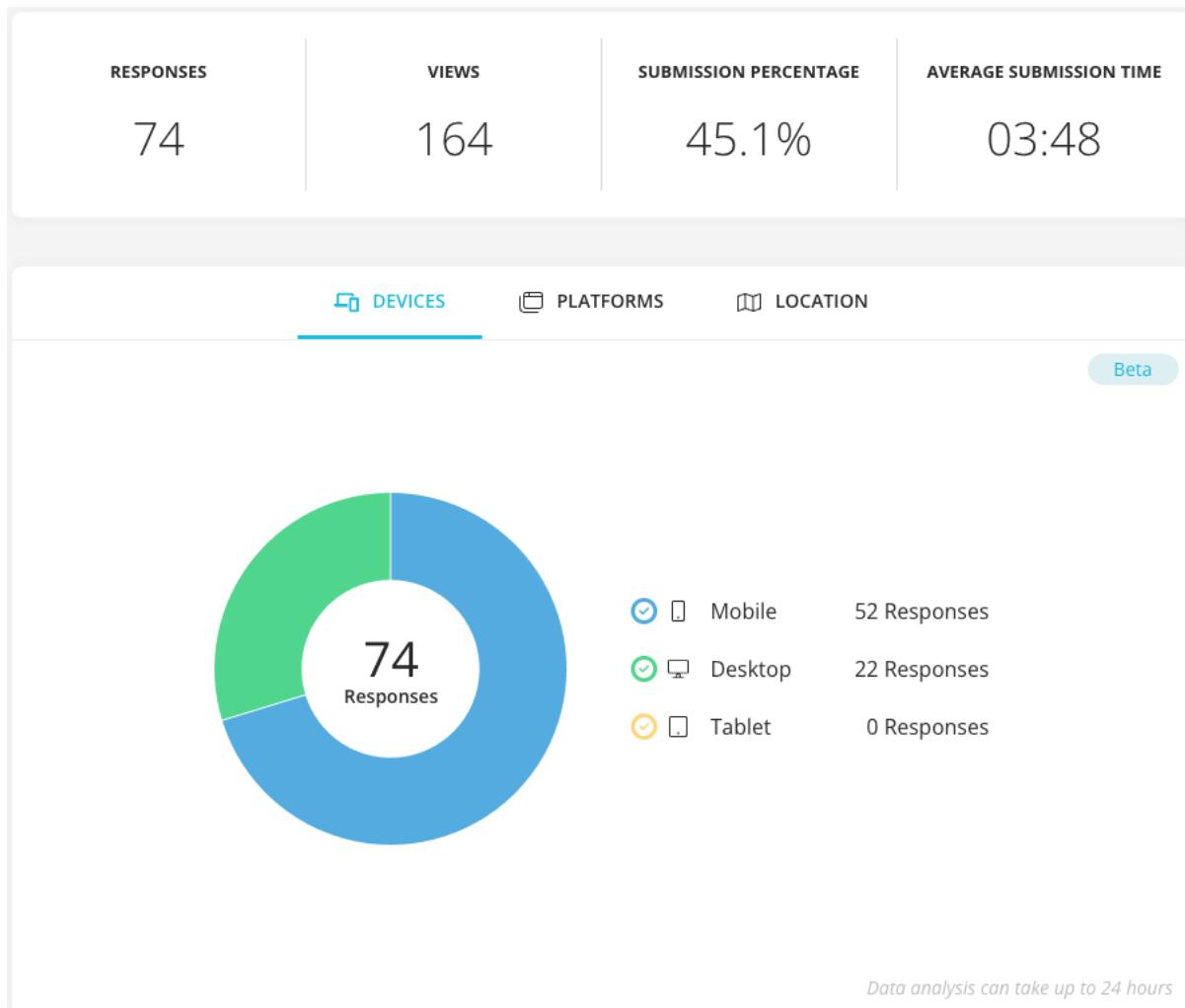


Figure 1-2 Career journey survey

For reference the responses on the survey can be found in the Appendices

Undergraduate Student: The undergraduate students had a response rate of 13.51%. Among half of them indicated that they will choose their career based on their interest, around 40% considered their domain expertise as the main factor for their career choice while the remaining 10% have other preferences.

Mid – Career (3 to 24 years of professional Experience): The majority of the survey respondents (67.57%) were mid-career professionals. More than half of them (54%) expressed their intention to change their career. Among those who wanted to switch their career, 52% based their decision on their domain expertise and another 36% on their interest. The remaining 12% had other reasons that were not included in the survey options.

Long Term Career (25 years and more professional Experience): Long term career professionals constituted 18.92% of the total survey respondents. Interestingly, 28.57% of them are still willing to change their career despite working for a long time. Among those who wanted to switch their career, the majority (57.14%) based their decision on their interest, while the rest (42.86%) on their expertise.

Career Counselling: The survey results showed that only 29.73% of the respondents had engaged in career consultation as part of their career development, while the majority (70.27%) had not. The reasons for not using career counselling varied, but some common themes emerged from the mid-career and long-term career professionals. These included lack of awareness of the availability and benefits of career counselling, lack of access to qualified and trustworthy counsellors, and lack of motivation or interest in seeking career guidance. Some representative quotes from these respondents were: “I had no chance to try it before and I don’t know where to get this service”, “I’ve never thought of career counselling but will definitely consider it now”, “Extra cost to go through such process” and “Did not know where and who to approach”.

Future Career Plan: The respondents’ career aspirations for the next 10 years were also explored in the survey. Only 36.49% of them indicated that they would continue working in their current field, while 40.54% of them expressed their intention to change their career path. The main reasons for this change were retirement or pursuing entrepreneurship opportunities. The remaining 22.97% of the respondents were either undecided or did not respond to this question.

Conclusion: From the survey data, the career decision making based on interest is significant for undergraduates and those who works for a long time. The other interesting finding is most of the mid-career participants wanted to change career, but they had not considered career consultation with various reasons given in the survey.

2 Project Objective

We present a simple-to-use system that can offer personalised career recommendations to people from various walks of life who are starting their career journey or seeking a career transition with the focus on their interest as main motivator.

The project aims to give individuals in all socio-economic groups an accessible, interactive, and alternative medium for initial career assessment. We intend to help individuals from college students to fresh graduates to those wishing a career switch, in determining which career path to take based on their interests or domain expertise and have them matched to the appropriate courses and educational programs to take to equip them through their upskilling or reskilling journey.

The system makes decisions by considering users' profile of career interests, preferred work conditions, experience, personal values, preferred salary and other related factors. This system leverages an existing knowledge base built in the field of occupational psychology as well as current data on the job market. In addition, the system will also recommend related educational programs so that users can decide where to further their education to meet the requirements of a chosen career path.

2.1 Target Audience

The target audience of this project consists of students who are either studying at the university or have recently graduated and are looking for a suitable career path or job opportunity. Another potential group of users are mid-career professionals who want to change their career direction and explore new possibilities. However, they should be aware that switching to a different career path may not guarantee their expected salary range and may require them to acquire new skills based on the recommendations of the tool.

3 System's Features

3.1 System Intelligence

The career path recommender system is an intelligent system that infers users' preferences based on their input answers to the questions provided by the chatbot. It considers a range of factors such as career interests, remuneration, work experience and education, and work context.

- Career interest is determined by personalized descriptions of an ideal job
- Remuneration is determined by fresh-graduate salary and job market salary.
- Work experience and education are determined by job descriptions, work tasks, and education.
- Work context is determined by work hours, workplace environment, workplace hazards, and social environment.

3.2 Ease of Access and Scalability

The CareerWise Chatbot is built on a web-based application that is seamlessly deployed and accessible to all users with an internet connection. The system can be accessed using any device such as personal computers, tablets, or mobile phones.

3.3 User Management

To ensure that the recommendation is tailored specifically to the intended user and its preference. A login account is needed to be created prior to use of the system. We have used Vue Application to for our frontend which features the login page.

4 Knowledge Modelling

Knowledge modelling involves identifying the concepts and relationships that exist within a domain and representing them in a way that can be used by computer systems.

The goal of knowledge modelling is to create a shared understanding of the concepts and relationships within a domain so that they can be used to support decision-making, problem-solving, and other activities.

Three main stages of knowledge modelling (Schreiber, et al, 2001)

- i. Knowledge Identification
- ii. Knowledge Specification
- iii. Knowledge Refinement

4.1 Knowledge Identification

Knowledge identification lays the foundation for the subsequent stage of knowledge management, which involves knowledge specification. During this stage, relevant information sources are identified and deemed useful in preparation for knowledge acquisition.

There are several sources from which we derive knowledge about the relevant factors required for career advice. The team has identified 2 main sources for knowledge elicitation. These are:

1. Career service expert
2. O*NET database

Because O*NET, which provides the most comprehensive occupational analysis datasets, derives its data from the US job market, there may be concerns that it may deviate from the local Singapore context due to economic, social, and cultural conditions. Thus, we also mined datasets from local sources, to gain insights on factors of interest in the local Singaporean context, such as salaries and popular relevant educational programs. These sources include:

1. Data.gov.sg, the one-stop portal for public data in Singapore, for Singapore graduate tracer study data
2. MyCareersFuture, a government web portal for job postings

The table below shows the complete summary of sources used in the project.

No.	Source of Information	Insights from Information Source	Knowledge Acquisition Technique
1	Career Service Expert	The Subject Matter Expert (SME) with years of experience will be able to provide insight on: <ul style="list-style-type: none">- Techniques which could be used to better provide a career path	Manual Elicitation by acquisition of tacit knowledge from a SME

No.	Source of Information	Insights from Information Source	Knowledge Acquisition Technique
		<p>recommendation to an individual</p> <ul style="list-style-type: none"> - Focusing the scope and intended audience 	
2	O* NET Database	<p>Occupational Information Network (O*NET) is United States primary source of occupational information. It could provide information relative to the project such</p> <ul style="list-style-type: none"> - Interest and Occupation - Job Titles - Level of relatedness between jobs - Job details, such as work context, required level of education or experience, work values, career interest, work tasks and activities 	Knowledge Elicitation through documented knowledge from surveys conducted by occupational psychologists and related social science experts.
3	MyCareersFuture	<p>It's designed to provide the most updated and reliable information about</p> <ul style="list-style-type: none"> - Job availability here in Singapore - Job description - Job requirements - Compensation details 	Web scraping to obtain publicly available data.
4	Singapore Graduate Data	<p>It has a data set that provides information on Singapore base universities, possible compensation, degrees and equivalent overall employment rate</p>	Downloaded from data.gov.sg
5	Generic Population	<p>To validate and support the claims from the Subject Matter Expert with data</p>	Elicitation of tacit knowledge through analysing results of a survey from the general population

Table 4-1 Knowledge Source and Acquisition Technique

4.2 Career Service Expert

Interviews were conducted with a career counsellor from NUS-ISS Career Services. The aim is to identify the crucial factors that are taken into consideration for career advice, and ways of judging the effectiveness of a career service. The questions posed during the interview are as follows:

- What is the process the student goes through in a career consultation?
- What are the questions you usually ask?
- Do you have a metric for how accurate a job recommendation is?
- What are the challenges / problems in recommending a career path?
- What challenges do you see students / clients often face when choosing a career?
- What theories (RIASEC, Super's) are most applicable in the Singapore context?

To summarize the findings of the interview, we found that recommending a career is multifaceted involving not just career interests of the client, but also the expected remuneration, job expectations and the nature of the work, which may encourage or discourage a client from undertaking a career path.

Secondly, we also concluded from the interview that career decision-making is a more complicated process for working professionals wanting to expand within their existing field of expertise, as such decisions require high levels of specificity; as compared to fresh graduates or those desiring a career switch, who are starting with a blank slate. We decided to follow the career counsellor's advice and pursue a blank slate career recommendation as it is more practical to obtain the data to make higher quality recommendations for this target audience.

The career counsellor also gave us some caveats of using the RIASEC model which posits that individuals' career interests. She elaborated that one's career interest may take a back seat when other practical factors are taken into consideration such as salary and work-life balance.

Third, the career counsellor also said that it was difficult to judge whether a career recommendation was satisfactory or not. This differs widely on a case-by-case basis. Clients may end up not choosing a career because of practical concerns, such as the new career path not earning enough to support the family, or the unwillingness to give up their current, more lucrative position after thorough consideration.

4.3 O* NET Database

The O*NET database consists of data about jobs collected by occupational psychologists containing detailed profiles of about 1110 occupations. The surveys are conducted on professional occupational analysts or incumbents working in the specific jobs being surveyed in the United States. Each occupation is identified by an O*NET Standard Occupational Classification (SOC) Code. Most of the listed occupations have information on relevant career interests, work values, work contexts, work tasks and work activities. The O*NET database was mined to provide us with information on the correlations between jobs.

We accessed the database via the O*NET API, providing us with the required datasets to make our career recommendations.

Domain	Description	Source file
Occupation data	Job titles, job description and O*NET SOC codes	https://www.onetcenter.org/dl_files/database/db_27_2_text/Occupation%20Data.txt
Job zones	Zones describe the level of education and preparation required, from high school (Zone 1) to post-graduate (Zone 5)	https://www.onetcenter.org/dl_files/database/db_27_2_text/Job%20Zones.txt
Education, Training and Work Experience	Shows the distribution of levels of education for each occupation	https://www.onetcenter.org/dl_files/database/db_27_2_text/Education%2C%20Training%2C%20and%20Experience.txt
Interests	Career interests according to RIASEC Model, a career interest personality theory by John Holland	https://www.onetcenter.org/dl_files/database/db_27_2_text/Interests.txt
Work values	Work values such as independence, achievement, etc.	https://www.onetcenter.org/dl_files/database/db_27_2_text/Work%20Values.txt
Work styles	Work styles such as independence, social orientation, achievement orientation, leadership qualities	https://www.onetcenter.org/dl_files/database/db_27_2_text/Work%20Styles.txt
Work activities	Activities involved in a job	https://www.onetcenter.org/dl_files/database/db_27_2_text/Work%20Activities.txt
Work tasks	Text statements on the tasks required for each occupation	https://www.onetcenter.org/dl_files/database/db_27_2_text/Task%20Statements.txt
Work context	Work environment, such as work hours, occupational hazards, social contact, and others	https://www.onetcenter.org/dl_files/database/db_27_2_text/Work%20Context.txt
Career pathways	Classification of related jobs into career pathways	https://www.onetonline.org/find/career/All_Career_Clusters.csv?c=0&fmt=csv

*Table 4-2 O*NET Occupational Profile Domains Dataset*

4.4 Knowledge Specification

4.4.1 Occupational Analysis

Data mining was performed on each of the datasets. The data was downloaded from O*NET. O*NET provides the dataset in the schema shown below, where each row represents a category of a variable, for example the first row in the sample below is Required Level of Education - Category 1.0.

	O*NET-SOC Code	Element ID	Element Name	Scale ID	Category	Data Value	N	Standard Error	Lower CI Bound	Upper CI Bound	Recommend	Suppress	Date	Domain	Source
0	11-1011.00	2.D.1	Required Level of Education	RL	1.0	0.00	27	0.00	NaN	NaN	N	07/2014	Incumbent		
1	11-1011.00	2.D.1	Required Level of Education	RL	2.0	0.00	27	0.00	NaN	NaN	N	07/2014	Incumbent		
2	11-1011.00	2.D.1	Required Level of Education	RL	3.0	0.00	27	0.00	NaN	NaN	N	07/2014	Incumbent		
3	11-1011.00	2.D.1	Required Level of Education	RL	4.0	6.05	27	4.43	1.28	24.19	N	07/2014	Incumbent		
4	11-1011.00	2.D.1	Required Level of Education	RL	5.0	4.23	27	4.27	0.50	27.80	N	07/2014	Incumbent		

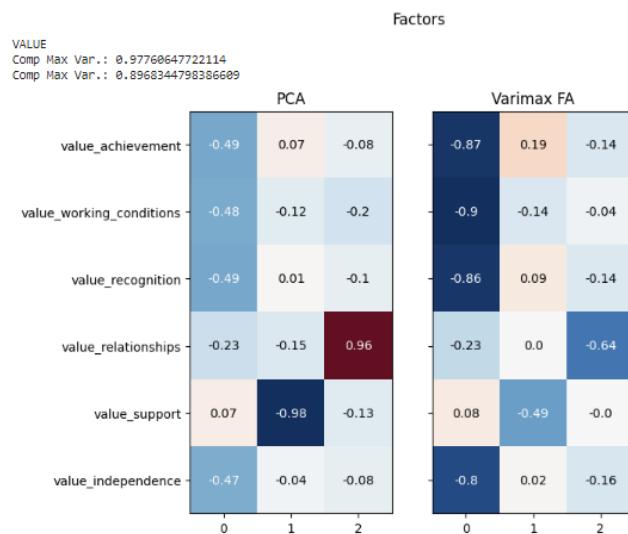
Figure 4-1 Example of raw data for Education, Training and Experience from O*NET

To make the data suited for analysis, we transformed the data in each dataset in the methods below. The full implementation can be viewed in the Jupyter Notebook in the project folder ('/SystemCode/data_processing/1a_gather_occupation_data.ipynb').

1. Group the rows by O*NET SOC
2. Convert the category rows into columns, e.g. the rows for Required Level of Education Categories 1.0, 2.0, 3.0 through 12.0 (last category) are converted to columns 'required_level_of_education_1', 'required_level_of_education_2', 'required_level_of_education_3', etc. The value of each cell is the corresponding "Data Value", see sample above. In this case it is the percentage of people falling into the respective category
3. The values from each of the datasets from the table of O*NET Occupational Profile Domains above are merged into a master dataset with each row representing an O*NET occupation (identified by O*NET-SOC code) and each column representing the career variables derived from step 2.
4. Data cleaning was performed, whereby occupations that are missing all values for multiple domains are excluded. For those missing only some values in a domain, data imputation via K-Nearest Neighbour was performed.
5. To discover patterns in the data, dimension reduction techniques, namely Principal Component Analysis (PCA) and Factor Analysis (FA) with varimax rotation were carried out on each of the domains. (Note: varimax rotation maximises the variance explained by each factor or dimension, allowing us to discover multidimensional patterns)
6. Cluster analysis was also performed to see if classes exist in the domains which can give us guidance on the recommendations.

O*NET-SOC Code	Title	Description	Zone	Date	Source	required_level_of_education_1	required_level_of_education_2	required_level_of_education_3	required_level_of_education_4	required_level_of_education_5	...	context_work_schedules
11-1011.00	Chief Executives	Determine and formulate policies and provide direction...	5.0	07/2014	Analyst	0.00	0.00	0.0	6.05	4.23	...	1.30
11-1011.03	Sustainability Officers	Communicate and coordinate with management, staff...	5.0	08/2021	Analyst	0.00	0.00	0.0	0.00	0.00	...	1.08
11-1021.00	General and Operations Managers	Plan, direct, or coordinate the operations of...	4.0	07/2015	Analyst	0.00	15.20	9.0	19.18	8.59	...	1.15
11-1031.00	Legislators	Develop, introduce, or enact laws and statutes...	4.0	06/2008	Analyst	NaN	NaN	NaN	NaN	NaN	...	NaN
11-2011.00	Advertising and Promotions Managers	Plan, direct, or coordinate advertising policy...	4.0	08/2018	Analyst	6.16	9.82	0.0	7.67	8.04	...	1.20
...
55-3014.00	Artillery and Missile Crew Members	Target, fire, and maintain weapons used to destroy...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
55-3015.00	Command and Control Center Specialists	Operate and monitor communications, detection,...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN

*Figure 4-2 Example of processed O*NET data*



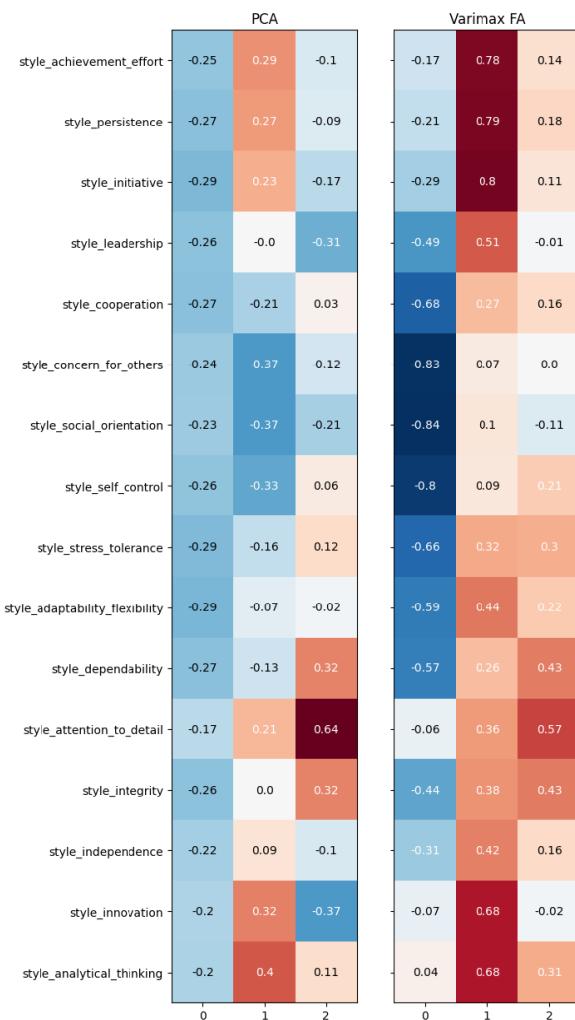


Figure 4-3 PCA and Factor Analysis results for (from top to bottom) Values and Work Styles

From the findings from data mining, we derived several conclusions from the data.

We find that the jobs are collapsable into a more manageable number of dimensions. The analyses were repeated by trying out different maximum numbers of dimensions. Figure 4-3 shows a sample of the analyses carried out on the dataset. For the full results, the reader can refer to the Jupyter notebook file. In the example above, the Work Values achievement, working conditions and recognition loads onto the same dimension, implying a common factor related to “achievement orientedness”. The remaining Work Values variables seem to be independent. For Work Styles, several variables like social orientation, concern for others, self-control, and cooperation load significantly onto the same dimension (shaded dark blue), suggesting a common tendency towards a single “social orientedness” dimension. The other major dimension that stands out are achievement, persistence, and initiative (in red), suggesting a common “achievement orientedness” dimension, similar to the Work Values dimension discussed above. With Work Activities, Work Contexts, and Interests, we repeated the steps above and discovered common factors – the implications of which will be discussed in the following section. For the full results and visualization, see ('/SystemCode/data_processing/1a_gather_occupation_data.ipynb')

4.5 Job Descriptions Dataset

4.5.1 Salaries Dataset

Salaries datasets were obtained by scraping the MyCareersFuture website.

In the present study, the process of data mining is subject to certain limitations, wherein the scope is constrained by a specific dataset acquired under the MyCarreerFuture platform. Specifically, the scope is limited to (I) February 2023, (II) the date and time of the extraction, (III) the amount of data that is available, and (IV) the compensation offered based on the job posting. As such, the system offers a snapshot of job availability and salary ranges, pertaining to a fixed number of postings. Nonetheless, the research team opines that the closed list of approximately 8000 entries is adequate, given the time complexities involved in extracting a colossal amount of data from the site. The implementation of the web scraping can be found in the project Jupyter Notebook file ('SystemCode/data_processing/2a_mcf_web_scrape.ipynb').

Using the BeautifulSoup library, the contents of every currently available page on <https://www.mycareersfuture.gov.sg/> website was scraped. The data was loaded into a Pandas dataframe with the features as shown below, namely company, job title, location, minimum salary, maximum salary, salary type, job description, job requirements, and source URL.

cf_id	company	job_title	location	min_salary	max_salary	salary_type	job_description	job_requirement	url
1	ANIRADUS PTE. LTD.	Web Application Developer #43393	Islandwide	\$3,500	\$6,500	Monthly	Job Description/Industry/ Organization Type... ASP.NET/Discipline/Oral Communication Skills/J...		https://www.mycareersfuture.sg/job/information...
2	ANIRADUS PTE. LTD.	Warehouse Assistant #50208	Islandwide	\$2,200	\$2,600	Monthly	Job Description/Industry/ Organization Type... [Microsoft Office/Microsoft Excel]/Warehouse Ma...		https://www.mycareersfuture.sg/job/others/ware...
3	ELLIOTT MOSS CONSULTING PTE. LTD.	SAP Basis Consultant	HIGH STREET PLAZA, 77 HIGH STREET 178433	\$7,000	\$8,000	Monthly	Job Description/Industry/ Organization Type... [Troubleshooting/Oracle/Unto/Administration/SA...		https://www.mycareersfuture.sg/job/information...
4	ANIRADUS PTE. LTD.	Quality Engineer #51945	Islandwide	\$3,000	\$4,000	Monthly	Job Description/Industry/ Organization Type... [ISO Requirements/Quality Control/Quality Mana...		https://www.mycareersfuture.sg/job/engineering...
5	ANIRADUS PTE. LTD.	Technician #60107	Islandwide	\$1,400	\$1,600	Monthly	Job Description/Industry/ Organization Type... [have discipline/Discipline/discipline checks/...		https://www.mycareersfuture.sg/job/engineering...
6	ANIRADUS PTE. LTD.	Admin Assistant #64017	Islandwide	\$1,800	\$2,500	Monthly	Job Description/Industry/ Organization Type... [Microsoft Office/Microsoft Excel]/Inventory/...		https://www.mycareersfuture.sg/job/admin/admin...
7	JJ CONSULTING SERVICES	Senior Software Engineer (f)1	FAR EAST FINANCE BUILDING, 14 ROBINSON ROAD 04-	\$6,000	\$8,000	Monthly	[Our Client is an established company in Sing... [Design Patterns/provide feedback/Software Eng...		https://www.mycareersfuture.sg/job/information...
8	ANIRADUS PTE. LTD.	Draftsman #62588	Islandwide	\$2,500	\$4,000	Monthly	Job Description/Industry/ Organization Type... [Revit/Microsoft Office/Construction/Adult D...		https://www.mycareersfuture.sg/job/manufacturi...
9	ANIRADUS PTE. LTD.	Sales Executive (Freight Forwarding) #60101	Islandwide	\$2,900	\$3,200	Monthly	Job Description/Industry/ Organization Type... [have discipline/Sales/Discipline/Microsoft Of...		https://www.mycareersfuture.sg/job/sales/sales...
10	ANIRADUS PTE. LTD.	Customer Service Officer (Freight Forwarding) ...	Islandwide	\$2,200	\$2,600	Monthly	Job Description/Industry/ Organization Type... [have discipline/Discipline/Microsoft Officel...		https://www.mycareersfuture.sg/job/customer-se...

Figure 4-4 MyCareersFuture dataset sample

To enable the data to be useful for exploration, we matched the MyCareersFuture dataset to the O*NET dataset. There is no direct lookup table available that matches these two sources of data, therefore we had to employ Natural Language Processing (NLP) procedures to do so. This implementation is found in the project Jupyter Notebook file ('SystemCode/data_processing/2b_onet_job_w2v.ipynb')

Our team chose Word2Vec, a procedure involving neural networks that vectorize a corpus of words into a set of dimensions or vectors, converting them into numerical form, thus enabling pairs of words to be compared geometrically in terms of semantic meaning. The Gensim Python library was used to implement Word2Vec. The 'GoogleNews-vectors-negative300' pretrained vectors loaded to do the string similarity matching. Pairwise matching with Gensim Word2Vec cosine similarity function was used to compare the occupational titles from the O*NET Occupational Dataset (shown previously in Table 1) with the job titles in the MyCareersFuture dataset. Only the highest four matches for each job title were retained in the final results. This yielded a total of 5,056 matches. This gives us a set of current offered salaries data for each O*NET SOC occupation. This will aid in giving job

recommendation with salary ranges to a fixed number of postings (please refer to the project scope for more details).

4.6 Generating the Recommendation Scores

Having gathered and explored the data, we proceeded to organize the data to make them suitable for use in our recommendation system. To organize the process of data refinement, we classify the knowledge acquired into four broad categories:

1. Career Interest
2. Remuneration
3. Work Experience and/or Education
4. Work Context

The interrelationship between the domains is presented in *Figure 4-5 Knowledge Representation*. The flowchart also shows the dependencies for our career recommendation system. Our project uses the content-based recommendation method. This method was chosen over other methods like collaborative filtering because career recommendations typically rely on expert advice on the various factors typically found in the O*NET database. As mentioned in the earlier section on Knowledge Identification, the subject matter expert has encouraged us to adopt a multi-faceted approach to the recommendation, which led us to the factors listed in the *Figure 4-5 Knowledge Representation*.

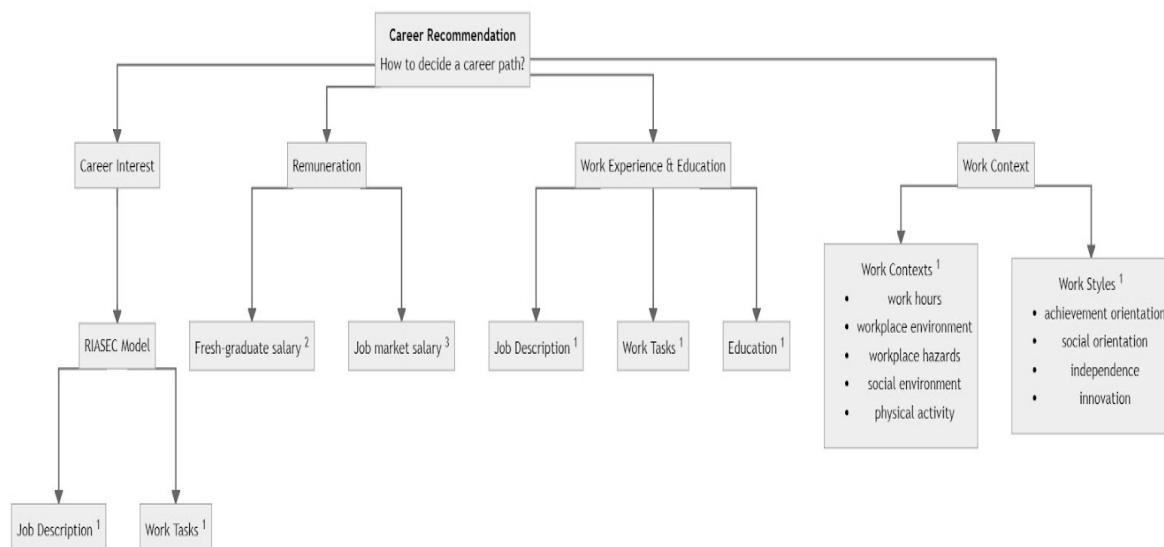


Figure 4-5 Knowledge Representation

4.6.1 Career Interest

There are various ways in which career advice can be generated from one's career interests.

We contemplated whether Career Interest should be represented as dimensions of RIASEC, with each column representing each dimension of the six-letter classification; or as a column of text describing the job description and work tasks. To make this decision, each group

member attempted the RIASEC questionnaire on O*NET's career profiler website called "My Next Move O*NET Interest Profiler" <https://www.mynextmove.org/explore/ip>. After taking the test consisting of 60 questions, each with a 5-point rating scale (from 1 - Strongly Dislike to 5 - Strongly Like), we found that classifying ourselves into only one of six categories and then recommending jobs based on the categorization produced results that we found to be too general. For example, being placed in the Artistic (A) will give you jobs from Interior Designers to Music Therapists. Secondly, the questions in the questionnaire seemed to favour certain categories, for example, our colleagues found that it was easy to get categorized as Artistic (A) and Social (S), even though based on their actual interests and current occupation, they should be classified as Investigative (I).

Therefore we opted to use a text input where users can freely key in a paragraph describing their ideal job, and using text vectorization and cosine similarity to compare the input with each of the job descriptions in our database, we can rank the user's most favoured occupations. This method also gives users a more personalized career recommendation that fits the description they inputted, rather than recommendations based on a broad category.

The text embeddings are obtained using the following methods, as detailed in the Jupyter Notebook

'SystemCode/data_processing/1b_prepare_text_analysis_for_job_descriptions.ipynb'. First, the columns Job Description and Work Tasks from the O*NET occupation database table were concatenated to form a new, complete jobs description column. Hugging Face's Bidirectional Encoder Representations from Transformers (BERT) Sentence Transformer was used to convert each job description into a vector of 768 dimensions. The pretrained model "sentence-transformers/all-mnlp-base-v2" was loaded from Hugging Face via the Sentence Transformer library to obtain the sentence embeddings.

In the final version, the question in the chatbot is an open-ended question. The user keys in a paragraph of his or her ideal or dream job. Content similarity is based on cosine similarity of sentence embeddings between the user's input and each occupation's job description and work tasks.

4.6.2 Remuneration

Remuneration data is derived from the job market salary salary range mined from MyCareersFuture in the previous section. The question in the chatbot is an open-ended question. The user keys in their minimum and ideal expected salary as a fresh graduate. Jobs with minimum salaries that deviate from the user's stated minimum are penalized and will appear at the bottom of the recommendation list.

4.6.3 Work Experience and Education

The question in the chatbot is an open-ended question. The user keys in a paragraph of his or her work experiences and strongest subjects when undertaking their highest educational program. Content similarity is based on cosine similarity of sentence embeddings between the user's input and each occupation's job description and work tasks. The method for accomplishing this is the same as the one explained in the 4.6.1 Career Interest section,

except, the user's description of their experience is used instead of the user's description of their ideal job.

Domain	Variable name	Question
Salary	salary_minimum	What is your minimum expected monthly salary as someone new to a job?
Salary	salary_ideal	What is your ideal monthly salary as someone new to a job?
Education	education_level	What is your current level of education?
Education	education_subject	What subjects or fields are you good at?
Interest	interest_open	Describe in detail what you would like to do for a living.
Experience	experience_open	Describe your past work experience. (If you're studying, you can talk about your extra-curricular activities)
Priority	priority	Which of the following is most important to you when choosing a career path?

Table 4-3 Open-ended questions and priority option

4.6.4 Work Context

We decided to use Multiple Choice Questions to match the user to relevant jobs. The implementation of the web scraping can be found in the project Jupyter Notebook file ('SystemCode/data_processing/1a_gather_occupation_data.ipynb').

Work context is more difficult to determine via text similarity matching as detailed descriptions of work context are not available for each job. However, the O*NET database provides detailed data on the various variables which we had previously processed using tools like PCA and Factor Analysis. Even though the domain is named "Work Context", we decided to include for consideration, the data from multiple facets of the O*NET dataset, i.e., work experience and education work context, work styles and work activities. Because there is significant, patterned overlap in these variables, we are able to identify some common dimensions, as discussed in the Knowledge Specification - Occupational Analysis section.

To make this data useful as for content-based similarity, we downloaded the questionnaires from the O*NET database for preparation of the MCQs. Each variable in the dataset corresponds with at least one question in the O*NET questionnaire booklet, which would sum up to 209 questions, for all the domains included in O*NET Occupational Profile Domains Dataset in Table 2.

This is untenable as a questionnaire for a chatbot, as users will become bored or restless with the number of questions, leading to increased rates of dropping out before finishing. We used the knowledge derived from dimension reduction carried out in the previous section to select the most representative questions. For example, for Work Styles, several variables like social orientation, concern for others, self-control, and cooperation load significantly onto the same dimension, suggesting a common tendency towards a single “social orientedness” dimension. Variables under this common dimension are expected to correlate highly with one another, therefore we can safely select one variable (i.e., question) from the four to represent “social orientedness”; the other three can be safely assumed to evoke very similar responses compared to the representative question.

The 10 representative questions that we picked out, due to high factor loadings in the Factor Analysis, low level of repetitiveness and overlapping meaning, and ease of understanding.

The results

are in the table below. Each MCQ uses a Likert scale of 1 (Lowest) to 5 (Highest) except for the work hours question.

Variable name	Question	Options
style_stress_tolerance	How high is your stress tolerance?	1 (Lowest) to 5 (Highest)
style_achievement_effort	Is it important for you to work hard to achieve a goal?	
style_social_orientation	Is it important for you to be able to form social connections with others?	
style_independence	Is it important for you to be allowed to work independently?	
style_innovation	Is it important for you to be allowed to innovate and try new ways to do a job?	
context_duration_of_typical_work_week	How many hours do you expect to work in a week?	<ul style="list-style-type: none"> - Less than 40 hours - 40 hours - More than 40 hours
context_outdoors_exposed_to_weather	How often are you willing to work outdoors, exposed to the weather?	1- Never 2- A few times every year 3 - A few times every month 4 - A few times every week 5 - Every day
context_spend_time_walking_and_running	How often are you willing to spend time doing physical activities (e.g. running, walking)?	
context_deal_with_unpleasant_or_angry_people	How often are you willing to deal with unpleasant, angry, or discourteous people?	

Variable name	Question	Options
context_exposed_to_hazardous_conditions	How often are you willing to expose yourself to hazardous conditions?	

Table 4-4 Work context questions

4.7 Similarity Score Aggregation

The similarity metrics of the above domains for each occupation were obtained and aggregated using the following:

$$S = W_c X_c + W_i X_i + W_e X_e + W_s X_s$$

Similarity scores:

1. **Career Interest (X_i)** - Cosine similarity of sentence embeddings for user's ideal career input compared with each occupation's job description plus work tasks.
2. **Work Experience (X_e)** - Cosine similarity of sentence embeddings for user's work experience input compared with each occupation's job description plus work tasks.
3. **Work Context (X_c)** - Euclidean distance between the 10 work context variables in Table 5 (a 10-item vector) and that of each occupation
4. **Salary penalty (X_s)** - normalize difference between expected minimum salary and the minimum salary of each occupation; positive values are set to zero, while negative values are used as a penalty score against.

Weighs:

5. **Career Interest weight (W_i)** - User selected. Depending on whether the user prioritises career interest ($W_i = 0.8$) or not ($W_i = 0.4$)
6. **Work Experience weight (W_e)** - User selected. Depending on whether the user prioritises experience ($W_e = 0.8$) or not ($W_e = 0.4$)
7. **Work Context weight (W_c)** - Set to 0.1. Based on experience a high weight on context leads to recommendations that deviate from user's descriptions as jobs in different domains may share similar contexts.
8. **Salary penalty weight (W_s)** - Set to 0.1. This value is adequate for penalization of occupations that fall below the user's expected minimum salary, without deviating from the user's described preferences.

The complete data processing workflow is shown in the *Figure 4-8 Data architecture for recommender system*.

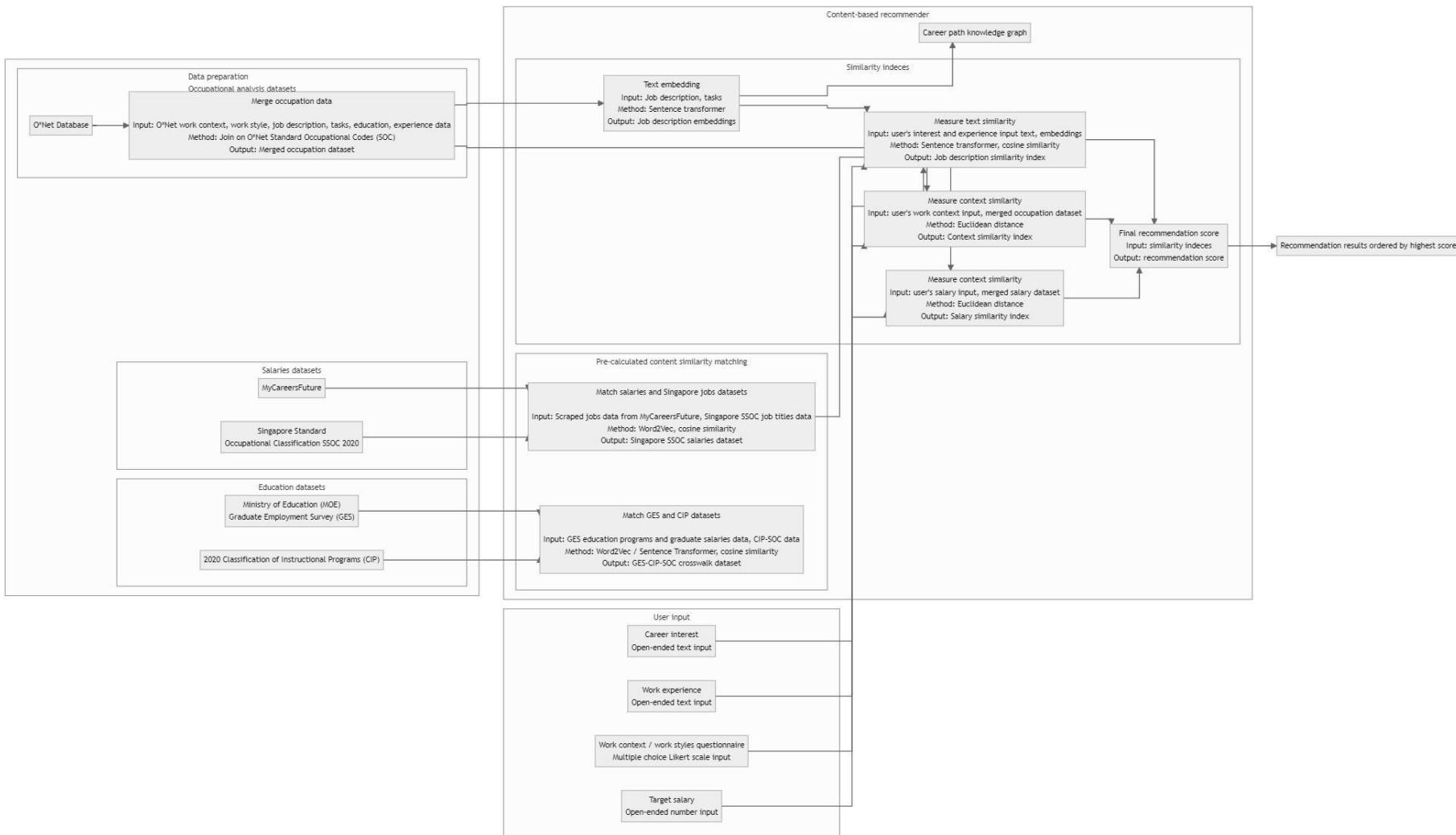


Table 4-5 Data architecture for recommender system

4.8 Additional Recommendations

To provide a value-added service to users, for each occupation recommendation, we provide

1. a career path knowledge graph
2. recommended education programs

4.8.1 Career Path Knowledge Graph

The career path knowledge graph is intended to: 1. Introduce users to more jobs related to their current recommendations; and 2. Allow users to map out their career progression from a low-end job (requiring less experience and education) to a high-end job (requiring more experience and education).

We construct the knowledge graph using the two variables, namely 1. the median experience level in a related field, and 2. the career pathway classification of a job,

Career pathway is a classification provided by the National Career Clusters Framework included in the O*NET database. Each occupation is classified into one of the Career Pathways. The Career Pathways are presented in the table below:

Career Cluster	Career Pathways
Agriculture, Food and Natural Resources	Agribusiness Systems, Animal Systems, Environmental Service Systems, Food Products and Processing Systems, Natural Resources Systems, Plant Systems
Architecture and Construction	Construction, Design/Pre-Construction, Maintenance/Operations
Arts, Audio/Video Technology and Communications	Audio and Video Technology and Film, Journalism and Broadcasting, Performing Arts, Printing Technology, Telecommunications, Visual Arts
Business Management and Administration	Administrative Support, Business Information Management, General Management, Human Resources Management, Operations Management
Education and Training	Administration and Administrative Support, Professional Support Services, Teaching/Training
Finance	Accounting, Banking Services, Business Finance, Insurance Services

Career Cluster	Career Pathways
Government and Public Administration	Foreign Service, Governance, National Security, Planning, Public Management and Administration
Health Science	Biotechnology Research and Development, Diagnostic Services, Health Informatics, Support Services
Hospitality and Tourism	Lodging, Recreation, Amusements, and Attractions, Restaurants and Food/Beverage Services
Human Services	Consumer Services, Counselling and Mental Health Services, Early Childhood Development and Services, Family and Community Services
Information Technology	Information Support and Services, Network Systems, Programming and Software Development, Web and Digital Communications
Law, Public Safety, Corrections, and Security	Correction Services, Emergency and Fire Management Services, Law Enforcement Services, Legal Services, Security and Protective Services
Manufacturing	Maintenance, Installation, and Repair; Manufacturing Production Process Development; Production
Marketing Sales and Service	Marketing Communications; Marketing Management; Merchandising; Professional Sales; Service Marketing
Science Technology Engineering Mathematics (STEM)	Engineering & Technology; Science & Mathematics
Transportation Distribution and Logistics	Facility & Mobile Equipment Maintenance; Health Safety & Environmental Management; Logistics Planning & Management Services; Sales & Service

Table 4-6 Career pathways in O*NET

The nodes that each edge connects are the occupations.

- Source node, is a lower-end occupation classified under the same Career Pathway by O*NET
- Target node, is a higher-end occupation classified under the same Career Pathway by O*NET
- Edge, specifies the “upgrades to” relationship, which is from lower-end of the Career Pathway to higher.

We calculated the item-item similarity matrix, i.e., the similarity matrix between each occupation’s job description concatenated with work tasks, and that of the other occupations in the list.

For each edge, only occupations having the same Career Pathway and having cosine similarities above 0.6 are selected as node items.

The end result is a knowledge graph of occupations related by the “upgrades to” relationship.

4.8.2 Education Programs

We also provide educational program recommendations via a content-based recommendation. This full implementation is found in the project Jupyter Notebook file ('SystemCode/data_processing/3b_merge_sg_course_cip_to_onet.ipynb').

The education programs are obtained by first matching the O*NET jobs to Certified International Property Specialist (CIPS) certifications data using the O*NET SOC-CIPS crosswalk (Source: <https://nces.ed.gov/ipeds/cipcode/post3.aspx?y=56>). The CIPS data comes from the US National Center of Educational Statistics, which classifies and lists the educational programs and certifications provided in the US and their corresponding O*NET occupations.

To enable this data to be applicable to the Singaporean context we matched each CIPS program with the names of educational programs provided in Singapore in the tracer study obtained from data.gov.sg. BERT Sentence Transformer was used to extract the vector embeddings of each entry and cosine similarity between the CIPS and Singapore educational program names were calculated. Pairs with cosine similarity above 0.6 were selected as recommendation items and joined with the occupation dataset we obtained in the previous steps.

4.9 Knowledge Refinement

In this step we have conducted model validation and verification by running simulation test cases in the system (refer to the [Appendix B](#) for the sample test scripts). We have validated the results with the Subject Matter Expert (SME) who can advise for potential areas of improvement which can be seen on [Section 9: Improvements](#) of the documentation.

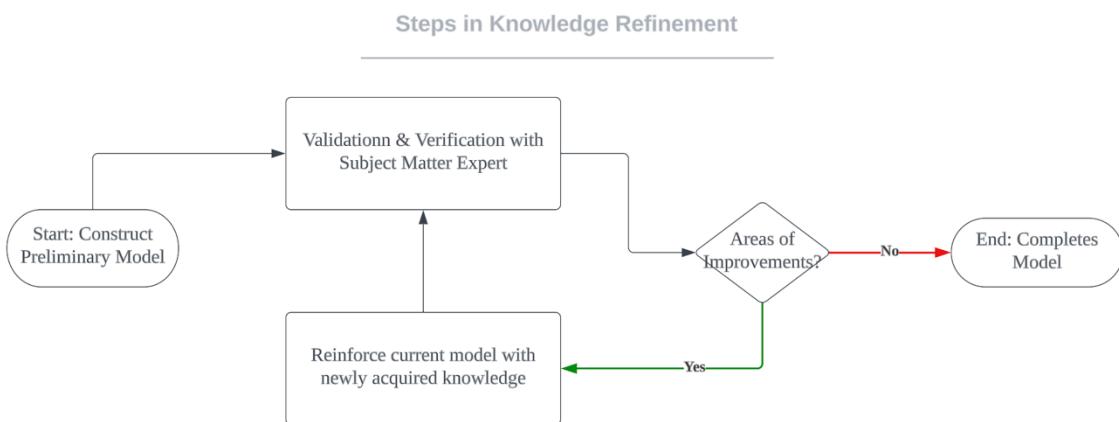


Figure 4-6 Steps taken in knowledge refinement

5 Solution and System Architecture

The Chatbot product is built using Vue 3 framework on the frontend and Flask API on the backend. PostgreSQL was chosen for the database. The user interface is kept as minimal and as simple to use as possible. A summary of the process flow is as follows:

1. First, the user registers for an account.
2. They are brought into the chat window. The chatbot greets the user and briefly introduces the service.
3. User answers the questions posed by the chatbot. When all questions have been answered, a button appears which, when pressed, generates the recommendations.
4. Recommendations are displayed to the user in an index.
5. The user can select a recommended occupation in the index to reveal more details about the occupation. Each occupation details page contains information about the job tasks, a Sankey diagram showing the network of related jobs, job salaries, and recommended educational programs.

Each step will be covered in detail below.

5.1 User Registration / Login

Users are brought into the login page. They can log in with their email and password. They can select a button to “Register for an Account” if they have not created an account. If they chose to register, they must key in their username, email and password and submit the form.

The screenshot shows a 'Sign Up' form with three input fields: 'User name', 'Email', and 'Password'. Each field has a red asterisk indicating it is required. The 'Password' field includes a small eye icon for password visibility. Below the fields is a large blue 'Sign Up' button. At the bottom of the form, there is a link 'Have an account? Log in now'.

Figure 5-1 Screenshot of login/register page

5.1.1 Technical Implementation

Vue 3 Frontend. The AuthPage.vue component contains the account login and signup forms – LoginForm.vue and SignupForm.vue. LoginForm.vue component has inputs which take the values, email and password, and sends them to the API POST endpoint, `/api/login`. SignupForm.vue component has inputs which take the values username, email and password, and sends them to the API POST endpoint, `/api/register`. Both endpoints return a JSON web token (JWT) which will be stored in the browser local storage. For subsequent API calls, bearer authentication method is used to authorize the user to view the data, in other words, the JWT token string will be appended to the request header field, “Authorization” before being sent to the backend API for user authentication.

Database. The user accounts are handled by the `users` table.

- id – Primary key
- username – User's unique username
- email – User's unique email address
- password – User's password
- last_chat_session_id – User's last visited chat session

Flask API. The user management module has the following API endpoints.

- LoginApi, POST '/api/login' - Logs the user in and receives a JWT token
- LogoutApi, POST '/api/logout' - Logs the user out and removes the JWT token
- RegisterApi, POST '/api/register' - Creates the user account and receives a JWT token
- UserApi, GET '/api/user' - Gets the user information

5.2 Chatbot Window

The user is directed to the chatbot window where they are greeted by the chatbot and briefly introduced to the service. The sidebar on the left shows the user's chat history. The input section at the bottom allows the user to input their replies to the chatbot.

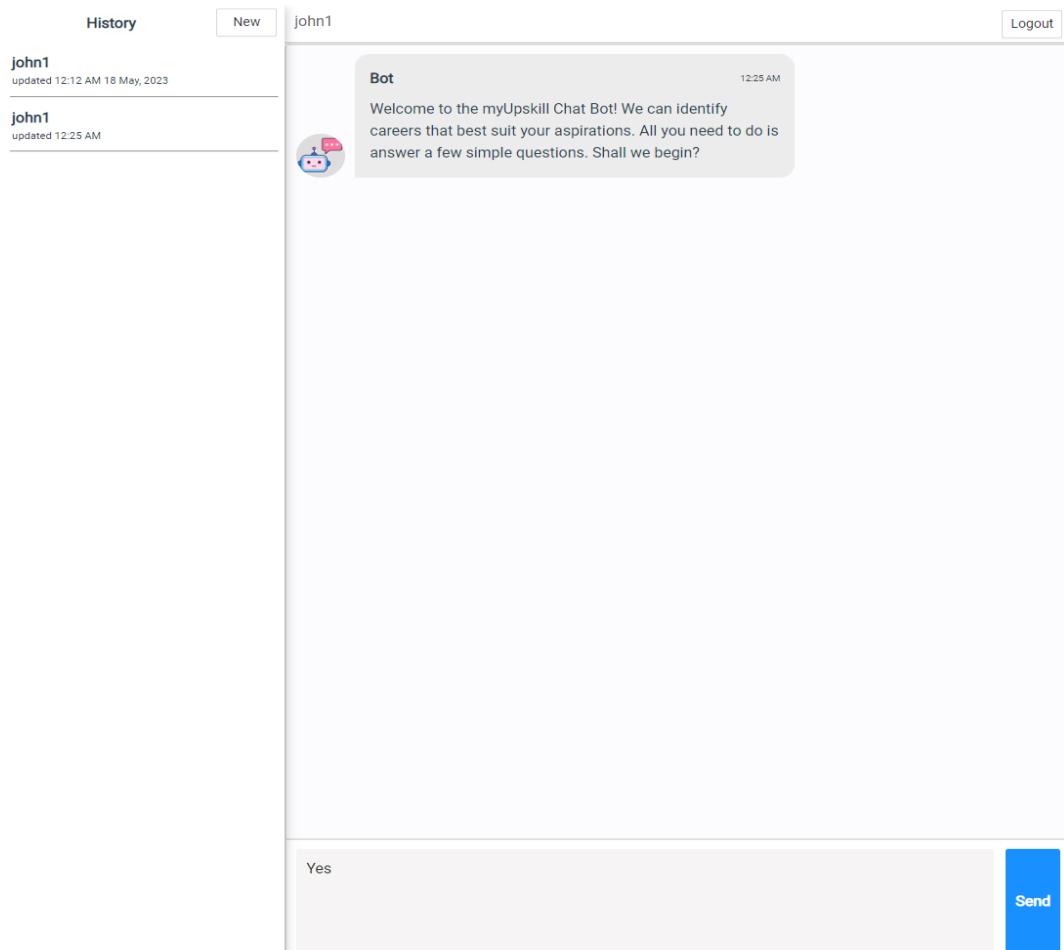


Figure 5-2 Screenshot of chatbot window at the beginning of the session. The sidebar on the left shows the user's chat history.

5.2.1 Technical Implementation

Vue 3 Frontend. The ChatWindow.vue component contains the sidebar chat sessions history component, ChatList.vue. Chat messages, ChatMessage.vue, are shown in a similar view to popular apps like Whatsapp and Telegram. ChatList.vue component is populated by all the fetched chat sessions via the API GET endpoint, '/api/chat_sessions'. The main section of the chat window is populated with information from the current active chat session (fetched via '/api/chat_sessions/<int:id>'). The API call also returns all the child chat messages which are displayed via the ChatMessage.vue component.

Database. The chat sessions are handled by the `chat_sessions` table.

- id – Primary key
- user_id – foreign key referring to the user that owns the chat session
- name – name of chat session
- end_time – timestamp of when the chat has ended

- status – status of the session, can be either “start”, “in progress”, “ready” (all questions answered and ready to generate results), “completed” (results generated), “cancelled” (session disabled)

Each chat session has children from `chats` table, which are the individual chat messages.

- id – Primary key
- user_id – foreign key referring to the user that owns the chat message
- chat_session_id - foreign key referring to the parent chat session
- message_text – the chat message text
- created_by – indicates whether the chat message was generated by the user or by the chatbot
- question_id – the current question that the user is answering with this chat message
- option_id – the ID of the option selected by the user (for MCQs only)

Flask API. The user management module has the following API endpoints.

- ChatSessionsApi, GET '/api/chat_sessions' - show all chat sessions by the user
- ChatSessionsApi, POST '/api/chat_sessions' - creates a new chat session
- ChatSessionApi, GET '/api/chat_sessions/int:id' - get a chat session, along with its child chat messages

5.3 Question-answering session

Once the user types in their response to the welcome message. The chatbot begins asking the user a series of questions (the questions were formulated in the section on Knowledge Specification and Refinement).

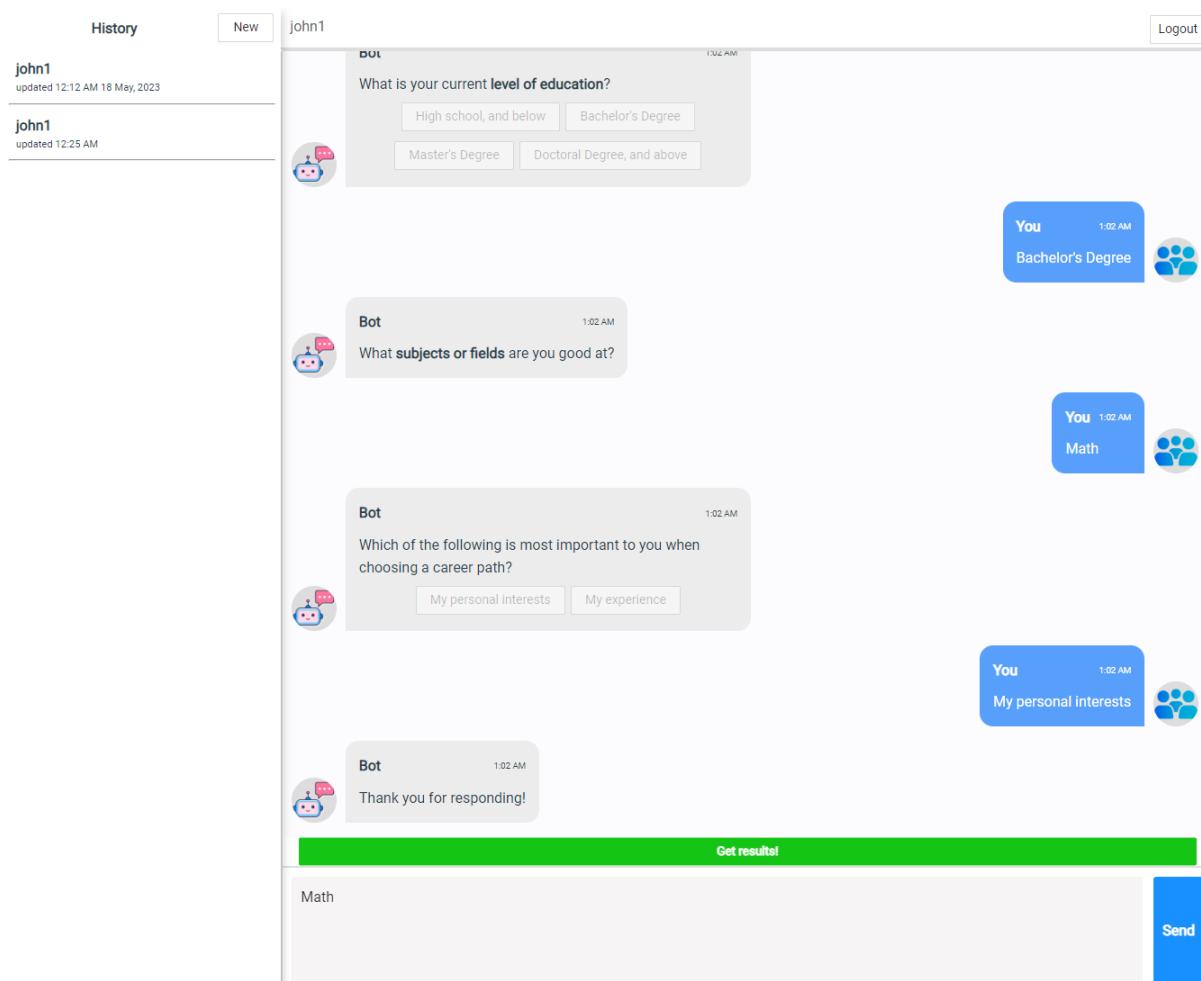


Figure 5-3 Screenshot of chatbot window at the end of the session. A flashing green block button appears above the input section, prompting the user to click to generate results.

5.3.1 Technical Implementation

Vue 3 Frontend. In the ChatWindow.vue component, users type in their response and when they click “Send”, it is submitted to the POST `/api/chats` endpoint. In the backend, the logic is handled as shown in the question and answer chatbot process flow *Figure 5-4 question and answer chatbot process flow* below. The response chat from the Bot contains the question, chosen from the questions table in the database. MCQ questions have child options objects, which will be displayed in the ChatMessage.vue component for the user to select.

Database. The questions are stored in the `questions` table, as below:

- id – Primary key
- order – order in which the questions will be shown in ascending order
- group – question group, specifies the domain in which the question belong, e.g.: Work Context, Salary, Interest, etc.
- code – question’s unique code, corresponds with the O*NET variables in the dataset
- text – Contents of the question

- min_response_length – for open-ended questions, minimum length of response

Each MCQ question has many 'options' children, outlined below.

- id – Primary key
- order – order in which the options will be shown in ascending order
- value – value of the option, e.g. 1 to 5 for Likert scale items
- data_type – data type, i.e. text or number or boolean
- label – text to be shown to the user for this option, e.g. "Very Important", "Somewhat important"
- question_id – foreign key referencing the parent question

Flask API. The user management module has the following API endpoints.

- ChatsApi, POST '/api/chats' - submit a chat message, and subsequently generates a Bot response message (usually in the form of a question)

Figure 5-4 question and answer chatbot process flow depicts a process flow for a survey chatbot interface. The chatbot collects user input on user preferences and then displays a question. The response type can be either open-ended or multiple-choice. If the response type is open-ended, the user inputs text, and the chatbot checks the length of the response. If the response is adequate, the chatbot saves the response. If the response is inadequate, the chatbot prompts the user for more information, and the user inputs more text. If the response type is multiple-choice, the user selects an option by clicking one of the option buttons, and the chatbot saves the response. The user can also key in their responses in the text input and click "Send", the backend automatically selects the answer from the question options with the closest Levenshtein distance. The user experience process flow restarts by selecting the next question in the list after the chatbot saves the response.

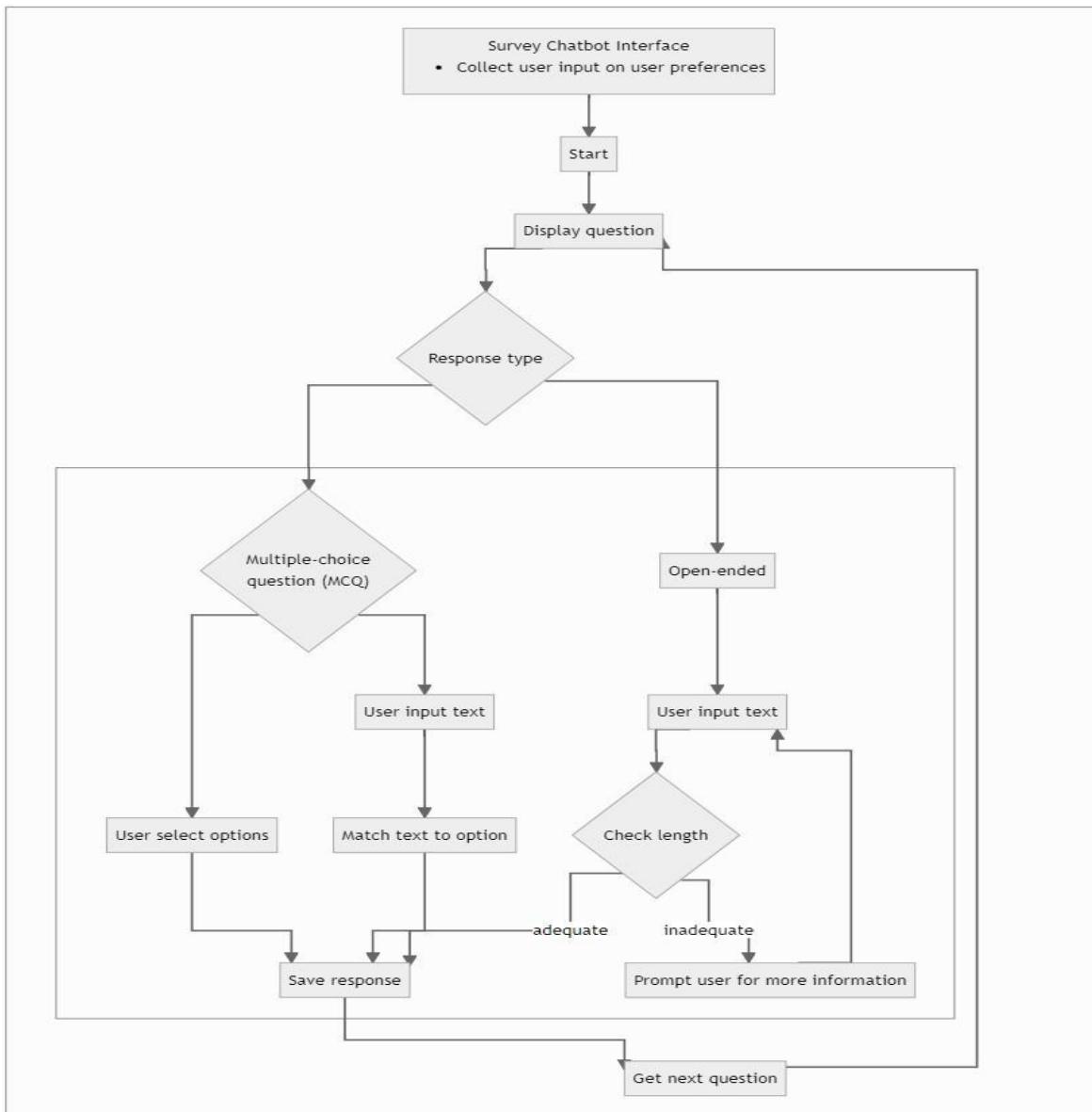


Figure 5-4 question and answer chatbot process flow

After all questions have been answered, the system now has sufficient user profile data to generate a recommendation. A flashing green block button labelled “Get Results!” appears above the input section, prompting the user to click it to generate results.

5.4 Recommendations Results Index

The backend calculates the similarity score for each O*NET occupation in the dataset obtained in the previous section. The occupations are sorted by similarity scores in descending order, where the highest signifies greatest match with the user’s profile.

A list of occupations is shown to the user. Each occupation is an O*NET occupation.

Careers

Personal Financial Advisors

O*Net SOC 13-2052.00

Advise clients on financial plans using knowledge of tax and investment strategies, securities, insurance, pension plans, and real estate. Duties include assessing clients' assets, liabilities, cash flow, insurance coverage, tax status, and financial objectives. May also buy and sell financial assets for clients.

[More](#)

Securities, Commodities, and Financial Services Sales Agents

O*Net SOC 41-3031.00

Buy and sell securities or commodities in investment and trading firms, or provide financial services to businesses and individuals. May advise customers about stocks, bonds, mutual funds, commodities, and market conditions.

[More](#)

Investment Fund Managers

O*Net SOC 11-3031.03

Plan, direct, or coordinate investment strategy or operations for a large pool of liquid assets supplied by institutional investors or individual investors.

[More](#)

Financial and Investment Analysts

O*Net SOC 13-2051.00

Conduct quantitative analyses of information involving investment programs or financial data of public or private institutions, including valuation of businesses.

[More](#)

Financial Quantitative Analysts

O*Net SOC 13-2099.01

Develop quantitative techniques to inform securities investing, equities investing, pricing, or valuation of financial instruments. Develop mathematical or statistical models for risk management, asset optimization, pricing, or relative value analysis.

[More](#)

Financial Risk Specialists

O*Net SOC 13-2054.00

Analyze and measure exposure to credit and market risk threatening the assets, earning capacity, or economic state of an organization. May make recommendations to limit risk.

Figure 5-5 Recommendation results index

5.4.1 Technical Implementation

Vue 3 Frontend. When the user presses the generate result button, the POST API `/api/chat_sessions/int:chat_session_id/results` will be called, which activates the SimilarityCalculator in the backend. The calculator generates a similarity score for each occupation using the formula obtained in Similarity Score Aggregation section in the Knowledge Refinement chapter. The resultant scores are stored in the `results` table and sorted in descending order of similarity. The resultant occupations list is returned. In the ResultsIndex.vue component, these occupations will be shown as an index for users to select.

Database. The results information from the similarity score calculations are stored in the `results` table, as below:

- id – Primary key
- occupation_id – Foreign key referring to an O*NET occupation
- interest_similarity – Cosine similarity of between user interest input and job description + work tasks
- experience_similarity – Cosine similarity of between user experience input and job description + work tasks
- context_similarity – Normalized Euclidean distance between user work context answers (MCQ option values) and the corresponding occupation variables
- salary_similarity – Normalized difference between user's expected minimum salary and occupation's minimum salary
- chat_session_id – Parent chat session
- score – final aggregated similarity score
- user_rating – user rating of how good the match is

The `occupations` table is as follows:

- id – primary key
- title – job title
- description – job description
- job_zone – job zone
- education_mode – Modal O*NET education category
- education_median – Median O*NET education category
- experience_mode – Modal O*NET experience category
- experience_median – Median O*NET experience category
- style_stress_tolerance – Stress tolerance score (part of Work Context)
- style_achievement_effort – Achievement/effort score (part of Work Context)
- style_social_orientation – Social orientation score (part of Work Context)
- style_independence – Independence score (part of Work Context)
- style_innovation – Innovation score (part of Work Context)
- context_duration_of_typical_work_week – duration of work week score (part of Work Context)
- context_outdoors_exposed_to_weather – exposure to weather/outdoors score (part of Work Context)
- context_deal_with_unpleasant_or_angry_people – dealing with angry people score (part of Work Context)
- context_exposed_to_hazardous_conditions – hazardous conditions score (part of Work Context)

- task – work tasks
- min_salary – minimum salary of related jobs in MyCareersFuture
- max_salary – maximum salary of related jobs in MyCareersFuture

Flask API. The user management module has the following API endpoints.

- GenerateResultsApi, GET '/api/chat_sessions/int:chat_session_id/results' - generates the similarity scores and returns the 20 matching occupations

5.5 Result Details Page

The user can select a recommended occupation in the index to learn more details about the occupation. Each occupation details page contains the following sections:

- **Job Descriptions** - O*NET job descriptions and work tasks
- **Job Market** - The job titles and salaries section shows the occupation description and related jobs (from MyCareersFuture) and salary ranges
- **Career Path** - The Sankey diagram shows how this occupation is related to other occupations and the path it takes from a lower level occupation (left) to a higher level occupation (right)
- **Educational Programs** - shows related educational programs and fresh graduate employability and salary trends. Each entry shows the name of the course, e.g. “Bachelor of Accountancy”, the faculty, “Nanyang Technological University, School of Business”; the most current fresh graduate salary and the percentage of graduates hired in full-time employment within the first 6 months of graduation. The trend line on the right shows the fresh graduate salary movement (as far back as 2013, where data is available).

X

Investment Fund Managers

O*Net SOC 11-3031.03

Plan, direct, or coordinate investment strategy or operations for a large pool of liquid assets supplied by institutional investors or individual investors.

> Work Tasks

▼ Job Titles & Salary

Title	ISCO Code	Min. Salary	Max. Salary
Head of Underwriting (Maritime Services) (SFw-Sea Transport) 🔍	1346	\$7,000	\$7,000
Claims Manager (Operations) (SFw-Financial Services) 🔍	1346	\$3,500	\$3,500
Chief Financial Officer (Financial Accounting/Management Accounting) (SFw-Accountancy) 🔍	1211	\$3,000	\$3,000
Director (Project Financing) (SFw-Engineering Services) 🔍	1211	\$3,000	\$3,000
Audit senior manager 🔍	1211	\$3,000	\$3,000

[◀](#) [1](#) [2](#) [3](#) [4](#) [5](#) ... [19](#) [▶](#) [5 / page](#) [▼](#)

Figure 5-6 Result details page shows the occupation description and related jobs (from MyCareersFuture) and salary ranges



Figure 5-7 Sankey diagram shows how this occupation is related to other occupations and the path it takes from a lower level occupation (left) to a higher level occupation (right)

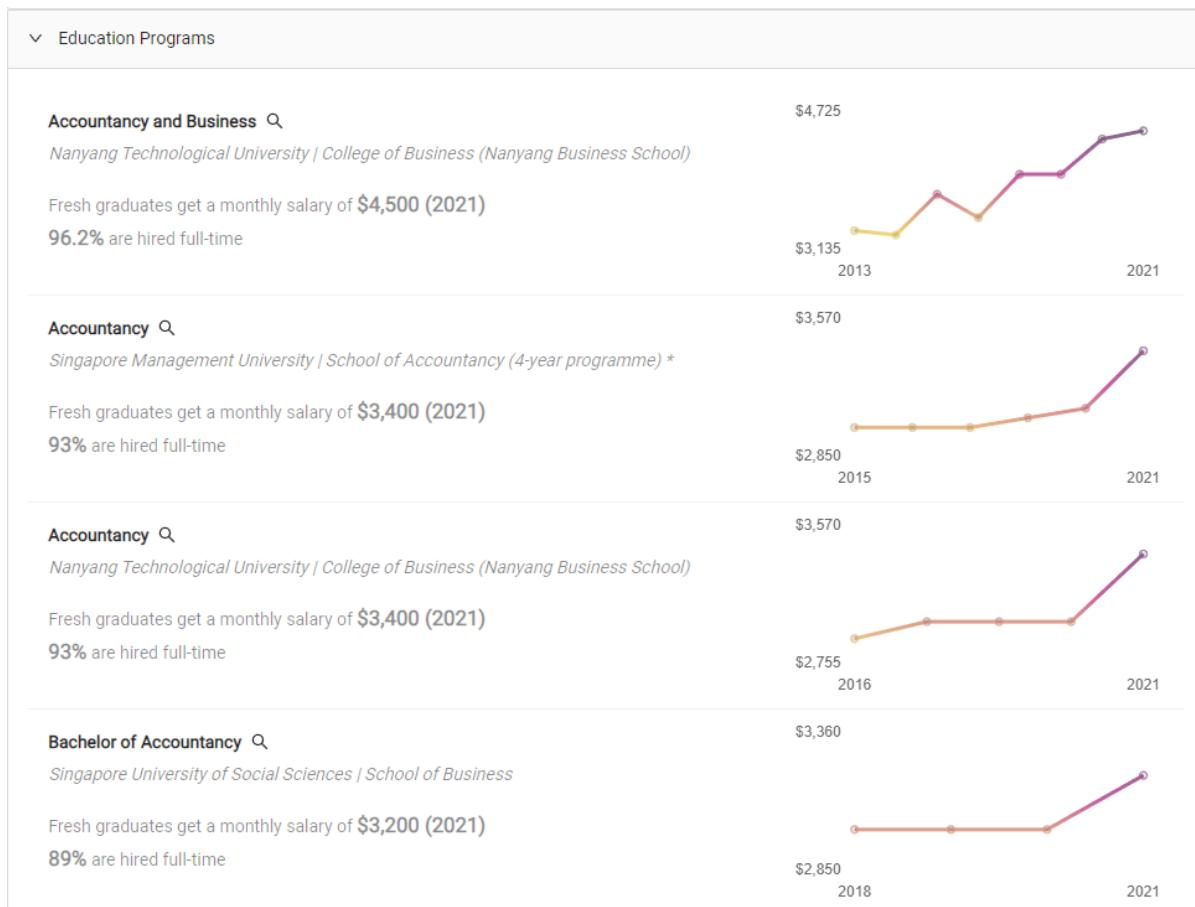


Figure 5-8 Educational Programs and Fresh Graduate Trends

5.5.1 Technical Implementation

Vue 3 Frontend. The ResultDetail.vue component shows the data fetched from the occupations GET API ('/api/occupations/string:id') endpoint.

Database. The occupations table has been discussed in the previous step. We will now describe the children of each occupation, namely the educational programs recommendations table, the career paths knowledge graph, and the job market or salaries information.

The educational programs are stored in the 'programs' table, and each program has many 'program_trends' children which display the fresh graduate salaries and employment rates each year according to the Singapore universities tracer study. The programs are referenced by each occupation via two many-to-many intermediary tables 'cip_occupation' and 'cip_program'.

The 'programs' table:

- id – Primary key
- university – university name

- school – faculty or department name
- degree – degree name

The `program_trends` table:

- id – Primary key
- year – Year of tracer survey
- program_id – Foreign key referencing the program
- employment_rate_ft_perm – fresh graduate full-time employment rate
- gross_monthly_median – fresh graduate median gross monthly income (SGD)

The career path knowledge graph edges are stored in the `career_paths` table. The nodes that each edge connects are the occupations.

- id – Edge primary key
- source_id – source node, is an occupation ID, referring to a job requiring lower level of experience
- target_id – target node, is an occupation ID, referring to a job requiring higher level of experience

The MyCareersFuture dataset is stored in the SSOC Jobs table `ssoc_jobs`

- id – primary key
- occupation_id – foreign key referencing the parent occupation
- ssoc_code – Singapore SOC code of the job (matched using Word2Vec)
- ssoc_job_title – Job title, as shown in MyCareersFuture
- isco_code – International Standard Classification of Occupations code
- min_salary – minimum offered salary
- max_salary – maximum offered salary

Flask API. The user management module has the following API endpoints.

- OccupationApi, GET '/api/occupations/string:id' - displays the current selected occupation details

6 Overall System Architecture

As shown in the previous section, the system consists of 2 main parts: The front-end web interface developed using Vue 3 Javascript Framework and the backend which uses Python Flask, connected to a PostgreSQL database as shown below.

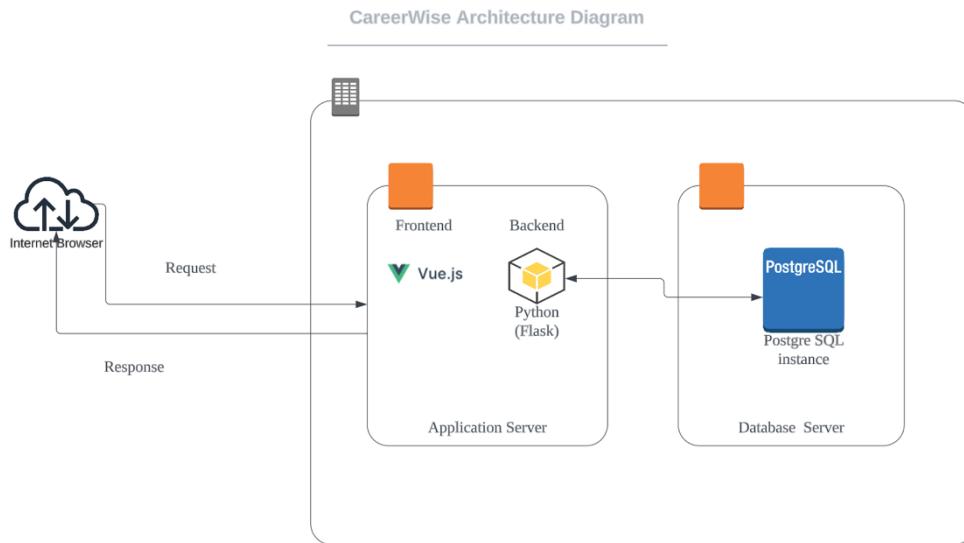


Figure 6-1 Overall System Architecture Diagram

The following Entity Relationship Diagram shows how the data presented in each step, from user registration to occupation recommendation, are related to one another.

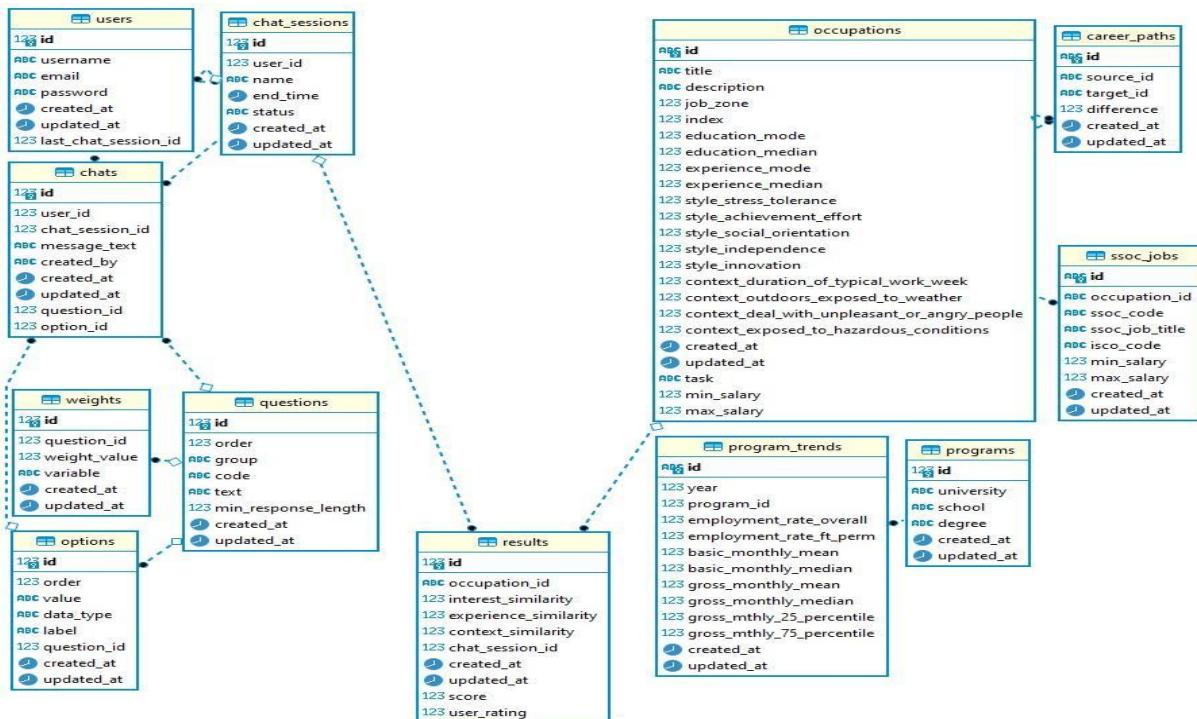


Figure 6-2 PostgreSQL Entity Relationship Diagram

7 Limitations

Although the system has given career path recommendations based on the user preference/interest or expertise, these recommendations can be improved upon with more input from experts in the field. We relied on the O*NET Career Pathways classifications to build the knowledge graph, but the appropriateness of this classification varies across career clusters, with some sectors like scientific fields yielding results that are somewhat unrealistic, e.g. Sociologists and Astronomers are grouped into the same career pathway in O*NET even though they require very different training and preparation.

It is still up to the user if he/she is willing to undergo the needed upskilling or reskilling in relation to the recommended careers. The tool is intended to help user for their initial assessment and result can be used as reference or if further consultation is intended, we can connect to a professional career service practitioner.

8 Conclusion

The team had gained useful skills along the way when creating the whole project. We encountered setbacks and had overcome them. The techniques learned during the Intelligent Reasoning System courses were proven useful such as Cosine Similarity, Word2Vec, Jaccard Similarity and the use of, vectorization, word embedding and sentence transformer were some of the techniques used to match our datasets. Building the system itself presented a whole new set of learning curves, allowing our main developer in the team to apply his practical knowledge on Vue, Python Flask, PostgreSQL and Dockerization.

Due to time, resource and cost constraints, the original plan to deploy the web application in AWS cloud has been postponed for future opportunities for further improvements and enhancements. But the Cloud formation template was made available in the repository.

9 Improvements

User login are being managed at the local PostgreSQL Database at the moment. Integrating the Single Sign On login page to popular technologies like Google, Facebook, Cognito, Auth0 and Azure etc. could help in gaining user traction.

If a user would like to have a further discussion or any clarification about the career path and opt to have a live chat with a career service practitioner this platform can be leveraged to initiate the first point of connection with a career service practitioner and the user. By introducing a live chat feature.

The Subject Matter Expert (SME) has highlighted several potential areas for improvement that could enhance our project:

- Revising the project name to better align with and reflect the overall project's objectives.
- Enhancing user assistance by providing additional information related to the educational program. This could involve sharing details such as the estimated cost of the program or the time commitment required for individuals to upskill or reskill themselves.
- Offering flexible options for educational programs, including full-time, part-time, or online formats, to cater to diverse user preferences and needs.
- Exploring the utilization of a paid subscription model, which could involve providing users with access to live 24/7 interactive chatbot sessions with career practitioners. Alternatively, we could leverage advanced technologies like GPT (Generative Pre-trained Transformer) to deliver interactive responses to users.
- Considering the provision of face-to-face career coaching sessions for individuals willing to invest in personalized services.

These suggested improvements demonstrate our commitment to refining the project by incorporating user-oriented enhancements, expanding information accessibility, and exploring innovative approaches to cater to various user preferences and requirements.

10 Bibliography

MyCareersFuture: <https://www.mycareersfuture.gov.sg/>

Graduate Data: <https://data.gov.sg/dataset/graduate-employment-survey-ntu-nus-sit-smu-suss-sutd>

O' NET

RIASEC Interest & Occupations:

<https://www.onetcenter.org/dictionary/27.2/excel/interests.html>

Job Titles: https://www.onetcenter.org/dictionary/27.2/excel/occupation_data.html

Level of relatedness between

jobs: https://www.onetcenter.org/dictionary/27.2/excel/related_occupations.html

Metadata: https://www.onetcenter.org/dictionary/27.2/excel/content_model_reference.html

Job Description: https://www.onetcenter.org/dictionary/27.2/excel/occupation_data.html

Chatot backend and UI: <https://buffml.com/web-based-chatbot-using-flask-api/>

Flask boilerplate: <https://github.com/realpython/flask-boilerplate>

Chang Ye Han, Chua Zhen Liang Desmond, Chee Jia Wei, Ganesh Kumar, Goh Yu Chen, Jeanette Tan Yi Wen. BTO Recommender System – Project Report Master of Technology NUS-ISS Singapore.

Schreiber, G., Akkermans, H., Anjewierden, A., Wielinga, B. J., Hoog, R.D., Shadbolt, N., & Velde, W. V. (2001). Knowledge Engineering and Management – The CommonKADS Methodology. Cambridge, Massachusetts: The MIT

11 Appendix

APPENDIX A1: Post Interview with Career Service Expert

We had a 2-hour long interview with a NUS Career Service Expert to get her input about our project's proposal objective, and scope of project. She had pointed several valid inputs which helped us recalibrate our intent and scope in the project. After our discussion she had reiterate here points within below email.

RE: Group Project: Career Service Chatbot

↪ ↲ ↢



Lim Meow Kuan Adeline <adeline_lim@nus.edu.sg>

Friday, 17 March 2023 at 2:28 PM

To: Angelie Borromeo; Cc: chuajackyune@gmail.com; kwattivy@gmail.com; +2 more



To protect your privacy, some external images in this message were not downloaded.

[Download external images](#)

[Go to Settings](#)

Hi Angelie and team,

As discussed, the team should identify **what stage of a person's career 'pain point' will your chatbot solve.**

Various chatbots are available to help individuals in different stages of their careers. Some examples are:

1. Mya: Mya is an AI-powered chatbot that assists job seekers in finding suitable job openings and provides them with personalized recommendations. It can also answer various questions related to the job search process.
2. Olivia: Olivia is a chatbot that helps individuals create resumes and cover letters. It can also help job seekers prepare for job interviews by providing tips and answering interview-related questions.
3. MyInterview: MyInterview is an AI-powered chatbot that helps job seekers prepare for job interviews. It provides personalized feedback on the candidate's answers and suggests improvements.
4. JobPal: JobPal is a chatbot that helps individuals to find suitable job openings based on their skills and experience. It also provides personalized career advice and suggestions for improving their resume and cover letter.

Some areas to think about,

- If the idea is to help a fresh graduate to seek for potential job opportunities, what prompts will the chatbot need to map out to identify 'best-fit' job opportunities?
- If the idea is to help a mid-career to identify possible next career, what prompts will the chatbot need to map out to identify 'best-fit' job opportunities? *Think about prior/current experiences / domain expertise + new skills à potential job roles.*
- If the chatbot is to focus more on what an individual request for / wants, will the results be realistic and suitable? i.e. *high salary range, four days work week, job location only in the east.*

The team needs to prioritize if the chatbot is to solve the pain point(s) of an individual jobseeker *OR* that of a career services consultant?

Regards,

Adeline Lim :: Career Services :: NUS-ISS

APPENDIX A2: Survey Results

The anonymous survey result was uploaded in the teams shared drive and accessible through below link.

LINK:

https://docs.google.com/spreadsheets/d/1w6I2Fn7OU9hZo3kVbzNrcaoYQtQqFVm1/edit?usp=share_link&ouid=107036657032408819531&rtpof=true&sd=true

APPENDIX B: Map of System Functionalities against Modules

System Functionalities	Modules
Recommended Career Paths Knowledge Graph Career Chatbot	Machine Reasoning <ul style="list-style-type: none"> - Knowledge graphs - Chatbot
Recommendation of occupations based on user input on ideal job, work experience, work context/values/styles Recommendation of relevant educational programs	Reasoning Systems <ul style="list-style-type: none"> - Content-based recommendation - Measuring similarity: Cosine similarity, Euclidean distance
Text analysis and similarity based matching of job description + work tasks	Cognitive Systems <ul style="list-style-type: none"> - Natural Language Processing

Table 11-1 System Functionalities - Modules Mapping

APPENDIX C: Sample test script used to answer CareerWise Chatbot Open ended question (data below were generated using GPT Generated Text)

Prompt: write a 300 word essay about "My dream job" involving math, science, biology research and medicine

Response: I have always been fascinated by the field of mathematics and its applications in various fields such as physics, engineering, economics, finance, etc. I believe that my strong foundation in these subjects will be valuable when pursuing a career in research and medicine. In addition, my interest in biology has led me to explore the subject further, with a particular focus on genetics and evolution. This knowledge can also be applied in medical research and development. Overall, I am excited at the prospect of combining my interests in math, science, and biology into a single career path.

Prompt: write a 300 word essay about "My dream job" involving modeling, acting, music, and fashion

Response: As a creative individual, I have always had a passion for modeling and acting. The ability to express myself through different mediums like clothing, makeup, and props has allowed me to develop skills in areas outside of academics. For example, I enjoy experimenting with different styles and techniques in photography and videography, which are essential tools for creating high-quality visual content. Additionally, my love for music and dance has inspired me to pursue a degree in performance arts, where I hope to become a professional musician or dancer. Finally, my fascination with fashion design has led me to study fashion merchandising and marketing, enabling me to create unique and innovative products that reflect my personal style. All of these passions combined make me feel fulfilled and motivated to continue learning and growing as an artist.

Prompt: write a 300 word essay about "My dream job" involving accounting, finance and managing client's investment portfolio

Response: Throughout my academic journey, I have developed a deep **understanding** of financial management and accounting principles. My goal is to use this knowledge to manage clients' investment portfolios effectively. As a result, I aspire to work in a role that involves overseeing multiple accounts and making informed decisions based on market trends and economic indicators. Furthermore, I am passionate about developing strategies to maximize returns while minimizing risk. Ultimately, I aim to build long-term relationships with clients and help them achieve their financial goals.

Prompt: write a 300 word essay about "My dream job" involving athletics and sports, namely basketball

Response: Basketball has always held a special place in my heart, and it continues to inspire me today. Growing up playing the sport, I learned early on the importance of teamwork, communication, and hard work. These values are crucial not only in basketball but also in life. Basketball taught me how to work together towards a common goal, communicate effectively with teammates, and adapt to changing situations. It was during my time in college that I discovered my true passion for the game. From then on, I dedicated myself to improving my skills and honing my craft. Today, I strive to embody those same qualities in all aspects of my life, including my future endeavors.

Prompt: write a 300 word essay about "My dream job" involving athletics and sports, namely ping pong, basketball and judo

Response: Growing up, I was always drawn to sports and physical activities. Ping pong became one of my favorite games due to its fast pace and strategic elements. Playing against friends helped me improve my coordination and reflexes. Basketball offered me a chance to hone my shooting skills and dribbling abilities. Judo provided me with discipline, strength, and mental toughness. Each sport taught me important lessons about perseverance, dedication, and self-improvement. Through each experience, I gained confidence in my own abilities and potential. With each new challenge, I grew stronger and more skilled. Now, I look forward to using my newfound skills and experiences to pursue my dreams and live out my passion for sports and physical activity.

Prompt: write a 300 word essay about "My dream job" involving competing with others, building a successful company and leadership

Response: Competing with others is something I've always enjoyed doing. Whether it's in school competitions or friendly matches, I thrive on pushing myself to excel beyond what I thought possible. Building a successful business requires both determination and creativity. Starting from scratch, I poured everything I had into launching a venture that would change the world. Every day brought new challenges and opportunities to grow and learn. But ultimately, it was worth every struggle and sacrifice because now I'm proud of what we accomplished together. Leadership comes naturally to me, thanks to my upbringing in a family of entrepreneurs. We were constantly pushing boundaries and taking risks, so I learned early on that persistence pays off. And even though our journey wasn't easy, we never lost sight of our ultimate goal - to succeed and leave a lasting impact on society.

Prompt: write a 300 word essay about "My dream job" involving software engineering, computer science and AI

Response: Software engineering, computer science, and artificial intelligence (AI) have always fascinated me. At first glance, they seem like vastly different fields, but upon closer examination, there are many similarities between them. Both require problem-solving skills, attention to detail, and a willingness to push boundaries. Software engineers must create complex systems that seamlessly integrate hardware and software components, ensuring seamless functionality across devices. Similarly, computer scientists must understand the underlying mechanisms behind digital technology and ensure efficient coding practices. AI, on the other hand, combines these two disciplines by harnessing machine learning algorithms to analyze vast amounts of data and provide insights that humans could never imagine. By studying these fields, I hope to gain a deeper understanding of the complexities involved in programming and contribute to advancements in the tech industry.

Prompt: write a 300 word essay about "My dream job" involving cooking and running a restaurant

Response: Cooking and running a restaurant has always been a dream of mine. Cooking allows me to experiment with flavors and ingredients, perfecting dishes to satisfy customers' cravings. Running a restaurant offers me the opportunity to bring people together and foster community through food. I want to create a space where everyone feels welcome and appreciated, just like a home away from home. To do this, I'll need to balance customer service with culinary expertise, ensuring guests receive exceptional meals without feeling overwhelmed. Running a restaurant also teaches me about balancing profit margins with providing quality food, keeping costs low while maintaining high standards. Ultimately, I aspire to combine my passions for cooking and hospitality, creating a thriving establishment that brings joy to patrons and fosters meaningful connections.

Prompt: write a 300 word essay about "My dream job" involving electrical engineering, civil engineering

Response: My dream job would be a combination of electrical engineering and civil engineering. I have always been fascinated by the way electricity and civil engineering come together to create amazing structures and systems.

As an electrical engineer, I would be responsible for designing and developing electrical systems that power buildings, machinery, and other electrical devices. As a civil engineer, I would be responsible for designing and building the physical structures that support these systems, such as buildings, bridges, and roads.

Working in both fields would allow me to bring together my passion for electrical engineering and my interest in civil engineering. I would be able to use my knowledge of electrical systems to design and build structures that are both efficient and sustainable. I would also be able to use my understanding of civil engineering to design and build structures that are safe, durable, and aesthetically pleasing.

One of the most exciting aspects of this dream job would be the opportunity to work on large-scale projects. As an electrical engineer and civil engineer, I would be able to work on projects such as skyscrapers, airports, and other large-scale infrastructure. These projects would require me to use my skills and knowledge to design and build structures that are both functional and beautiful.

Overall, my dream job would be a fulfilling and challenging career that would allow me to use my skills and knowledge to make a meaningful contribution to society. Whether I was working on a small-scale project or a large-scale infrastructure project, I would be able to use my passion for electrical engineering and civil engineering to create amazing structures and systems that would benefit people for years to come.

APPENDIX D: Individual Report

Chua Jack Yune (A0269363U)

1. Personal Contribution

My personal contribution to the project consists of several parts. As I am the main programmer in the project, my job scope includes implementing the frontend, backend, database management, data gathering and model building.

For the frontend, I designed and built the user interface on Vue 3. I had a hand in building the user login and registration, chatbot and recommendations display pages. I built the backend API on Flask, including the chatbot logic and the recommender model.

From the data gathering side, I also suggested the use of O*NET and data.gov.net Singapore universities tracer study data for recommending occupations and educational courses. I proceeded to gather and process these datasets with the help of my colleagues.

On the modeling side, I suggested and implemented the use of content-based recommendation system, based on the fact that we have ample data on O*NET from Subject Matter Experts, who are occupational or organizational psychologists. This is based on my past experience as a researcher in psychometrics for two and a half years, where I had experience helping in developing and analyzing a paper-and-pencil test for career guidance. I also built the model based on the input from the career counselor and the O*NET data available, guided by my past experience in the subject.

2. Most useful skill learnt

The most useful skill I learnt was building a content-based recommendation system from start to finish using Python tools like Pandas, ScikitLearn, Word2Vec and Hugging Face's BERT Sentence Transformer.

In the beginning, I learnt much with regards to the process of data wrangling and transforming the data into a format that can serve as meaningful input to a machine learning model. My skill with cleaning and formatting data with Pandas improved significantly, as I was less familiar with this library in the beginning. Methods of aggregating data and efficient ways of applying different functions to data, and the usage of vectorized operations. I researched ways of optimizing the performance of Pandas operations, such as using vectorized methods, which can be useful when dealing with large datasets where processing time can be long.

When it comes to choosing or building a model, I have made many mistakes and learnt a lot about the varying performance of models in getting the results that I want. Because we wanted to build a model that lets users write and describe their dream job, instead of a multiple choice questionnaire, I had to study more on Natural Language Processing (NLP) for matching semantic content. The choice of models can involve some trial and error, where I

have to set a standard of the acceptable degree of error and try out different models from simple (Jaccard's similarity) to increasingly more complex (Word2Vec, BERT Sentence Transformers), in order to get a satisfactory match. Implementation of a model is also a challenge as one must learn the strengths and caveats of using each model, such as what data needs to exist and how complex the model would be (i.e. how much computing power would be needed, and will our machine succeed in running it in a timely manner).

3. How I can apply the skills learnt

I can apply the skills learnt in many ways. I currently work for a real estate software company. A content based recommendation system and NLP can be useful if a user has a specific type of home they want to buy or sell, and the chatbot recommendation system can be used to match a user's description of their ideal property, e.g. "2 bedrooms, 1 bathroom. Suitable for the family. Facing North. Located in CBD."; then the system can find matching properties of that specification for the user.

Conversely, if a user wants to sell a property, they can also input their specifications into a chatbot to find comparable properties to help them price their home. Comparable properties are a very important part of the property valuation process as people tend to look at properties of similar size, location, amenities and other factors and use those similar properties as a benchmark for their own property pricing. A content-based recommendation system can help with that, and a chatbot interface with intelligent text-based recommendation can greatly improve engagement as opposed to a dropdown filter that we see in PropertyGuru and other well-known sites.

The strategies learnt in this project can improve these processes and make the real estate transaction process smoother for clients.

Kwatt Ivy (A0269639H)

1. Personal Contribution

This is my first implementation of a machine learning application, so without much knowledge what and how it should be designed and implemented, I used my general project process as reference, I tried to apply my project knowledge and system integration and testing skills during the initial planning phase of the discussion. Given such a short time frame for the project and resource is limited and only one expert out of four of us in this domain, I opt for the viable project idea, which is having a ready dataset at hand to start with.

I provided our team a centralised file repository in google drive, as I already subscribed for personal and work use, so I created a project folder where everyone can use with no limit to the storage.

For data collection, I volunteered for a module on getting the dataset by webscraping and clean the data and use the word2vec model to match the onet data to web scraped data. I also try to get survey data through our questionnaire.

For the testing part, I proposed to use Katalon studio to do the test automation based on the chatbot rules, but there was not enough time to perform this part, and so I just verify the installation procedures and help to review and format the report.

2. Most useful skill learnt

I learned how to do web scraping, which is one of the many ways to collect data. As a fresh learner of machine learning, I understand the importance of having a good and reliable and updated dataset to have better predictions and recommendations.

I started off using an example code from github and modify for the “MyCareersFuture” site as target. I could not perform this task any faster, as it will take weeks to finish scraping 9000+ pages, so I decided to reduce to 3000+ pages. I also have to break it into few hundred pages at a time so it would not take too long, and I can use my laptop to start processing the data. I understand why having a powerful computer is necessary for machine learning implementations.

The other experience I enjoyed was when trying to find a good model to match the onet dataset job description to the MyCareersFuture job description, it was not a good result when I used Jaccard coefficient, as I believe it is matching word similarity rather than meaning similarity. So, I used the word2vec with google pre-built data to match the job description. It yielded a much better result.

3. How I can apply the skills learnt

Firstly, I can experience how a machine learning application was implemented, and by participating in the process, I can understand and appreciate what are the capabilities and limitations of a machine learning system and can apply the knowledge to the next project.

Secondly, lesson learnt from this project was to identify the gaps in order to make the process flow efficiently and final product meeting a certain quality. From design to implementation to integration and testing and deployment, we spend much time on learning the basics and concepts, there is little time on integration and testing. Also, the skills to implement full stack from backend database design and creating services to front end interface need has slightly more understanding and able to try out on the next project.

Borromeo, Angelie Quiapo (A0270177A)

1. Personal Contribution

I assumed the role of the primary project manager, responsible for overseeing various tasks such as gathering essential information and ensuring that target timelines were met. To

understand the project's business value, I conducted thorough investigations and research into its background. I facilitated discussions and consultations with Subject Matter Experts and utilized surveys as valuable resources for the project. Additionally, I actively participated in discussions to design the project's theme and content.

In an effort to process educational programs from the Singapore dataset (Graduate Data) into O*NET's Course Instructional Program scheme, I explored the application of Jaccard Similarity and Word2Vec techniques. However, the matching process did not yield satisfactory results, leading the main developer to adopt an alternative approach.

My contributions extended to writing the final report, as well as providing direction and editing support for marketing materials, user guides, and the system design video. Furthermore, I made an attempt to integrate the entire solution into a public cloud environment, and even wrote an AWS CloudFormation Template to facilitate this process. However, due to time constraints and unforeseen issues encountered along the way, the decision was made to implement the solution locally on the machine instead.

2. Most useful skill learnt.

I have recognized the practical applications of the fundamental knowledge and technological concepts acquired in my classes, particularly in the realm of intelligent recommendation systems. I have gained a deeper understanding of knowledge modeling and knowledge acquisition techniques, to enable the project in constructing extensive knowledge base. This knowledge has proven invaluable in addressing various challenges and translating technical theories into commercially viable projects.

Moreover, my experience has not been limited to technical expertise alone. I have also honed my soft skills, such as creating compelling marketing promotion videos and refining my writing abilities. Additionally, I have developed effective teamwork skills, allowing me to collaborate seamlessly with others. I have also acquired project reporting and presentation skills, which have enhanced my ability to communicate project progress and findings effectively.

Taking on the role of a project team leader has further refined my coordination skills, as I have assumed responsibility for overseeing the project as a whole. This leadership experience has provided me with valuable insights into the intricacies of project implementation and deepened my understanding of the holistic nature of project management.

3. How I can apply skills learnt

First and foremost, I value the knowledge I have gained in Intelligent Reasoning System, as it holds significant meaning for me in constructing future reasoning systems. Throughout this project, I've actively applied various concepts learned in class, such as text processing and Jaccard Similarity. While these specific methods were not ultimately employed, the practical experience we gained by working with them proved invaluable.

Moving forward, I intend to leverage the insights garnered from this project in my future endeavors, adapting and applying technology flexibly to tackle complex problems. I recognize the importance of engaging in active discussions and collaborations with others, as it fosters a collaborative environment conducive to innovative solutions.

The project also equipped me with essential skills in basic data processing and data visualization, both of which are crucial for aspiring product managers. Additionally, the project enhanced my ability to conduct comprehensive research on project backgrounds and ascertain their business value. Furthermore, honing my writing skills and capabilities has proven indispensable in the role of a product manager.

Secondly, the project afforded me the opportunity to develop my proficiency in creating impactful promotional videos, a skill that will undoubtedly contribute to future product publicity efforts.

Lastly, the project deepened my understanding of system concepts and honed my ability to analyze business value. This knowledge will undoubtedly prove advantageous in my involvement with any future projects.

APPENDIX E: Installation and User Guide

Following video presentation was created by the team to give a demo on installation and Usage of Carrier Chatbot.

Title: CareerWise Chatbot Installation and Usage Guide

- I. Introduction:
- II. Installing and Configuring CareerWise Chatbot:
- III. Launching the CareerWise Application:
- IV. Creating an Account:
- V. Engaging with CareerWise Chatbot:
- VI. Career Path Recommendation:
- VII. Conclusion:

Link: https://drive.google.com/file/d/13NQYipt-6jEE9UmYIh8rRoNrPyelWmnB/view?usp=share_link

NOTE: The User Guide documentation can be found under the repository “ProjectReport” directory.