

**INTENT-BASED NETWORKING: POLICY TO SOLUTIONS RECOMMENDATIONS**

By

Low Jun Sheng

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION TECHNOLOGY (HONS)

COMMUNICATIONS AND NETWORKING

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2020

## REPORT STATUS DECLARATION FORM

**Title:**           **Intent-Based Networking: Policy to Solutions Recommendations**

**Academic Session: 202001**

**I                   Low Jun Sheng**

**(CAPITAL LETTER)**

declare that I allow this Final Year Project Report to be kept in  
Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.



(Author's signature)

Verified by,



(Supervisor's signature)

**Address:**

\_\_\_\_ Jalan Universiti\_\_\_\_  
\_\_\_\_ Bandar Barat\_\_\_\_  
\_\_\_\_ 31900, Kampar, Perak\_\_\_\_

**Aun Yichiet**

\_\_\_\_  
Supervisor's name

**Date:**           **21/4/2020**  
\_\_\_\_\_

**Date:**           **23/4/2020**  
\_\_\_\_\_

**INTENT-BASED NETWORKING: POLICY TO SOLUTIONS RECOMMENDATIONS**

By

Low Jun Sheng

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfilment of the requirements

for the degree of

BACHELOR OF INFORMATION TECHNOLOGY (HONS)

COMMUNICATIONS AND NETWORKING

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2020

## DECLARATION OF ORIGINALITY

I declare that this report entitled “**INTENT-BASED NETWORKING: POLICY TO SOLUTIONS RECOMMENDATIONS**” is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.



Signature : \_\_\_\_\_

Name : Low Jun Sheng

Date : 21/4/2020

## **ACKNOWLEDGEMENTS**

I would like to express my sincere thanks and appreciation to my supervisors, Dr. Aun Yichiet who has given me this bright opportunity to engage in an Intent-Based Networking project. It is my first step to take part in the development of intent-based networking project. A million thanks to you.

Besides, not forget to thanks to my parents and my family for their love, support and continuous encouragement throughout the course. Without their support and encouragement, I won't have the strength and motivations to finish the whole project.

Last but not least, million thanks to my friends that always stands beside me no matter what troubles that I have encountered during the project development. They remind me no matter what difficulty that I faced, they will be standing my back and support me.

# **ABSTRACT**

Network design and solution architecting becomes challenging when multiple constraints are involved to comply with individual network policy. The semantic diversity of policies written by people with different IT literacy to achieve certain network security or performance goals created ambiguity to otherwise straightforward networking solution implementations. In this project, an intent aware solution recommender is designed to decode semantic cues in network policies written by various demographics for robust solution recommendations. A novel policy analyzer is designed to extract the inherent intents using a custom ML model to recognize network constraints and goals to provide context-specific recommendations. There are two components: (1) a custom intent recognizer A.I. trained with network logs first normalize spectrums of policies ranging from layman to domain-specific to detect entities of interests; such as data quota, access-controls, sharing permission, etc. (2) a recommendation system based on crowd-sourced ground truth to suggest optimal solutions to achieve the goals outlined in these policies. The experimental results showed that the proposed expert system is effective in general-purpose recommendations with an average score of 69% precision for different use cases and workload types.

# TABLE OF CONTENTS

<b>REPORT STATUS DECLARATION FORM</b>	<b>i</b>
<b>TITLE PAGE</b>	<b>ii</b>
<b>DECLARATION OF ORIGINALITY</b>	<b>iii</b>
<b>ACKNOWLEDGEMENTS</b>	<b>iv</b>
<b>ABSTRACT</b>	<b>v</b>
<b>TABLE OF CONTENTS</b>	<b>vi</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF TABLES</b>	<b>xi</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xii</b>
<b>CHAPTER 1 INTRODUCTION</b>	
1.1 Problem Statement	1
1.2 Background and Motivation	1
1.3 Project Objectives	3
1.4 Proposed Approach/Study	3
1.5 Highlights what have been achieved	4
1.6 Report Organization	4
<b>CHAPTER 2 LITERATURE REVIEW</b>	
2.1 VIVoNET: Visually-Represented, Intent-Based, Voice-Assisted Networking	5
2.2 Techniques of Semantic Analysis for Natural Language Processing – A Detailed Survey	7
2.3 Type of Recommendation System and techniques Used	10
2.4 ONOS Intent Monitor and Reroute service: Enabling Plug & Play Routing Logic	12
2.5 Towards End-to-End spoken intent recognition in smart home	14
2.6 Tab Completion	16
2.7 Policy Ambiguity	16
2.8 Intent vs Goal	17
<b>CHAPTER 3 SYSTEM DESIGN</b>	
3.1 System Overview	19
3.2 Train with Latent Dirichlet Allocation (LDA) technique with Amazon Comprehend	20
3.3 Computational function to look for intent through Lambda function	37
1. Store Policy in S3 Storage	37
2. Comprehend extract text file from S3 and run the computational job	38
3. Determine the intent of the policy	40

<b>4. Recommendations are given based on policy</b>	<b>43</b>
<b>CHAPTER 4 PRELIMINARY WORK</b>	<b>54</b>
<b>CHAPTER 5 RESULT AND DISCUSSION</b>	
<b>5.1 Calculations of precision metric</b>	<b>56</b>
<b>5.1.1 Comparisons of results</b>	<b>58</b>
<b>5.2 Calculations of Personalization Metric</b>	<b>63</b>
<b>5.3 Calculations of Coverage Metric</b>	<b>65</b>
<b>5.4 Finding and Implications</b>	<b>66</b>
<b>CHAPTER 6 CONCLUSION</b>	
<b>6.1 What we have achieved?</b>	<b>67</b>
<b>6.2 Future Direction</b>	<b>68</b>
<b>BIBLIOGRAPHY</b>	<b>69</b>
<b>APPENDIX</b>	<b>80</b>
<b>POSTER</b>	<b>84</b>
<b>PLAGIARISM CHECK RESULT</b>	<b>85</b>
<b>CHECK LISTS</b>	<b>88</b>
<b>FINAL YEAR PROJECT WEEKLY REPORT</b>	<b>89</b>



## LIST OF FIGURES

<b>Figure Number</b>	<b>Title</b>	<b>Page</b>
Figure 1.4.1	The proposed approach of the whole project	3
Figure 2.1.1	The workflow for Alexa	6
Figure 2.2.1	Example of RDF description graph	9
Figure 2.2.2	Example of RDF Ontology	9
Figure 2.3.1	The proposed architecture of a recommendation system	11
Figure 2.3.2	The equation used in context aware technique	11
Figure 2.4.1	IMR with ONOS and with the OPA	13
Figure 2.5.1	Pipeline Intent Recognition Method	14
Figure 2.5.2	SLU analysis and intent recognition for lighting up a bulb	15
Figure 2.6.1	In Command Prompt, when we type the first letter 'D' and type 'Tab' continuously, the system will show a few choices of directory which start with the letter of 'D'.	16
Figure 3.1.1	Basic flow of the intent recognition model	19
Figure 3.1.2	Flow of the system in Amazon Web Services (AWS)	19
Figure 3.2.1	Data Sampling for label WEB_ACCESS_CONTROL	30
Figure 3.2.2	Data Sampling for label AUTHENTICATION	31
Figure 3.2.3	Data Sampling for label PERFORMANCE	31
Figure 3.2.4	Data Sampling for label SELF_HEALING	31
Figure 3.2.5	Data Sampling for label NETWORK_SECURITY	32
Figure 3.2.6	The Python libraries needed for processing the dataset	32
Figure 3.2.7	The import of the full dataset	32
Figure 3.2.8	The sample text for the 5 intents	33
Figure 3.2.9	Converting the sample text into a single text file in Python	33
Figure 3.2.10	The converted text file is stored in the same directory with the Jupyter program file.	33
Figure 3.2.11	Entity list which consists of keywords and its label (intent)	34
Figure 3.2.12	Storing of both training documents and entity list	34
Figure 3.2.13	Give the name for the recognizer and define the custom entity type	35

Figure 3.2.14	Specify the method used in training the recognizer and the s3 path for storing the entity list and the training documents.	35
Figure 3.2.15	The intent classifier has successfully built	36
Figure 3.2.16	The sample input	36
Figure 3.2.17	The output from the intent classifier	36
Figure 3.3.1	The workflow of the intent-based recommendation model in AWS	37
Figure 3.3.2	The lambda function for putting an object into a S3 bucket	37
Figure 3.3.3	The input of the policy	38
Figure 3.3.4	The full Lambda function for trigger Amazon Comprehend	38
Figure 3.3.5	The computation job status in Amazon Analysis job portal	39
Figure 3.3.6	The output from the intent classifier model will be stored at a predefined S3 path.	40
Figure 3.3.7	The function that can read the content of a compressed zipped file.	40
Figure 3.3.8	The output from the intent classifier model	41
Figure 3.3.9	The function that able to extract the classified label	41
Figure 3.3.10	The function that looks for the intent	42
Figure 3.3.11	The function returns the intent after the computation	42
Figure 3.3.12	The table in Dynamo DB services	51
Figure 3.3.13	The details for attribute 'AUTHENTICATION'	51
Figure 3.3.14	The details for attribute 'NETWORK_SECURITY'	52
Figure 3.3.15	The details for attribute 'PERFORMANCE'	52
Figure 3.3.16	The details for attribute 'SELF_HEALING'	52
Figure 3.3.17	The details for attribute 'WEB_ACCESS_CONTROL'	52
Figure 3.3.18	The function that creates a connection to the database to extract the recommendations.	53
Figure 3.3.19	The output of the intent recommendation system.	53
Figure 4.1	The output for the intent 'WEB_ACCESS_COTROL'	54
Figure 4.2	The output for the intent 'AUTHENTICATION'	54
Figure 4.3	The output for the intent 'NETWORK_SECURITY'	54
Figure 4.4	The output for the intent 'SELF_HEALING'	55
Figure 4.5	The output for the intent 'PERFORMANCE'	55

Figure 5.1.1	The survey form for respondents	57
Figure 5.2.1	The formula that used to calculate similarity matric	64
Figure 5.2.2	The python code that used to calculate similarity matric	64
Figure 5.2.3	The result of the similarity metric.	65

## LIST OF TABLES

<b>Table Number</b>	<b>Title</b>	<b>Page</b>
Table 2.1.1	Sample utterance to invoke	6
Table 3.2.1	Table for specifying the source and explanation of the keywords for each of the intents.	30

## LIST OF ABBREVIATIONS

<i>IBNS</i>	Intent-Based Networking System
<i>CLI</i>	Command Line Interface
<i>IT</i>	Information Technology
<i>IoT</i>	Internet of Things
<i>NaaS</i>	Network-as-a-Service
<i>API</i>	Application Programming Interface
<i>A.I.</i>	Artificial Intelligence
<i>NLP</i>	Natural language Processing
<i>LDA</i>	Latent Dirichlet Allocation
<i>AWS</i>	Amazon Web Services
<i>SDN</i>	Software Defined Networking
<i>OvS</i>	Open vSwitch
<i>ASK</i>	Alexa Skills Kit
<i>DPID</i>	Data Path Identifiers
<i>CPU</i>	Central Processing Unit
<i>LSA</i>	Latent Semantic Analysis
<i>SVD</i>	Singular Value Decomposition
<i>RDF</i>	Resource Description Framework
<i>NLIDB</i>	Natural Language Interface to Database
<i>QoS</i>	Quality of Services
<i>QoE</i>	Quality of Experience
<i>PBNM</i>	Policy-Based Network Management
<i>PIP</i>	Policy Information Point

<i>PDP</i>	Policy Decision Point
<i>PEP</i>	Policy Enforcement Point
<i>ONOS</i>	Open Network Operating System
<i>ONF</i>	Open Networking Foundation
<i>IMR</i>	Intent Monitor and Reroute
<i>OPA</i>	Off-Platform Application
<i>CRR</i>	Clustered Robust Routing
<i>E2E</i>	End-to-end
<i>SLU</i>	Spoken language Understanding
<i>ASR</i>	Automatic Speech Recognition
<i>NLU</i>	Natural Language Understanding
<i>ACL</i>	Access Control List
<i>CSV</i>	Comma Separated Value
<i>ASCII</i>	American Standard Code for Information Interchange

## **CHAPTER 1 INTRODUCTION**

### **1.1 Problem Statement**

Network solution architecting is important to identify a set of tools to address certain networking goals and security requirements. Among a vast of solutions and providers, these options become a selection dilemma for DevOps to choose the right implementation considering the cost of deployment and interoperability issues. The workload for network administrator is getting bigger as the configurations of the network nowadays are getting more complex. With the introduction of Intent Based Networking System, it allows administrator to have an automated network management to achieve what they needed. Network administrator only need to specify their intention on network and IBNS will automate at scale automatically. (Butler 2017) Traditionally, if administrator would like to perform actions on network, they would need to have a background or knowledge in networking to control the network devices through Command Line Interface. Apart from that, the configurations of the whole network could be big issues for a network administrator to manage especially when there is always dynamical network change. The introduction of Intent-Based Networking allow network administrators need not worry about the solutions for any incident as the system will give the best recommendations that align the internal network continuously and dynamically business needs. (Intent-based Networking explained 2018)

### **1.2 Background and Motivation**

Intent-Based Networking is a technology that able to receive the desires from user and leverage the techniques of machine learning and user analytics to maintain the performance of network continuously. The technology is not new in the market as there are many similar precedent technologies.

The first digital voice assistant in the world, Siri has been released by Apple in 2011, which has taken a great step in human's machine learning technology. Siri was embedded iPhone 4s when released in the market. Users can use natural language to trigger actions from Siri and Siri can perform tasks on behalf of an individual. (O'Boyle 2019) The invention of Siri has progressed the Natural Language Processing (NLP) technology. With the advance of Technology, smart assistant has become a new trend as it could assist human in performing various tasks. Furthermore, there are getting more advanced voice assistant technologies invented such as Google's Home and Amazon Alexa. These technologies even utilize in the

application of IoT devices. Nevertheless, technology companies still working to improve the capability of voice assistant technologies to cater even more tasks from users. (Statt 2019)

Today, Cisco came out with an idea to adopt intent-based networking in business environment as a channel between IT and business. IBNS provide an abstract layer that enable network administrators to express their intent without having to worry about the underlying network architecture. The believed by 2020, IBNS will be the main stream for more than 1000 enterprises to be used in production. (Butler 2017).

IBNS simplify the network operations by having an automated network management. It can ease the workload for administrator as it seamlessly manages all the network devices in an interface. The realization of IBNS was came with the advancement of machine learning and orchestration technologies. The operation would start from translate high-level human language into machine understandable code and eventually turn it into command to manage network devices. (Intent-based Networking explained 2018)

There are two types of intent-based networking:

**General-purpose** - is an intent extraction method that uses an average weighted score of network entities to recognize underlying intents. In a network policy, multiple intents or sub-intents sometimes co-exist and conflicting. In this method, the algorithm simply takes the weighted sum of occurrence; or using term frequency to recognize the dominant intent among a set of intents. This is useful to remove some less important intents that are not directly related to networking, or could have been taken care of when the main intent is implemented.

**Special purpose** - extract intents comprehensively at a finer level. Depending on contexts, special purpose recognition is needed for robust solution recommendations to fulfill multiple constraints. This method removes the assumption of one size fit all solution, instead, it recommends a set of solutions and delegates fine controls to them for better coverage. Besides, to be more computationally intensive, special purpose can sometimes ‘over’ recommend due to model overfitting especially if there are many noises in the written policies.

In this project, a general-purpose solution recommender is designed to cater for wider use case while preventing overfitting issues.

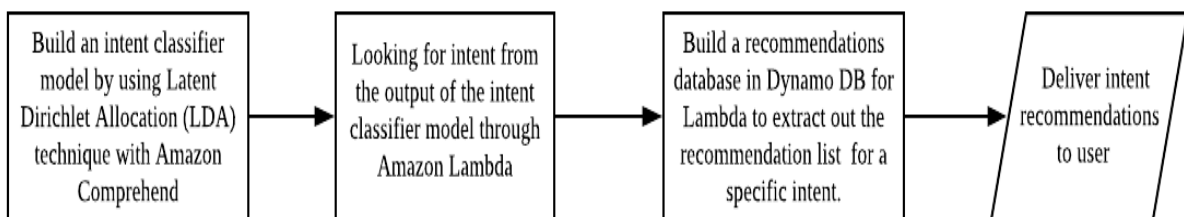


### 1.3 Project Objectives

There are two major process in this intent recommendation project. In the first stage, an intent aware model will be developed to carry out semantic analysis and determine the intention of an input. The second stage would be recommending solutions based on the context and scenario of a use case. A few objectives have been set and defined as following:

- i. To develop a network-domain specific analyzer by using the techniques of Natural Language Processing (NLP) for the automatic synthesis of network policies to extract the contextual networking meaning.
- ii. To develop an intent-aware recommender system using domain knowledge for automated network management for simplified network management and architecting
- iii. To design an empirical evaluation method-based crowd-sourced ground truth for the recommender.

### 1.4 Proposed Approach/Study



**Figure 1.4.1 The proposed approach of the whole project.**

In this project, an intent classifier model will be built for analyzing a policy text and extract out any keywords that are related to the intents. The training of the intent classifier model will be using one of the services in AWS which named Amazon Comprehend. Amazon Comprehend was using Latent Dirichlet Allocation (LDA), which is one of the techniques in Natural Language Processing (NLP) to map every document delivered to a topic and the words in each document are mostly captured by those predefined topics.

After having the output from the intent classifier, AWS Lambda will be the central processing management point to extract out the label consisted in the sample input and calculate the frequency of the label appear in the input policy. The label which has the highest frequency will be assumed as the main intent of the policy.

To have a recommendation database, Dynamo DB table was used to match different intents with the recommendations. After Lambda got the intent, the Lambda function will then extract out the recommendations from the Dynamo DB table.

In the last phrase, after having the intent and recommendations, the Lambda function will deliver the output to users.

### 1.5 Highlights what have been achieved

In Project I, an intent recognizer model has been developed to predict the intent for a specific policy. The technologies that used to build this recognition model were using Amazon Comprehend with the assist of Python Programming Language. Amazon Comprehend was a text analyzer that use topic modeling techniques to analyze a text input by using entity detection.

The intent recognition process will start when the user inputs an utterance to the recognition model for them to identify the intent. The recognition model will first identify the keywords in the utterance and classify them into different categories. For example, when user key in the utterance *“Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that are a violation of applicable law or company policy.”*, the recognition should be able to identify *‘fraudulent’*, *‘harassing’*, *‘obscene’*, *‘threatening’*, these are the keywords related to Web Access Control and there is a high possibility that requires recommendations that suggest network administrator control the web access.

### 1.6 Report Organization

In this report, it organized in the following structure: 1) 1) In Chapter 1, we will study the background of the project and determine the scope and objectives for the whole project development. 2) In Chapter 2, existing work that have been carried out by other researchers will be studied and we will analyze the effects can bring from these technologies. 3) In Chapter 3, we will outline a proposed method of the project development. Based on the proposed method, a detailed showing of the works toward the realization of the project will be discussed. 4) In Chapter 4, the preliminary work done from the proposed method in Chapter 3 will be presented. 5) In Chapter 5, a discussion on the result will be carried out based on the output obtained from the intent recommendation system and we will compare the output to a real-time scenario by conducting a survey. Besides, other metrics also will be used to evaluate the recommendation system. 6) In Chapter 6, we will make a summarization of the project that we have done and highlight any future direction that could be achieved.

## **CHAPTER 2 LITERATURE REVIEW**

### **2.1 VIVoNET: Visually-Represented, Intent-Based, Voice-Assisted Networking (Amar, Amrita, Atharva, Dewang, Lakshay, Levi, Rahil and Sapna 2019)**

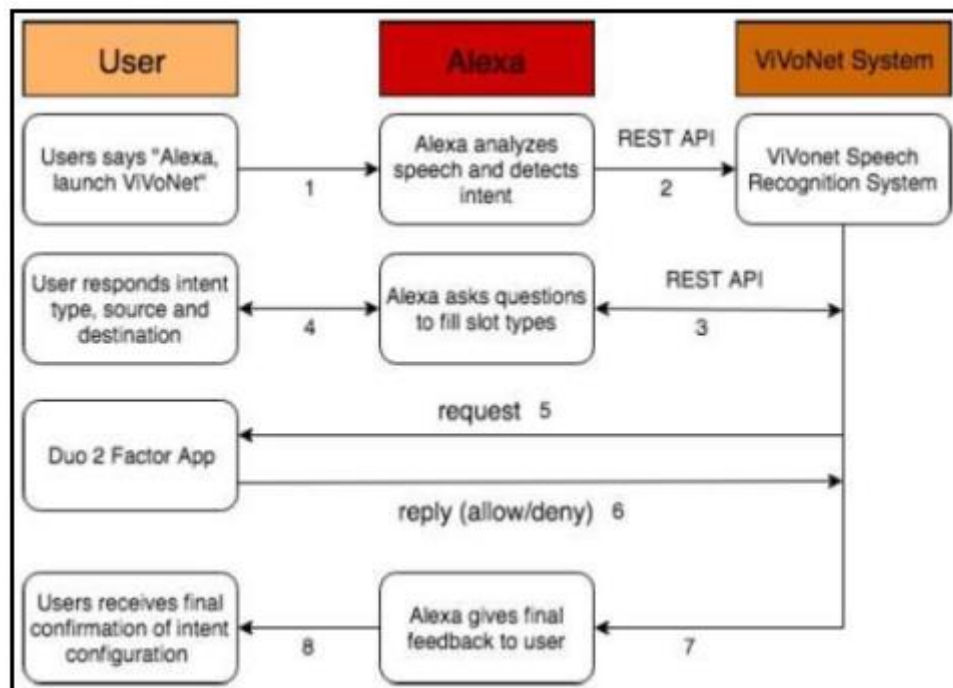
In this research paper, the author has proposed an Intent-Based Networking System (IBNS) which embedded with voice assistance that can visualize the underlying network architecture and network administrator can perform modification on the network based on intents through voice commands. With the integration of Software-Defined Networking (SDN), it allows the administrator to implement policies without having specific low-level details. (Y. Han, J. Li, D. Hoang, J. Yoo & J. Hong 2016)

The proposed system was using software named VivoNet. The system will accept voice inputs from the user and dynamically convert it into intents and implement them on suitable network infrastructure. VivoNet was developed by using Python programming language. The front end and the REST API were developed by Django, which was a Python Based model template architecture that can ease the creation of complex websites. (Meet Django 2018) The components of VIVoNet are virtualized on a virtual machine on VMWARE ESXi hypervisor. The virtual switches used in this development are Open vSwitch(OvS) that running OpenFlow 1.3 and there is a centralized SDN controller to manage all the OvS. The development of the project is believed to have a significant impact on network administrators as well as visually-impaired audiences.

To enable the system to accept voice input to the system, Amazon Echo Dot (2<sup>nd</sup> Generation) has used to enable user utterances to be identified. Echo Dot with access to the Internet has connected Amazon-Alexa service over the Wi-Fi network. With Amazon Alexa, the Speech Recognition system can directly accept these voice commands. A custom skill has been created by VIVoNEt using Alexa Skills Kit (ASK). (What Is Alexa? n.d) The custom skill collects a set of sample utterances to invoke specific intents and slot types to stimulate the input from the user. In the VIVoNet system, the intent Engine responsible for converting user-requested intents into the translated network configuration. The Engine will create appropriate flows and the controller will push them out using OpenFlow protocol to the switches. The switches greatly depend on the communication between the controller and OpenFlow else it will forward traffic based on the flow tables.

The flow of the process will start from accept voice input from the user and Alexa will analyze the user utterance and identifies the corresponding intent. After that, the user input will

first convert into text and Alexa will match the text with all the sample utterances from the pre-configured intents. If there is a match, it will match user input with slot-types by asking additional questions. The slot-types values and the intent will be sent to REST API by a POST request.



**Figure 2.1.1 The workflow for Alexa**

Slots Name	Slots Type	Sample Slots Values
{intent_type}	intent_type	least hopcount, least latency, high bandwidth
{from_city}	AMAZON.City	Denver, San Francisco (any U.S. city)
{to_city}	AMAZON.City	New York, Chicago (any U.S. city)
{confirmation}	confirmation	Yes/No

**Table 2.1.1 Sample utterance to invoke**

The Speech Recognition system will first provide the required intent to the Intent Engine. The intent engine retrieves the switch Data Path Identifiers (DPID) from the controller, and it will find the suitable paths between Data Path Identifiers (DPID) by querying the

controller again. After the controller computes the best available path, it will use OpenFlow's protocol to fetch the required information to the respective switches. Once the flow was successful, the intent and flows are will be written to the database. The intent processing system will return True if success and return False if failure.

From the results obtained by the developer, it is shown that VIVoNet can achieve agility as it required 0.0016 seconds on average to create, push and verify flow for one intent. Besides that, the VIVoNet system enables administrators to customize their requirements without having to rely on vendors. As network nowadays becoming getting larger and complex, VIVoNet greatly helps operators with visual impairment by making information more accessible and thus allowing them to communicate with the network infrastructure. Network administrators always can alter network behavior through voice commands.

In the future, VIVoNet would like to extend its capability through its voice assistance capabilities. It will be owing to retrieve network device health conditions such as memory and CPU utilization for better performance. Besides, it will even go further by having a network monitoring feature.

## **2.2 Techniques of Semantic Analysis for Natural Language Processing – A Detailed Survey (Hanumanthappa 2016)**

Semantic Analysis is one of the important parts in Natural Language Processing (NLP) system as it could determine the meaning of the given input through a detailed analysis. The exact meaning could be extract from the semantic interaction with different linguistic levels. In this paper, a study was carried out on the process of NLP and the current practices that using Semantic Analysis in different use cases.

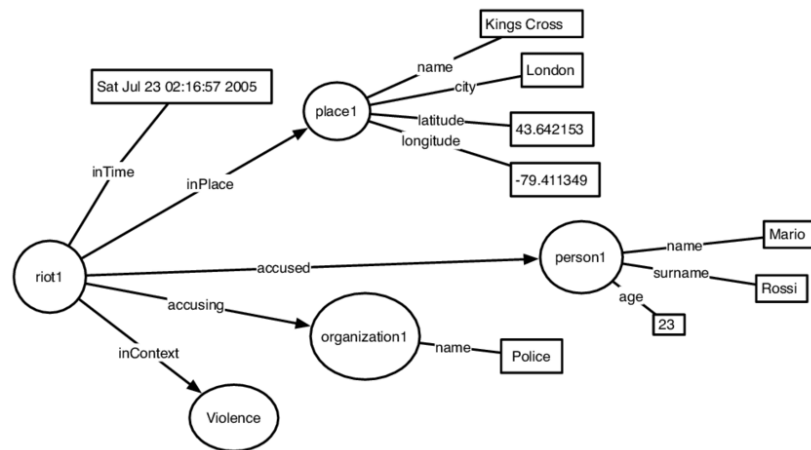
Natural Language Processing is a technique that makes the computer system able to understand the text used in natural language. It involved 4 stepes which are Morphological processing and Lexical Analysis, Syntax Analysis (parsing), Semantic Analysis and Pragmatic Analysis. First of all, morphological prcessing will first separate a sentence into a set of chunk words. At the meantime, Lexical analysis will take charge to divide the whole chunk of text into paragraphs, sentences, and words. (Natural Language Processing (NLP), n.d.). The separated chunk words will be categorized based on their grammatical class. After that, Syntax analysis will analyze the relationship between different chunk words. (Redd & Hanumanthappa

2014). At last, semantic analysis will be analyzing the meaning of the text and pass the result to Pragmatic analysis to understand the actual meaning of a text based on specific context.

Semantic analysis will retrieve the information and understand the meaning of words with the particular context. It might use technologies such as data mining and machine translation to capture the meaning of the text more accurately. It will first assign the tokenize words into different grammatical class and structure it to find out the real meaning of the input. Let's take an example, a train schedule can be related to the departure time, estimated arriving time, train model, or platform number. The process of semantic analysis required the system developer to predefined the dataset and train a semantic model that could understand the actual meaning of the text. (Expert System Team 2017)

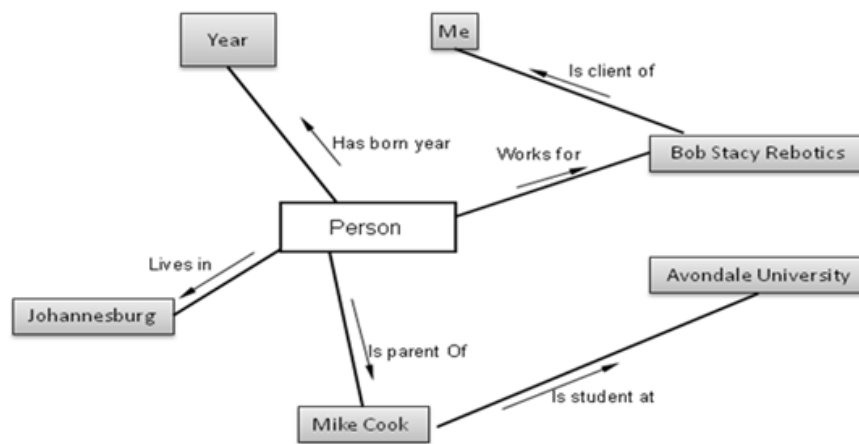
A few real-life applications of semantic analysis have been listed out in this paper. First of all, Peter W.Foolitz provide a technique for writing a summary by comprehending a textual information by using Latent Semantic Analysis (LSA). LSA will first analyze the relationship between a set of words as it assumes the words that have the similar meaning will occur in the same pieces of text. After that, it will construct a matrix which consist of the occurrence of unique word from the text. A mathematical approach named Singular Value Decomposition (SVD) will then decomposes the word-by-document matrix into a set of  $k$ . After that, the set of  $k$  are compared by taking the cosine of the angle between the two vectors formed by any two columns. Values close to 1 represent text with similar meaning while values close to 0 represent very dissimilar meanings. By using LSA, researchers found out that it's well-matches for researchers in education to look for textual resources which have semantic similarity. (Foltz 1996)

Apart from that, researchers Harish Jadhao, Dr. Jagannath Aghav, and Anil Vegiraju using semantic tools to extract information from unstructured data and visualize it through spring graph. In today world, numerous information spreads across the Internet such as social media,news and video content and those information will eventually become worthless if we did not locate and extract the usable information from unstructured data. Structured data will have a set of entities within a domain and relationship that links these entities. These entities will be eventually represented into RDF (Resource Description Framework) graphs.



**Diagram 2.2.1** Example of RDF description graph.

Before the entities structured into RDF graph, the author used Ontology learning method to define the concepts and relationships between entities. It will extract the domain terms such as the object, action and subject from textual data. The extracted entities will eventually be represented in spring graph. The spring graph could also be use in information extraction after semantic analysis. To conclude, the information extraction could be represented by using RDF graph with the use of Ontology method or visualize in a spring graph after semantic analysis. (Zhang and Mostafa 2002)



**Diagram 2.2.2** Example of RDF Ontology.

Avinash J. Agrawal and Dr. O. G. Kakde used domain Ontology for to organize information from unstructured data. Ontology will represent the types, properties and the associations between entities. The author uses semantic analysis on natural language queries and store the information extracted into Natural Language Interface to Database (NLIDB). The use case used in this research is use inquiry. Most of the time, we will come across different

kind of inquiry such as railway inquiry, bank inquiry, business inquiry, etc and it's impossible for the staff to answer all those inquiry from time to time. To simply the work, the query will be pre-processed by using semantic analysis and the result will be stored in NLIDB. For example, railway inquiry often involves the domain words such as train, station, fare, etc. By gathering these information, it will be described with the associated set of answer. These information will be stored in NLIDB. If someone ask for the same inquiry, the system could extract the information from the database and present the information to user. (Agrawal & Kakde 2013)

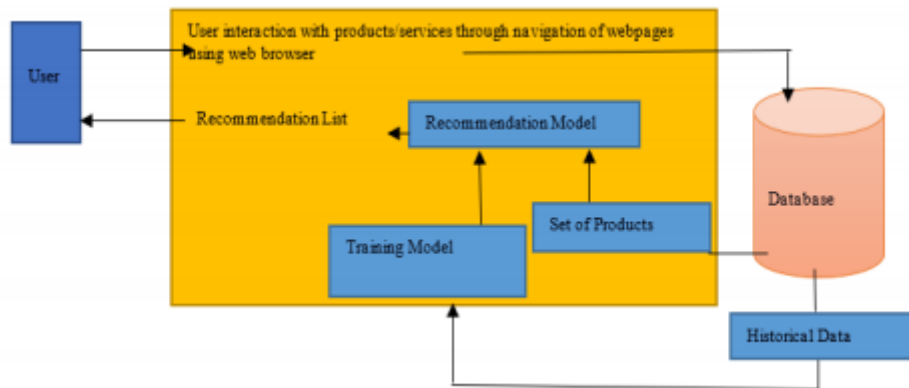
Semantic analysis could be used in many applications and semantics was mainly focus on the study of meaning. From the example given in this paper, it is believed that semantics is necessary when understanding a text. In future implementation, ontology could be the main tools to extract usable information from unstructured data.

### **2.3 Type of Recommendation Systems and Techniques Used**

Every time user browses the system to look for new items, famous online website such as Amazon, Netflix or Yahoo would provide suggestions based on their past history and demographic profile. (Afkhamizadeh, Avakov, and Takapoui 2013) With the exponential growth of information across the web, it is difficult to locate the information.

Modeling and analyzing of user navigation behaviour help the system to understand what online users demanded. To perform that, 'Information filtering' is the method to manage the flow of the large information, Information filtering involves the techniques of web usages mining to present the relevant information to user. The recommendation system will identify user's interest based on their previous experiences such as rating, browsed website or comments. For instance, if a user interest in photography then he/she may also interest in related photography website. Hence, the proposed recommendation system will predict the interest of a user but not ranked by the user. (Singh, Rishi, Awasthi, Srivastava and Wadhwa 2020)





**Figure 2.3.1 The proposed architecture of a recommendation system**

There are many ways to design a recommendation system. Developer first need to determine what kind of recommendation system want to build, whether is a personalized recommended system or non-personalized recommended system, a personalized recommended system would provide different recommendations list for different users while non-personalized recommendations system would only provide the same content based on the review or interest of user. (Singh, Rishi, Awasthi, Srivastava and Wadhwa 2020)

Collaborative Filtering (CF) Technique is the most widely used approaches for recommendation system. It can be classified into memory based and model based. Memory based collaborative filtering technique is based on the similarity among user, product or services to predict the possible interest of a user. Model-based collaborative technique uses machine learning technique to give the recommendations which has higher accuracy compared to memory-based collaborative filtering techniques. There are number of predictive models that using model based collaborative filtering techniques such as Deep learning method and Bayesian Network. (Singh, Rishi, Awasthi, Srivastava and Wadhwa 2020)

Another widely used approach in recommendation system is context aware technique. In this approach, user interest will be the basis of the situations and the new recommendations from the model will be the context. The working solution shown in Figure 2.3.2. R is the possible interest of user and D1, D2 are different situations for a context. (Singh, Rishi, Awasthi, Srivastava and Wadhwa 2020)

$$f_R = D_1 X D_2 X \dots X D_m \rightarrow R \dots\dots\dots$$

**Figure 2.3.2 The equation used in context aware technique.**

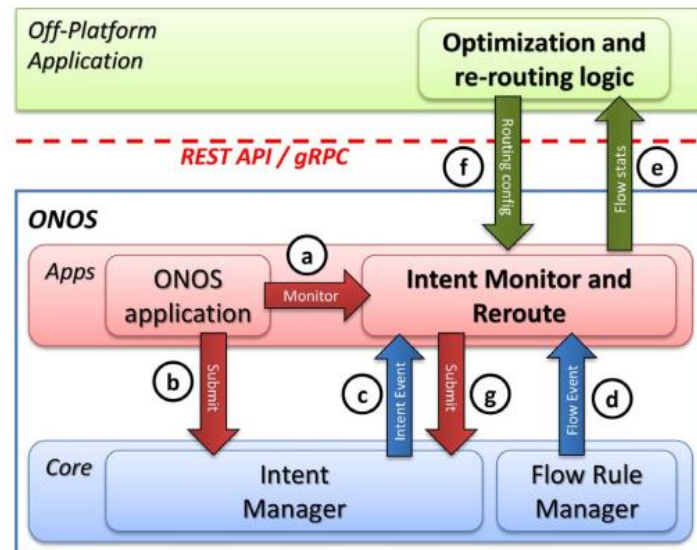
To evaluate the recommendation model, precision and recall is the most common metrics. Precision is the measure of users' preferences among the suggestion your recommendations system has given, it used to compute how much does the recommendation system meet user requirement. Recall is the another metrics that used to measure the sensitivity of the system, it is measure by having recommended item and used item over the overall item has been used by user. Both metrics is essential for a recommendation system when comes to evaluate the quality of a model. (Bondarenko 2019)

#### **2.4 ONOS Intent Monitor and Reroute service: Enabling Plug & Play Routing Logic (Davide, Daniele, Mattia, Ilario, Antonio 2018)**

An Open Network Operating System (ONOS) application developed by Open Networking Foundation (ONF), is an open-source project which provides an easy and high-performance way of maintaining the network state. The application just has to specify their intent without having to worry about any low-level functionalities. It able to meet the requirements of operators by creating and deploying dynamic network services with simplified programmatic interfaces. As ONOS supports both configuration and real-time control of the network, it eliminates the need to run routing and switching control protocols inside the network fabric. In the ONOS platform, it consists of a set of applications that act as an extensible, modular, distributed SDN controller. And it is a scale-out architecture to provide the resiliency and scalability required to meet the rigors of production carrier environments. (Technical Steering Team (TST), n.d.)

ONOS carry out their service by monitoring and rerouting of the intents. The system developer defines a set of Application Programming Interface (API) to make it interact with an Off-Platform Application (OPA). The OPA will implement the re-routing logic which ranges from classical optimization tools to Machine Learning or Artificial Intelligence approaches based on collected statistics.

In ONOS architecture, there is a service called Intent Monitor and Reroute (IMR) which allows ONOS applications and users to specify a set of intents whose statistics are monitored and exposed to an external routing logic. IMR will the central components to communicate with ONOS Intent manager and flow rule manager to monitor the flow between the intent and the corresponding flow rules.



**Figure 2.4.1 IMR with ONOS and with the OPA**

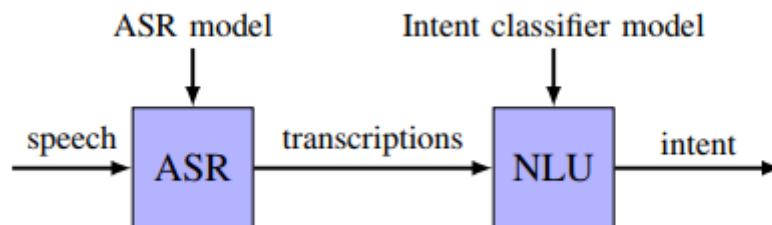
Based on Figure 2.4.1, Intent Monitor and Reroute (IMR) offer ONOS applications and users desire to monitor and re-routing of specific intent. ONOS monitoring will periodically retrieve the statistic of each low-level flow generated and apply corresponding intent in the network. IMR always has to interact with ONOS Intent Manager and Flow Rule Manager to keep track of the mapping between the intent and corresponding flow rules and other related information. IMR has 3 states for an intent that is '*Not Monitored*', '*To Be Monitored*' and '*Monitored*'. By default, the intent state is '*Not Monitored*' until when a user or application requires to monitor an intent, the state will either change to '*To Be Monitored*' or '*Monitored*'. If the state is in '*Monitored*', it will start tracking of its statistics. Based on the statistics retrieved from IMR, Off-Platform Application (OPA) can specify a particular path for each of the intent to optimize the network performance. Through this mechanism, IMR just has to submit the intent keys and the rest of the work will be left to OPA. OPA will collect the statistics, optimize the algorithms and reroute the intents if necessary. (Davide, Daniele, Mattia, Ilario, Antonio 2018)

Different network objectives will achieve with different routing logics by the OPA. The paper proposed Clustered Robust Routing (CRR) which is an adaptive Robust Traffic model that allows tuning the trade-off between dynamic routing and stable routing. The historical data will be feed into the optimization model over the training period. The model will compute a series of routing configurations that will be applied in the network. To deal with dynamic traffic, routings are trained with more robust over subsets of traffic matrix space.

With ONOS, the monitoring executed by OPA is important to handle any unexpected traffic scenarios. For each of the ONOS module, it can optimize the routing logic in ONOS applications based on intents, via an external plug and play routing logic with a few modifications to the application code. The performance can be easily improved by leveraging routing logic decoupled from the application which requested the intents.

## 2.5 Towards End-to-End spoken intent recognition in smart home (Desot, F. Portet and Vacher 2019)

In today's world, there are getting more smart home products featured in voice-based interaction to perform functions. These smart home products responsive to voice commands and able to identify the intent of the user through voice message. Intent recognition was relying on Spoken language Understanding (SLU) to understand the meaning from an input. (Dmitriy, Wang, Christian, Kumar, Liu, Bengio 2018) In this paper, the author presents an end-to-end (E2E) SLU approach which integrate both ASR and NLU in one model. Voice-based intent recognition is typically performed through Automatic Speech Recognition (ASR) and its outputs transcriptions that fed to the intent recognition module for intent extraction. Speech-Language Understanding (SLU) is a slot-filling approach that used to predict the intent from a user by extracting entities in a spoken utterance (slots and values).

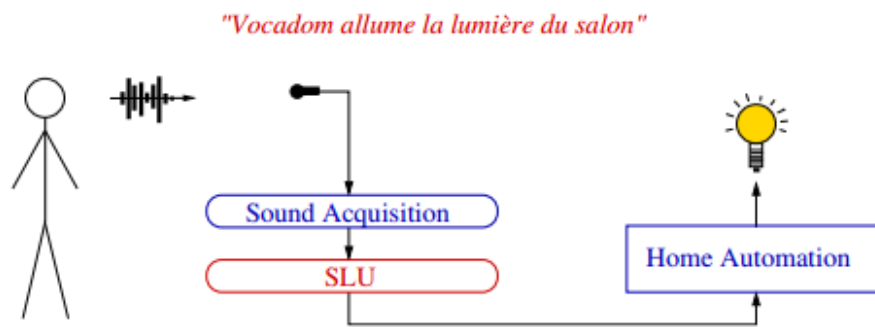


**Figure 2.5.1 Pipeline Intent Recognition Method**

The method proposed by the author consists of ASR and NLU. Based on the diagram above, the ASR system generates hypothesis transcriptions from a speech utterance and the transcriptions will be analyzed by the NLU module to extract meaning. The intent classifier model used in this approach is a seq2seq attention-based PyTorch model that encodes the sequence of words and decodes a sequence of symbols representing the global intent and the classifier performed jointly with slot recognition. (Liu, Ian 2016) Each slot contained in the sequence supports the intent classification. For example, if user given a sample utterances “Turn on the light”, the intent classifier will first generates the sequence intent[set\_device], action[TURN\_ON], device[light]. In this case, the intent is to set a device and the slots action

and device provide information about which entities are concerned with the voice command. The NLU model used in this approach used slot-filling as a sequence labeling task. A sequence labeling task requires each word in the transcription aligned with one unique slot label.

The pipeline Intent Recognition model has been tested in the case of voice command in smart homes. From the result, the utterance is captured and analyzed by the SLU module. If the intent to control the house, the semantics is extracted from the utterance and sent to the home automation system.



**Figure 2.5.2 SLU analysis and intent recognition for lighting up a bulb.**

Nevertheless, user could only give simple set of commands as it restricted with specific grammar and not enough to handle a large set of intents with a lot of syntactic and lexical variation. Since the amount of real-data is not sufficient for training, the author enhanced the recognition model by using a corpus generator of Desot et al to produce training data automatically labeled with intents, slots, and values.

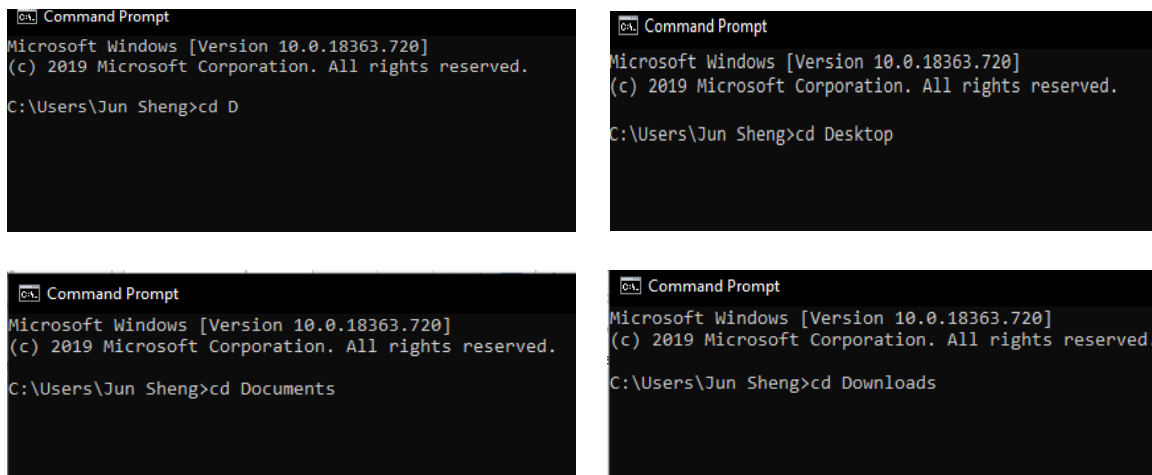
The E2E SLU approach heavily relies ASR performance, if the user command translated wrongly by ASR, it will affect the intent prediction by NLU. To reduce the error, sub-sampling of the minority and majority classes was implemented to the sampling data to improve the performance. The NLU model may use unaligned data to give the flexibility to infer slot labels and values from incomplete transcription to recognize the intent.

In conclusion, the E2E intent prediction approach could be implement in the smart home system as it consists of ASR and NLU. Further work to extend the intent recognition approach will be using transcription augmentation. Multi-task and transfer learning with models were also studied in future work.

## 2.6 Tab Completion

Tab completion is a technique that is mostly used in most of the command-line environment. It can increase user typing speed when typing commands. With a hit of tab button, the command can be complete automatically with your preferences of choice. For example, in Linux Command Line environment, if we have various complex and long file name which start with a letter of 'A', and we just want to access a folder. To avoid the typing of long command, we may just type A and press Tab to complete the file name. The system may give you more than one choice which match to your preferences. (Hoffman 2014)

Tab Completion is useful to complete a certain command for a given executable environment. It can save users' time from typing long text and assist user to know the available commands in a given command-line environment. Besides, it helps to prevent errors by showing the options available based on what user have already typed. (Creating a bash completion script 2018)



**Figure 2.6.1** In Command Prompt, when we type the first letter 'D' and type 'Tab' continuously, the system will show a few choices of directory which start with the letter of 'D'.

## 2.7 Policy Ambiguity

As network environment are getting complex, policy-based management facilitates policies provisioning to achieved the desired quality of service (QoS). However, conflicts of policy specification may lead to policy ambiguity and unpredictable effect on network performance. When there are multiple similar policies coexist, there is a possibility that policy will be in conflict either because of specification error or application-specific constraints. (Charalambides 2005)

Modality conflicts is one of the main causes of policy ambiguity. If two policies are specified using the same targets, subjects and actions but different set of policies applied. It may lead to overlap between the output of the policies. (Lupu & Sloman 1999) For example, when a network administrator wants to restrict only 1GB of the Internet bandwidth for every user within an organization network, he set the policy as the first network policy. However, the director is excluded from that policy and he has the privileges to used up to 5GB of Internet usage within the organization network. Hence, the network administrator once again set the policy that allow the director to used up to 5GB of Internet usage. Here is a conflict between the first and second policy as network administrator already mentioned at first all the user only can used up to 1GB which include the director as well, hence the second policy will not work even it is specified.

To resolve this issue, network administrator may consider resolving the policy conflicts by assigning precedence to policies. By establishing policy precedence, it allows two similar policies coexist within the network system and determine which policy should be applied. Policy conflicts could be substantially reduced by establishing the precedence and priority between different policies. Let take the previous example, if the network administrator changes the arrangement of the policies by setting the director having 5GB of Internet usage as the first policies, and the second policy set all other users only can used up to 1GB of Internet usage. There will be no conflict between these two policies as the first policy will have the priority over the second policy. (Lupu & Sloman 1999)

## 2.8 Intent vs Goal

Intent and goal have a strong relationship in understanding a policy and giving out recommendations. First of all, intention is a general term that try to bring up actions around, and goal is a more concrete actions that want to accomplish. (What's the Difference in Intentions vs Goals? 2020) When understanding a policy, we may first want to extract the intent as intent bringing up actions. For example, a network administrator wishes to improves the security of the organization network. Hence, '*improves security*' is the intent of the text as it is the objective of the text and could bring more actions such as installing firewall, setting ACL or setting VLAN. On the other hand, goal is the more specific term that we want to declared. For example, network administrator wants to deploy proxy for controlling the web access of employees. From the use case, we know that '*deploy proxy*' is the specific goal that the user wants to achieve, and it is the final deliverable of the intention of the text. To summarize, the

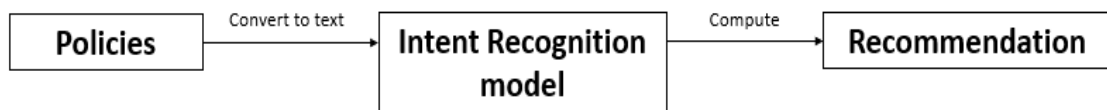
relationship between intent and goal is strongly correlated, the goals may come with an intention and an intention may bring up goals. In computing, semantic analysis could be one of the ways to understand the information of a text and figuring out what the actual intent was based on the context in the text. (Hanumanthappa 2016) Setting up goals will be based on the result from semantic analysis and giving out related recommendations to achieved the intention of a text. One of the possible methods for the system giving out goals or suggestion could be embedded neural network in recommendation system. (Koehrsen 2018)



## CHAPTER 3 SYSTEM DESIGN

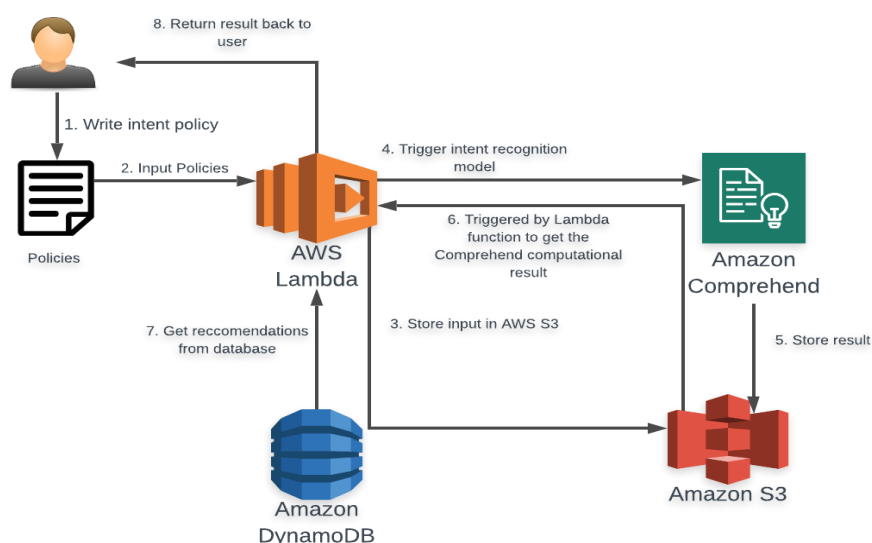
### 3.1 System Overview

The proposed approach used in this project will be going through two pipelines. Firstly, the policies will translate into intent through an intent recognition model. The intent recognition model will then identify the keywords in the input text and convert it into intent. After having the intent of the text, the next process is to give recommendations to the user based on the specific intent.



**Figure 3.1.1 Basic flow of the intent recognition model.**

The project will be solely based on the services in Amazon Web Services (AWS). The development of the system include the services such as Amazon Comprehend, Lambda, Simple Storage Services (S3) and Dynamo DB as the fundamental of the development model. The system will start with the user input a usage policy to Lambda, after that, a Lambda function will convert the policy into a text file and save it in S3 storage. The next following Lambda function will trigger Amazon Comprehend function to analyze the text file and extract the keywords in the text classifier into different labels. The Lambda function will accumulate the labels after this and compute for the intent. Next, the Lambda function will have some computation based on the intent and finally return a list of recommendations to the user. The detailed flow is shown below:



**Figure 3.1.2 Flow of the system in Amazon Web Services (AWS)**

Each of the Amazon Web Services (AWS) will have different function and capabilities, the specifications are listed as following:

**AWS Lambda** – a serverless computing platform that runs code in response to events and automatically manages the computing resources required by that code.(AWS Lambda n.d.) In this project, Lambda is used for computation and acts as central management to trigger the other services.

**Amazon Comprehend** - A natural language processing (NLP) service that uses machine learning to find the relationships in a text. It uses topic modeling to determine the entities consist the text. (Amazon Comprehend, n.d.)

**Amazon Dynamo DB** – A NoSQL database service that supports key-value and document data structures. It used to map the intent with recommendations in this project. (Amazon DynamoDB, n.d.)

**Amazon S3** – Used to provides object storage through a web service interface. (Amazon S3, n.d.)

In each of the services, there consist of different configurations and settings. In this chapter, every single step of the development of the recommendations model will be explained and it will start from: 1) Train with Latent Dirichlet Allocation (LDA) technique with Amazon Comprehend. 2) Computational function to look for intent through Lambda function. 3) Translation of intents into solution recommendation.

### **3.2 Train with Latent Dirichlet Allocation (LDA) technique with Amazon Comprehend**

The intent recognition model required data collection for training. The prepared dataset will be collected from different sources to build an entity list for training. For training an intent recognizer in Amazon Comprehend, an entity list and training documents are needed. Entity List is a list that consists of related keywords from the sample text and classifies the keywords into different categories while sampling document is the datasets that used to train the recognizer. Python programming language will be the main programming language in whole system development.

Data collection is the first step for training the recognition model. Before collecting the data, we have defined 5 labels for the intent recognition model to label which are

AUTHENTICATION, NETWORK\_SECURITY, PERFORMANCE, WEB\_ACCESS\_CONTROL, and SELF\_HEALING. Each of these labels required training sets for data sampling. The training datasets are getting from different sources such as organizations' Internet usage policies, websites or any other documentations. For each of the data sampling, we must first identify the sample sentences is belong to which labels. For example, one of the sample computer and network usage policy from Stanford University has the following policy, *"Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that are a violation of applicable law or University policy. In particular, contributing to the creation of a hostile academic or work environment is prohibited."*, when talking about restricting the browsing content of the user, we might relate to Access Control List (ACL) as according to Cisco Documentation, Access control lists (ACLs) able to carry out packet filtering to inspect the packet content and monitor the traffic of packets in a network. Packet filtering provides security by limiting the access of traffic into a network. Hence, we categorized ACL under WEB\_ACCESS\_CONTROL as it belongs to one of the approaches to control the web access and we take that network usage policy as our sampling documents.

Each of the labels requires sampling data for the intent recognizer able to map the entity list with the sampling data. To do that, we have prepared 5 sets of sampling data for different labels. For each of the intent, there will be different targeted keywords which can represent the intents. Each of the keywords was clarified from an authentic source that can map the keywords to the intent. Keyword mapping has a few matrixes which can determine the mapping of the intent, we take two of the matrixes as our baseline to map the keywords to intent which are relevancy and usefulness. Relevancy is the measure of the quantity or state of being closely connected or appropriate while usefulness is the measure of how informational that a keyword can bring to the label. (THATAGENCY 2020). Below is the table of intent, keywords, sources, and explanation.

INTENT	KEYWORDS	SOURCE	EXPLANATION
AUTHENTICATION	password	(Authentication, Authorization, and Accounting Configuration Guide, Cisco IOS Release 15SY 2019), (WS-Security UsernameToken authentication, n.d.)	Password authentication is the most common way in authentication as it appears in almost all the Internet usage policy and it is used for

			verification purposes.
	<b>identity</b>	(Identity Authentication – Are They Who They Say They Are? 2018), (Rouse 2018)	Authentication is the act of verifying one's identity, hence 'identity' can be one of the most relevant terms to Authentication.
	<b>username</b>	(WS-Security UsernameToken authentication, n.d.), (Security Authentication vs. Authorization   What's the Difference 2018)	Username and password commonly used as the identifier and key for accessing an account for authentication.
	<b>Authorization</b>	(Rouse 2006), (authorization, n.d.)	Authorization is the process of verifying what you have access to which is also. Authentication required authorization to determine which client as the access to login an account.
	<b>credentials</b>	(Young 2016), (Credentials Processes in Windows Authentication 2016)	Credentials have a strong relation with username and passwords as it's used for login an account
	<b>verification</b>	(WS-Security UsernameToken authentication, n.d.), (Security Authentication vs. Authorization   What's the Difference, 2018)	Verification of identity always used in authentication, hence, 'verification' in policies used as a keyword for Authentication intents
	<b>permission</b>	(Poremba 2017), (Rouse 2018)	Permission used to describe a user who has the right to gain access to an account.
	<b>domain controller</b>	(Credentials Processes in Windows Authentication 2016)	Domain controller used as an authenticating target when there is a

		, (Petters 2018)	domain-joined computer. It used to respond to security authentication requests within a computer domain.
	<b>authentication server</b>	(Credentials Processes in Windows Authentication 2016)	The authentication server performs the authentication of the credentials and instructs the authenticator to allow or reject the supplicant's traffic.
<b>WEB_ACCESS_CONTROL</b>	<b>pornography</b>	(Employee Computer Use Policies, n.d.), (Stanford Computer and Network Usage Policy 2014), (Internet And Email Access Policy 2015), (Sample internet usage policy, n.d.), (Acceptable Use Policy-Prohibited Activities 2019), (Sample Internet Policy, n.d.)	Pornographic content is restricted in most of the Internet policy, hence it categorized as Web Access Control.
	<b>monitoring</b>		'monitoring' in Internet policies usually refer to inspect the content of users' activity.
	<b>obscene</b>		Any obscene content will be restricted in most of the Internet policy and users are not allowed to use or send any obscene material.
	<b>prohibited</b>		Most of the Internet Policies used the term 'prohibited' to refer to some Internet activities are restricted within organizational network.
	<b>Internet activities</b>		Internet activities may restrict due to violation of organization policies.
	<b>personal use</b>		Some Internet policies may limit their users' personal

			use of the Internet during office hours.
	<b>gaming</b>		Gaming is prohibited in most of the Internet usage policy hence is categorized as one of the keywords for WEB_ACCESS_CONTROL intent.
	<b>non-work related</b>		Internet users are prohibited from initiating non-work-related Internet sessions using organizational information resources during working hours.
	<b>job-related activities</b>		Internet usage policy may restrict Internet users only allowed to engage in job-related activities during working hours.
	<b>Inspecting</b>		'Inspecting' in Internet policies represent monitor the users' activities on the Internet.
	<b>harassing</b>		Organization may restrict Internet users not send, view or download harassing messages or material that violate organization policy.
	<b>Internet access</b>		An organization may restrict or allow Internet access to a user or a specific group of users.
	<b>piracy</b>		Internet users are not allowed to download, copy pirated software and electronic files that

			are copyrighted or without authorization
	<b>illegal</b>		User must not access or download any material or content that are illegal.
	<b>offensive</b>		Internet user must not send any content that consists of offensive words.
	<b>fraudulent</b>		Internet policy may restrict a user from sending or accessing Fraudulent e-mail messages, web sites, and other fraudulent online materials.
<b>NETWORK_SECURITY</b>	<b>malicious network traffic</b>	(Malicious Traffic: Understanding What Does and Doesn't Belong on Your Network, n.d.), (Malicious Network Traffic, n.d.)	Malicious traffic is anomalous traffic that doesn't look normal which require the network administrator to take a countermeasure against unknown traffic.
	<b>Trojan Horse</b>	(What is a Trojan Virus, n.d.), (Avoiding a Trojan Virus: Keeping the Gates Closed, n.d.)	A type of malware that is often disguised as legitimate software, trying to gain access to users' systems. It required to take security countermeasures to prevent it.
	<b>prevent unauthorized access</b>	(How To Prevent unauthorized Computer Access 2018), (How To	Unauthorized access means invading the privacy of someone's computer

	<b>unauthorized network access</b>	Prevent unauthorized Computer Access 2019)	without their consent. Such a situation can arise if you leave your system without any security measure against malware and viruses.
	<b>Cyber threats</b>	(Tunggal 2020), (Taylor 2020)	The cyber threat is a possible danger that might exploit a vulnerability to breach security and seek to damage data, steal data, or disrupt the network operation. Hence, security measure is needed to prevent it.
	<b>cyber attack</b>	(What Are the Most Common Cyber Attacks, n.d.), (Fruhlinger 2020)	Cyber-attack is an attack launched from one or more computers against another computer, multiple computers or networks.
	<b>worm</b>	(Malware, n.d.), (What are malware, viruses, Spyware, and cookies, and what differentiates them, n.d.)	Viruses or worm is a type of program or file that can harm computer operation or devices, hence Network Security is needed to prevent it.
	<b>viruses</b>		
	<b>Denial of service</b>	(What Are the Most Common Cyber Attacks, n.d.), (What is a denial of service attack (DoS), n.d.)	An attack that shut down a machine or network, making it inaccessible to its intended users. It requires network security measure against these attacks.
	<b>mitigate threats</b>	(A Holistic Approach of Advanced Threat Mitigation 2019), (Rossi 2015)	Threat mitigation in Cyber Security refers to the prevention of unexpected security incidents when security attacks do happen.



	<b>block specific traffic</b>	(Internet Access Control, n.d), (Sangfor Internet Access Management, n.d)	Network access control is monitored by a network administrator and block any abnormal or unknown traffic if there trying to intrude to a network.
	<b>untrusted outside network</b>	(Danelle 2013), (Chapter 4 : Capturing Live Network Data, n.d.)	An untrusted outside network refers to any IP or traffic that is not within the organization and I not authorized.
<b>PERFORMANCE</b>	<b>Network Consumption</b>	(Hoffman 2019)	Refer to how much bandwidth and data do the devices on the network. It will slow down the network performance if one of the devices exceeds its bandwidth usage. Hence, it required bandwidth management to maintain network performance.
	<b>bandwidth utilization</b>		
	<b>monitoring bandwidth</b>	(Hoffman 2019) , (Bandwidth Monitor 3.4 build 757 released, n.d.)	Monitor network usages of computer it's installed on within the network, bandwidth usage included the download and upload speeds of an application over time.
	<b>limiting bandwidth</b>	(Hoffman 2019) , (Bandwidth Monitor 3.4 build 757 released, n.d.) (Sangfor Internet Access Management, n.d)	allocating specific traffic to a user or group of users. A network administrator may take necessary action to limit Internet usage for heavier users.

	<b>Download and upload speeds</b>	(Kinght 2014), (George 2019)	Download speed refers to the megabits of data per second it takes to download data from a server in the form of images, videos, text and more while upload speed refers to the megabits of data per second you can send information from your computer to another device or server over the Internet.
	<b>bandwidth resources</b>	(Bandwidth Monitor 3.4 build 757 released, n.d.)	Bandwidth resources are the maximum rate of data transfer across a given path. It must be controlled to maintain network performance.
	<b>alternative path</b>	(Deep & Karthik 2018)	Alternative paths in network refer to multipath routing which takes place if any of the single link failure. It can yield a variety of benefits such as load balancing and fault tolerance.
	<b>redirects routing</b>	(Jose & Somani 2003)	Multipath routing enables traffic to redirect if detect any of the single link failures. If a link failed, traffic rerouting for the restoration of the failed connections will be taking place to maintain the network performance.
	<b>distribute network traffic</b>	(What Is Load balancing?, n.d.)	The load balancer will distribute

			network traffic across multiple servers to optimize resource use, maximize throughput, minimize response time and avoid the overhead of one single resource.
<b>SELF_HEALING</b>	<b>interruption of network services</b>	(Reed 2019), (Hussain 2005)	Failure of Internet Service Provider or any of the equipment may result in interruption of network services. Hence, a recovery plan is necessary for ensuring the network operation always running.
	<b>equipment failure</b>	(Introduction to Business Continuity, n.d.), (Prescient 2017)	Network equipment failure will result in network services down, hence, a necessary countermeasure to resolve this issue is needed.
	<b>downtime</b>	(What is Business Continuity 2017), (Reed 2019)	Network downtime may result in enterprise suffer a huge amount of loss. Hence, a business continuity plan is needed.
	<b>Disaster</b>	(Rouse, n.d.), (Reed 2019)	Any form of disaster will affect the operation of a network. Hence, an IT organization should enact automatic disaster recovery plans such as hosting data backups on redundant servers for data resilience.
	<b>failover</b>	(Failover 2017), (Tobias 2019)	Failover is the constant capability

			to automatically and seamlessly switch to a highly reliable backup. The main purpose of failover is to eliminate, or at least reduce, the impact on users when a system failure occurs.
	<b>business continuity</b>	(Introduction to Business Continuity, n.d.), (What is Business Continuity 2017)	Business continuity is having a plan to ensure operations are not disrupted by a disaster or unplanned incident that takes critical systems offline.

**Table 3.2.1 Table for specifying the source and explanation of the keywords for each of the intents.**

WEB_ACCESS_CONTROL	Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that are a violation of applicable law
WEB_ACCESS_CONTROL	Every employee is responsible for the content of all text, audio, video or image files that he or she places or sends over the company's Internet and e-mail systems.
WEB_ACCESS_CONTROL	Inspecting and monitoring information and information resources may be required for the purposes of enforcing this policy, conducting University investigations or audits,
WEB_ACCESS_CONTROL	Company employees are expected to use the Internet responsibly and productively. Internet access is limited to job-related activities only and personal use is not permitted.
WEB_ACCESS_CONTROL	Internet access is provided primarily to use for the Trust's business. Limited personal use is permitted, but not for personal financial gain, upon the condition that it is done
WEB_ACCESS_CONTROL	This policy refers to monitoring user activity on the Internet, whether accessed via a PC, a shared departmental computer system or Internet café facilities provided by the Trust.
WEB_ACCESS_CONTROL	Staff must not access, transmit or download any offensive, obscene or indecent images, data or other material, or any data capable of being transformed into obscene or indecent
WEB_ACCESS_CONTROL	Staff must not access, transmit or download any content with obscene language, pornography, hostile material relating to gender, sex, race, sexual orientation, religious, political
WEB_ACCESS_CONTROL	All the employee within the company must comply with company Internet policy that cannot access any content with obscene, pornographic, threatening, or other messages or material
WEB_ACCESS_CONTROL	While it's common for employers and employees to use the Internet, your business will benefit having some regulations in place, including rules regarding personal use of the Internet.
WEB_ACCESS_CONTROL	Email is to be used for company business only. Please keep this in mind, also, as you consider forwarding non-work related emails to associates, family, or friends. Non-work related
WEB_ACCESS_CONTROL	Employees need to be wary of the content of all emails they may send. One email sent without thought as to the potential repercussions can have unintended consequences.
WEB_ACCESS_CONTROL	The Internet may not be accessed for personal use during normal hours of employment. Occasional use for personal reasons is allowed outside working hours, however the
WEB_ACCESS_CONTROL	The company/firm is particularly at risk when you have access to the Internet. The nature of the Internet makes it impossible to define all inappropriate use. However you are
WEB_ACCESS_CONTROL	The creation, internal or external transmission or downloading of any offensive, obscene or indecent images, or unlawful data or material is prohibited. Apart from that, gathering
WEB_ACCESS_CONTROL	Pornography (often shortened to porn) is the portrayal of sexual subject matter for the exclusive purpose of sexual arousal. Any employees watch any obscene content during work
WEB_ACCESS_CONTROL	Monitoring is not enough, indeed, controlling is also necessary. Surveilstar monitors all Internet activities including emails, programs, IM chats, screen snapshots, etc. Surveillance
WEB_ACCESS_CONTROL	Company employees are expected to use the Internet responsibly and productively. Internet access is limited to job-related activities only and personal use is not permitted.

**Figure 3.2.1 Data Sampling for label WEB\_ACCESS\_CONTROL**

AUTHENTICATION	Unless company staff have proper authorization, users must not attempt to modify or remove information or information resources that are owned or used by others.
AUTHENTICATION	Authorization is the function of specifying access rights/privileges to resources, which is related to information security and computer security in general and to access control in computer systems and networks.
AUTHENTICATION	Authorization in computer systems and networks rely on access policies. The access control process can be divided into the following phases: policy definition phase where access control policies are defined, policy enforcement phase where access requests are permitted or not permitted based on the policy, and policy monitoring phase where access control is monitored.
AUTHENTICATION	Most modern, multi-user operating systems include access control and thereby rely on authorization. Access control also uses authentication to verify the identity of consumers. Trusted consumers are often authorized for unrestricted access to resources on a system, but must be verified so that the access control system can make the access approval.
AUTHENTICATION	Authorization is the process of giving someone permission to do or have something. In multi-user computer systems, a system administrator defines for the system which users are authorized to access which resources. Authorization is seen as both the preliminary setting up of permissions by a system administrator and the actual checking of the permission values that have been set up when a user requests access to a resource.
AUTHENTICATION	Authorization is a security mechanism used to determine access levels or user/client privileges related to system resources including files, services, computer programs, data and other system resources.
AUTHENTICATION	Most web security systems are based on a two-step process. The first step is authentication, which ensures about the user identity and the second stage is authorization, which ensures that the user is authorized to access the resource.
AUTHENTICATION	Accessing company information that is not within the scope of one's work. This includes unauthorized reading of customer account information, unauthorized access of personal information, unauthorized use of company resources, and unauthorized disclosure of confidential information.
AUTHENTICATION	Authorization is the function of the policy definition phase which precedes the policy enforcement phase where access requests are permitted or not permitted based on the policy.
AUTHENTICATION	Authorization is a security mechanism used to determine user/client privileges or access levels related to system resources, including computer programs, files, services, data and other system resources.
AUTHENTICATION	In authentication, the user or computer has to prove its identity to the server or client. Usually, authentication by a server entails the use of a user name and password. Other authentication methods include digital certificates, biometrics, and smart cards.
AUTHENTICATION	Authorization is a process by which a server determines if the client has permission to use a resource or access a file. The type of authentication required for authorization may vary depending on the resource and the user.
AUTHENTICATION	Authorization is a process by which done through a username and password. The user enters their username, which allows the system to confirm their identity; this system may then prompt the user for a password.
AUTHENTICATION	Passwords aren't the only way to authenticate your users. While password authentication is the most common way to confirm a user's identity, it isn't even close to the most secure. Other authentication methods include digital certificates, biometrics, and smart cards.
AUTHENTICATION	Authentication confirms an online user's identity. Authentication means confirming your own identity, whereas authorization means being allowed access to the system. In authentication, the user or computer has to prove its identity to the server or client. Usually, authentication by a server entails the use of a user name and password. Other authentication methods include digital certificates, biometrics, and smart cards.
AUTHENTICATION	Authentication is about validating your credentials such as Username/User ID and password to verify your identity. The system then checks whether you are what you say you are. Authentication factors determine the many different elements the system uses to verify one's identity before granting the individual access to anything. An individual's identity is verified by a combination of factors such as something you know (password), something you have (smart card), and something you are (biometrics).
AUTHENTICATION	Two-Factor authentication requires a two-step verification process which not only requires a username and password, but also a piece of information only the user knows. U

Figure 3.2.2 Data Sampling for label AUTHENTICATION

PERFORMANCE	Staff offices are also monitored for bandwidth usage. The limit for office computers is typically one gigabyte of data per day (upload and download), although exceptions can be made.
PERFORMANCE	Bandwidth both within the company and in connecting to the Internet is a shared, finite resource. Users must make reasonable efforts to use this resource in ways that do not negatively impact the performance of the network.
PERFORMANCE	Network bandwidth is used when a person uploads or downloads data on the Internet. Network administrator will monitor bandwidth consumption to make sure that this shared resource is used efficiently.
PERFORMANCE	Download and upload speeds to the Internet are limited to 2 Mbps in order to provide a consistent quality of service for all devices on the network. Faculty and staff offices are also monitored for bandwidth usage.
PERFORMANCE	Network bandwidth is the capacity of a wired or wireless network communications link to transmit the maximum amount of data from one point to another over a computer network or other telecommunications system.
PERFORMANCE	All the user within the network must be aware of its network bandwidth usage. All the download and upload speeds to Internet are limited to 5Mbps per person to ensure network stability.
PERFORMANCE	Network bandwidth is used when a person uploads or downloads data to or from the Internet. Computing Services monitors university data network and Internet bandwidth consumption.
PERFORMANCE	Staff offices are also monitored for bandwidth usage. The limit for office computers is typically one gigabyte of data per day (upload and download), although exceptions can be made.
PERFORMANCE	Bandwidth both within the company and in connecting to the Internet is a shared, finite resource. Users must make reasonable efforts to use this resource in ways that do not negatively impact the performance of the network.
PERFORMANCE	Network bandwidth is used when a person uploads or downloads data on the Internet. Network administrator will monitor bandwidth consumption to make sure that this shared resource is used efficiently.
PERFORMANCE	Download and upload speeds to the Internet are limited to 2 Mbps in order to provide a consistent quality of service for all devices on the network. Faculty and staff offices are also monitored for bandwidth usage.
PERFORMANCE	Network bandwidth is the capacity of a wired or wireless network communications link to transmit the maximum amount of data from one point to another over a computer network or other telecommunications system.
PERFORMANCE	All the user within the network must be aware of its network bandwidth usage. All the download and upload speeds to Internet are limited to 5Mbps per person to ensure network stability.
PERFORMANCE	Network bandwidth is used when a person uploads or downloads data to or from the Internet. Computing Services monitors university data network and Internet bandwidth consumption.
PERFORMANCE	While servers can be reconfigured if needed, it's best to calculate your company's bandwidth requirements at the start to avoid major overhauls down the line. This means determining bandwidth requirements for all devices on the network.
PERFORMANCE	Sangfor IAM improves bandwidth utilization by more than 30% using three unique major traffic management solutions. Dynamic Traffic Control automatically adjusts traffic control policies based on network conditions.
PERFORMANCE	With viruses and malware often consuming out of the ordinary amounts of bandwidth, monitoring bandwidth utilization can also be invaluable in identifying security anomalies. By monitoring bandwidth usage, network administrators can identify unusual patterns of activity.
PERFORMANCE	Network bandwidth is used when a person uploads or downloads data to or from the Internet. Computing Services monitors university data network and Internet bandwidth consumption.
PERFORMANCE	While servers can be reconfigured if needed, it's best to calculate your company's bandwidth requirements at the start to avoid major overhauls down the line. This means determining bandwidth requirements for all devices on the network.

Figure 3.2.3 Data Sampling for label PERFORMANCE

SELF_HEALING	Any interruption of network services can affect an organization's ability to access, collect or use data and communicate with staff, partners and customers. Interruptions put
SELF_HEALING	Network equipment failure such as routers, switches, modems, gateways, or any other device can affect the performance of all other devices connected to them. It can cause a single point of failure.
SELF_HEALING	A single network consists of multiple routers, nodes, or switches. Failure of one network components might become overloaded and stop working, which can trigger a cascade of failures.
SELF_HEALING	Disasters of any type can significantly damage or even destroy your production center and virtual infrastructure, thus causing significant business losses. Any interruption of network services can affect an organization's ability to access, collect or use data and communicate with staff, partners and customers. Interruptions put
SELF_HEALING	Data centers must be able to quickly recover from equipment failure. This is especially true for applications such as e-commerce where a significant amount of money can be lost.
SELF_HEALING	There are three principles of systems design in reliability engineering which can help achieve high availability. First, we need to eliminate single points of failure. This means designing systems that exhibit truly continuous availability.
SELF_HEALING	Systems that exhibit truly continuous availability are comparatively rare and higher priced, and most have carefully implemented specialty designs that eliminate any single point of failure.
SELF_HEALING	A single point of failure is a part of a system that, if it fails, will stop the entire system from working. SPOFs are undesirable in any system with a goal of high availability or reliability.
SELF_HEALING	A highly available system should be able to quickly recover from any sort of failure state to minimize interruptions for the end user. Best practices for achieving high availability include redundancy and failover.
SELF_HEALING	The term downtime is used to refer to periods when a system is unavailable. Downtime or outage duration refers to a period of time that a system fails to provide or perform its intended function.
SELF_HEALING	Network downtime is never good – when IT stops, business stops. IT failure and unexpected downtime is one of the biggest issues in today digitally advanced business environment.
SELF_HEALING	Disaster recovery (DR) is an area of security planning that aims to protect an organization from the effects of significant negative events. Having a disaster recovery strategy is essential for business continuity.
SELF_HEALING	The goal of disaster recovery is for a business to continue operating as close to normal as possible. The disaster recovery process includes planning and testing and might involve backup and recovery of data.
SELF_HEALING	The need for the Spanning Tree Protocol (STP) arose because switches in local area networks (LANs) are often interconnected using redundant links to improve resilience and avoid loops.
SELF_HEALING	STP is a Layer 2 link management protocol that provides path redundancy while preventing loops in the network. For a Layer 2 network to function properly, only one active path should exist between any two network devices.
SELF_HEALING	Network redundancy is a process through which additional or alternate instances of network devices, equipment and communication mediums are installed within network infrastructure.
SELF_HEALING	Network redundancy is primarily implemented in enterprise network infrastructure to provide a redundant source of network communications to ensure network availability and reliability.
SELF_HEALING	Network redundancy is a process through which additional or alternate network path, equipment and communication mediums are installed within network infrastructure.
SELF_HEALING	Redundant links are used to prevent nasty network failure. These are used to provide redundancy, i.e. back up when a link fails, i.e. a frame can be forwarded out through an alternate path.
SELF_HEALING	Network path redundancy between cluster nodes is important for vSphere HA reliability. A single management network ends up being a single point of failure and can result in a complete loss of management.
SELF_HEALING	Any interruption of network services can affect an organization's ability to access, collect or use data and communicate with staff, partners and customers. Interruptions put

Figure 3.2.4 Data Sampling for label SELF\_HEALING

NETWORK_SECURITY	Central campus network and security personnel must take immediate action to mitigate threats that have the potential to pose a serious risk to campus information system resource
NETWORK_SECURITY	Unauthorized network access – An attacker needs access before being able to perform any attacks. An attacker can be a disgruntled employee, an employee that has become a victim
NETWORK_SECURITY	Unauthorized network access is accessing your network with intrusion techniques. Whether it be to protect yourself from malware or ensure that your private information is safe, h
NETWORK_SECURITY	A firewall is a network Security device that monitor incoming and outgoing network traffic and devices whether to allow or block specific traffic based on a defined set of Security ru
NETWORK_SECURITY	In computing, a firewall is software or firmware that enforce a set of rule about what data packet will be allowed to enter or leave a network. Firewall are incorporated into a wIDSe
NETWORK_SECURITY	A firewall is a network Security System that monitor and control over all your incoming and outgoing network traffic based on advanced and a defined set of Security rule. A firewall
NETWORK_SECURITY	A firewall is a network Security System designed to prevent unauthorized access to or from a private network. Firewall can be implemented a both hardware and software, or a com
NETWORK_SECURITY	Any hardware and/or software designed to examine network traffic using policy tatement (ruleset) to block unauthorized access while permitting authorized communication to or f
NETWORK_SECURITY	A firewall is a network Security device that monitor incoming and outgoing network traffic and devices whether to allow or block specific traffic based on a defined set of Security ru
NETWORK_SECURITY	A network firewall protect a network from unauthorized access. It might take the form of a hardware device, a software program, or a combination of the two. network firewall gua
NETWORK_SECURITY	Almost all network security systems operate by allowing selective use of services. An ACL or Access control list is a common means by which access to and deny any unwanted acces
NETWORK_SECURITY	Cisco Access Control Lists are the set of conditions grouped together by name or number. These conditions are used in filtering the traffic passing from router. Through these condi
NETWORK_SECURITY	Besides filtering unwanted access, ACLs are used for several other purposes such as prioritizing traffic for QoS (Quality of Services), triggering alert, restricting remote access, debug
NETWORK_SECURITY	A firewall is a system designed to prevent unauthorized access to or from a private network. You can implement a firewall in either hardware or software form, or a combination of l
NETWORK_SECURITY	Malicious traffic or malicious network traffic is any suspicious link, file or connection that is being created or received over the network. Malicious traffic is a threat that creates an ir
NETWORK_SECURITY	Malicious traffic detection technology continuously monitors traffic for possible signs of any suspicious links, files, or connections created or received. In order to identify malicious
NETWORK_SECURITY	Malicious traffic detection technology continuously monitors traffic for possible signs of any suspicious links, files, or connections created or received. In order to identify malicious
NETWORK_SECURITY	Malicious network traffic activity can refer to a number of different behaviors that involve abnormal access patterns, database activities, file changes, and other out-of-the-ordinary
NETWORK_SECURITY	A cyber or cybersecurity threat is a malicious act that seeks to damage data, steal data, or disrupt digital life in general. Cyber threats include computer viruses, data breaches, Denia

**Figure 3.2.5 Data Sampling for label NETWORK\_SECURITY**

After gather all the sampling data, the dataset was stored in .csv file format. the following step is prepare an entity list and a training documents. To achieve this, Python programming language will be used to structure the data. The high-level programming language able to assist in structuring the dataset provided it has comprehensive libraries and tools. The python editor that used to write all the codes is Jupyter Notebook . It embedded with the latest version of Python which is 3.8. After having all the tools, we import the required libraries in Python.

```
import csv
import codecs
```

**Figure 3.2.6 The Python libraries needed for processing the dataset**

The csv library was used to modify the data in csv file while the Codec library was used to encode and decode when import or export a file. The encoding format that we needed is 'UTF-8' as it is the format that Amazon Comprehend used.

```
dataset=[]

with open('FullDataset.csv','r') as csv_file:
    csv_reader=csv.reader(csv_file)

    for line in csv_reader:
        dataset.append(f"{line[1]}")
```

**Figure 3.2.7 The import of the full dataset.**

We import the '.csv' file which consists of the label and its sample text. After that, we extract out the sample text for converting into training documents.



```
In [3]: dataset
```

```
Out[3]: ['Unless company staff have proper authorization, users must not attempt to modify or remove information or information resources that are owned or used by others. ',
'Authorization is the function of specifying access rights/privileges to resources, which is related to information security and computer security in general and to access control in particular. For example, human resources staff are normally authorized to access employee records and this policy is usually formalized as access control rules in a computer system.',
'Authorization in computer systems and networks rely on access policies. The access control process can be divided into the following phases: policy definition phase where access is authorized, and policy enforcement phase where access requests are approved or disapproved. Authorization is the function of the policy definition phase which precedes the policy enforcement phase where access requests are approved or disapproved based on the previously defined authorizations.',
'Most modern, multi-user operating systems include access control and thereby rely on authorization. Access control also uses authentication to verify the identity of consumers. When a consumer tries to access a resource, the access control process checks that the consumer has been authorized to use that resource. Authorization is the responsibility of an authority, such as a department manager, within the application domain, but is often delegated to a custodian such as a system administrator. Authorizations are expressed as access policies in some types of "policy definition application", e.g. in the form of an access control list or a capability, or a policy administration point ',
'Trusted consumers are often authorized for unrestricted access to resources on a system, but must be verified so that the access control system can make the access approval decision. "Partially trusted" and guests will often have restricted authorization in order to protect resources against improper access and usage. The access policy in some operating systems, by default, grant all consumers full access to all resources. Others do the opposite, insisting that the administrator explicitly authorizes a consumer to use each resource.',
'Authorization is the process of giving someone permission to do or have something. In multi-user computer systems, a system administrator defines for the system which users are allowed to access the system and what privileges of use (such as access to which file directories, hours of access, amount of allocated storage space, and so forth).',
'Authorization is seen as both the preliminary setting up of permissions by a system administrator and the actual checking of
```

**Figure 3.2.8 The sample text for the 5 intents.**

Amazon Comprehend required the training documents is one documents per file (a single training set was stored in a single file), thus we need to convert the sample text in the '.csv' file into a single text file. All the converted training documents will be in *UTF-8* format.

```
x=0
for line in dataset:
    x+=1
    with codecs.open('Doc' + str(x) + '.txt', 'w',"utf-8") as output:
        output.write(line)
```

**Figure 3.2.9 Converting the sample text into a single text file in Python.**

Doc1.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc2.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc3.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc4.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc5.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc6.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc7.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc8.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc9.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc10.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc11.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc12.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc13.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc14.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc15.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc16.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc17.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc18.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc19.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc20.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc21.txt	3/2/2020 10:41 AM	Text Document	1 KB
Doc22.txt	3/2/2020 10:41 AM	Text Document	1 KB

**Figure 3.2.10 The converted text file is stored in the same directory with the Jupyter program file.**

Next, we need to prepare an entity list for the model to train and the model able to map the classifier(intent) to the related sample training documents. The entity list will only consist of the labels and its keywords.

authorized	AUTHENTICATION	
verify	AUTHENTICATION	
verification	AUTHENTICATION	
permissions	AUTHENTICATION	
Identity	AUTHENTICATION	
domain controller	AUTHENTICATION	
authentication server	AUTHENTICATION	
username and password	AUTHENTICATION	
User Identity Verification	AUTHENTICATION	
pornographic	WEB_ACCESS_CONTROL	
monitored	WEB_ACCESS_CONTROL	
monitoring	WEB_ACCESS_CONTROL	
gaming	WEB_ACCESS_CONTROL	
pornography	WEB_ACCESS_CONTROL	
prohitbited	WEB_ACCESS_CONTROL	
Internet activities	WEB_ACCESS_CONTROL	
personal use	WEB_ACCESS_CONTROL	
obscene	WEB_ACCESS_CONTROL	

**Figure 3.2.11 Entity list which consists of keywords and its label (intent).**

After the training documents and entity list are prepared, we can now begin to train the intent recognizer. The both documents have to be stored in Amazon S3 storage for the system to locate the file when using Amazon Comprehend services.

Upload



+ Create folder

Download

Actions

US East (N. Virginia)

Viewing 1 to 2

<input type="checkbox"/>	Name	Last modified	Size	Storage class
<input type="checkbox"/>	 dataset	--	--	--
<input type="checkbox"/>	 NewEntityList.csv	Feb 3, 2020 11:03:55 PM GMT+0800	3.0 KB	Standard

Viewing 1 to 2

**Figure 3.2.12 Storing of both training documents and entity list.**

As mentioned earlier, Amazon Comprehend use topic modeling to categorized the words into different labels. The service uses the techniques in NLP named Latent Dirichlet Allocation (LDA) to generate topics based on the mapping of topics and related documents. (Foltz 1996) We utilize the technique by using it to determine the intent of a policy.

After proceed to Amazon Comprehend, we first need to fill-up the form with details for the recognizer to map the labels with the related training documents. As shown in the diagram



3.2.13, we need to give a name to the recognizer and provide the label that wants the recognizer to be identified.

Recognizer name

intentRecognizer

Then name has to be unique and can be from 1 to 64 characters. Valid characters are a-z, A-Z, 0-9 and hyphen (-).

Custom entity type [Info](#)

A custom label or labels you want the recognizer to identify in your dataset. The entity type(s) must match one of the types in the annotations or entity list.

AUTHENTICATION [Remove type](#)

NETWORK\_SECURITY [Remove type](#)

WEB\_ACCESS\_CONTROL [Remove type](#)

SELF\_HEALING [Remove type](#)

PERFORMANCE [Remove type](#)

[Add type](#)

You can add 7 more entity type(s).  
The entity type must be upper-case and underscore-separated

**Figure 3.2.13 Give the name for the recognizer and define the custom entity type.**

Next, we have to specify which method are we using for training the intent recognizer. There are two options for the training type (shown in Figure 3.2.14). As for our case, ‘*entity list and training docs*’ will be the method for training our model. We also need to provide the entity list and training documents.

Training data [Info](#)

Training type

☐ Using annotations and training docs  
Annotations need to contain the following columns: file; line; begin offset; end offset; type.  
Example

File	Line	Begin Offset	End Offset	Type
documents.txt	0	0	12	ENGINEER
documents.txt	1	0	5	ENGINEER
documents.txt	3	25	30	MANAGER

At least 200 annotations per entity type are required

☒ Using entity list and training docs  
Entity examples need to contain the following columns: text; type.  
Example

Text	Type
Jo Brown	ENGINEER
John Doe	ENGINEER
Jane Smith	MANAGER

At least 200 matches per entity type are required

Entity list location on S3  
Paste the URL of an input data file in S3, or select a bucket or folder location in S3

s3://jsnewdataset/NewEntityList.csv [Browse S3](#)

Must contain the custom entity type you provided above. File(s) must be in csv format.

Training documents location on S3  
Paste the URL of an input data file in S3, or select a bucket or folder location in S3

s3://jsnewdataset/dataset [Browse S3](#)

**Figure 3.2.14 Specify the method used in training the recognizer and the s3 path for storing the entity list and the training documents.**

The training of the intent classifier takes time as it depends on the number of training documents provided. Once the classifier is successfully built, it will show the following result.

Recognizer details			
Name	ARN	Training started	Training documents
intentRecognizer	arn:aws:comprehend:us-east-1:367639164540:entity-recognizer/intentRecognizer	2/3/2020, 11:34:42 PM	s3://jsnewdataset/dataset <a href="#">↗</a>
Status		Training ended	Entity List
🟢 Trained		2/4/2020, 12:01:31 AM	s3://jsnewdataset/NewEntityList.csv <a href="#">↗</a>
Custom entity type	Number of trained documents	Language	Recognizer encryption
AUTHENTICATION, NETWORK_SECURITY, WEB_ACCESS_CONTROL, SELF_HEALING, PERFORMANCE	840	English	-
	Number of test documents		
	105		

Recognizer performance <a href="#">Info</a>		
Precision	Recall	F1 score
100	100	100

**Figure 3.2.15 The intent classifier has successfully built.**

To test the recognition model, a sample policy was input into the classifier that just trained through the Comprehend portal (*Analysis job -> create job*) and the computational result will need to wait for some time for the model to compute.

Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that are a violation of applicable law or company policy. Unless company staff have proper authorization, users must not attempt to modify or remove information or information resources that are owned or used by others.

**Figure 3.2.16 The sample input.**

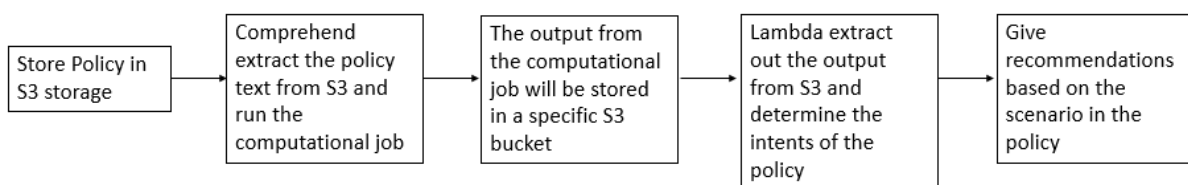
As shown in Figure 3.2.17, the intent classifier will classifier the keywords into different labels. For each of the keywords, it will specify the begin offset and end offset of the words from the text and the precision of the keywords is properly labeled.

```
{
  "Entities": [
    {
      "BeginOffset": 38,
      "EndOffset": 48,
      "Score": 0.9999929666519165,
      "Text": "fraudulent",
      "Type": "WEB_ACCESS_CONTROL"
    },
    {
      "BeginOffset": 50,
      "EndOffset": 59,
      "Score": 0.9999924898147583,
      "Text": "harassing",
      "Type": "WEB_ACCESS_CONTROL"
    },
    {
      "BeginOffset": 61,
      "EndOffset": 68,
      "Score": 0.9999842643737793,
      "Text": "obscene",
      "Type": "WEB_ACCESS_CONTROL"
    },
    {
      "BeginOffset": 76,
      "EndOffset": 88,
      "Score": 0.9999842643737793,
      "Text": "pornographic",
      "Type": "WEB_ACCESS_CONTROL"
    },
    {
      "BeginOffset": 225,
      "EndOffset": 238,
      "Score": 0.9998327493667603,
      "Text": "authorization",
      "Type": "AUTHENTICATION"
    }
  ],
  "File": "NewlyCreated2.txt"
}
```

**Figure 3.2.17 The output from the intent classifier.**

### 3.3 Computational function to look for intent through Lambda function

Even though the input can be manually sent to intent classifier for computing, it will be troublesome for the user as they first need to store their input file in S3 bucket, after that have to go through Amazon Comprehend portal for running an analysis job and need to specify where the input file stored and define the role for running the tasks. To simplify the tasks, we make all the tasks running in a single platform via Amazon Lambda. Hence, the user does not have to go through multiple platforms for running different tasks as everything was written in Lambda functions.



**Figure 3.3.1 The workflow of the intent-based recommendation model in AWS**

From the figure shown previously, the first step for the workflow will be storing the policy in S3 storage. Hence, in Lambda, we write a function shown below:

#### 1. Store Policy in S3 Storage

```

1  import boto3
2
3  def lambda_handler(event, context):
4      s3 = boto3.resource("s3")
5
6      Data=event['data'];
7      encoded_string = Data.encode("utf-8")
8
9      s3.Bucket("datahello").put_object(Key="NewlyCreated.txt",Body=encoded_string)
10
11     return "You're doing great"
  
```

**Figure 3.3.2 The lambda function for putting an object into a S3 bucket.**

We first need to import the boto3 library into the Lambda function. Boto3 is SDK for Amazon Web Services in Python. It allows programmer to configure and manage services in AWS such as S3 and DynamoDB. After the import of boto3 library, we declared a variable `s3` and define that variable as a connection to AWS S3 services by writing the line of code `'s3=boto3.resource("s3")'`.

After having a connection, a variable '*Data*' is declared to map the input. The input is called '*event*' in Amazon Lambda and the parameter '*data*' is the properties inside. After get the input, the line of code '*encoded\_string = Data.encode("utf-8")*' will encode the input into utf-8 format.

```
1 {
2   "data": "The internet may not be accessed for personal use during normal hours of employment."
3 }
```

**Figure 3.3.3 The input of the policy.**

s3 connection variable that declared earlier and specifies the S3 bucket, the name of the text file in the destination path, and the content of the text file (the input).

***s3.Bucket("datahello").put\_object(Key="NewlyCreated.txt",Body=encoded\_string)***

**S3 bucket name**                      **Name of the file in the destination path**                      **Content of the file**

## 2. Comprehend extract text file from S3 and run the computational job

After storing the input into S3 storage, now is turn for the intent classifier to run the computational task and analyse the policy text. To trigger the Amazon Comprehend services and let the classifier runs the task, the following code was written:

```
import boto3
from pprint import pprint
def lambda_handler(event, context):

    comprehend = boto3.client("comprehend")
    response = comprehend.start_entities_detection_job(
        InputDataConfig={
            'S3Uri': 's3://datahello/NewlyCreated.txt',
            'InputFormat': 'ONE_DOC_PER_FILE'
        },
        OutputDataConfig={
            'S3Uri': 's3://datahello',
        },
        JobName='ComprehendjobTest',
        EntityRecognizerArn="arn:aws:comprehend:us-east-1:367639164540:entity-recognizer/intentRecognizer",
        LanguageCode='en',
        DataAccessRoleArn="arn:aws:iam::367639164540:role/service-role/AmazonComprehendServiceRole-entity-recognizer",
        ClientRequestToken='autogenerated'
    )

    pprint(response)
```

**Figure 3.3.4 The full Lambda function for trigger Amazon Comprehend.**

Same as previous code, we first import boto3 library and define a variable used for connection to Amazon Comprehend services. After that, we trigger one of the methods in Amazon Comprehend which is `'start_entities_detection_job'` to start the computational job. In the method, we first specify the path of the input file and the input format. In our case, the input format is `ONE_DOC_PER_FILE` where each of the files is considered a separate document. After that, the output file path needs to specify the URI.

After having both input and output, other important attributes need to specify are `'JobName'`, `'EntityRecognizerArn'`, `'LanguageCode'`, `'DataAccessRoleArn'`, and `'clientRequestToken'`. `'JobName'` is the identifier you give for the computational job. `'EntityRecognizerArn'` is an attribute that identifies the specific entity recognizer to be used by the method. At the end of the code, there is a line of code `'pprint(response)'` report the status of the task.

After the task starts running, it will be shown at the Comprehend terminal following the details defined earlier.

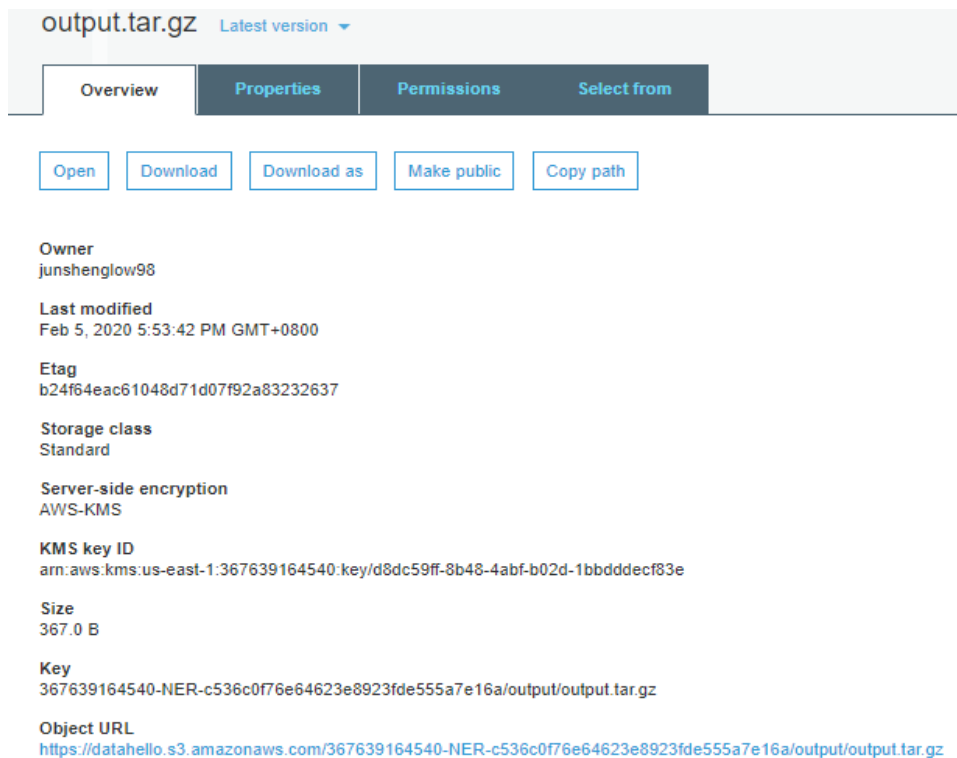
ComprehendjobTest

Copy

Job details

Name	Analysis type	Input data location
ComprehendjobTest	customEntities	s3://datahello/NewlyCreated2.txt <a href="#">🔗</a>
Status	Start	End
✔ Completed	2/5/2020, 5:45:29 PM	2/5/2020, 5:54:08 PM
ID	Recognizer arn	Recognizer name
c536c0f76e64623e8923fde555a7e16a	arn:aws:comprehend:us-east-1:367639164540:entity-recognizer/intentRecognizer	intentRecognizer
Job encryption		
-		

**Figure 3.3.5** The computational job status in Comprehend Analysis job portal.



**Figure 3.3.6** The output from the intent classifier model will be stored at a predefined S3 path.

### 3. Determine the intent of the policy

After the Comprehend recognizer completes its computational process, it will return a zipped file which is in tar.gz format. To be able to read it, the file must unzip and decompress first. The following python code able to unzip the file, read the content of the file and create a new text file which having the same content with the input text file.

```
import boto3
import tarfile
from io import BytesIO

def lambda_handler(event, context):

    s3=boto3.client('s3') #Connection to s3

    bucket = 'datahello' #Target bucket
    key = '367639164540-NER-a7afebb6bf099abb585302d9f3da07b4/output/output.tar.gz' #Target location
    response = s3.get_object(Bucket=bucket, Key=key) #Get the file from S3 storage

    input_tar_content = response['Body'].read() #Read the content of the file

    #Read the file content of the zipped file and extract it
    with tarfile.open(fileobj = BytesIO(input_tar_content)) as tar:
        for tar_resource in tar:
            if (tar_resource.isfile()):
                bytes_content = tar.extractfile(tar_resource).read() #Read the content of extracted file
                #Upload a new file in a new S3 path
                s3.upload_fileobj(BytesIO(bytes_content), Bucket = bucket, Key = "NewOutput.txt")
```

**Figure 3.3.7** The function that can read the content of a compressed zipped file.

Same as previous code, we import the boto3 library and create a variable named 's3' as a connection to s3 storage services. Next, the library tarfile is used for reading and writing tar archives and io library used for dealing with various types of I/O including text I/O, binary I/O and raw I/O. The source file path must be first specified for the s3 path to extract the folder and read the content.

After reading the content of the zipped file, the function extracts the file and read the content in it. The content of the zipped file will be stored in 'bytes\_content' variable. After that, a plain text file was re-upload with the same content in the zipped file to s3 storage. The new plain text file was placed in a new s3 location with a new name "NewOutput.txt".

As the output of the intent classifier is in JSON format, the function will identify the main intent of the text, we extract out the labels and take the labels with the highest frequency as the policy main intention.

```
{'Entities': [{'BeginOffset': 38, 'EndOffset': 48, 'Score': 0.9999929666519165, 'Text': 'fraudulent', 'Type': 'WEB_ACCESS_CONTROL'}, {'BeginOffset': 50, 'EndOffset': 59, 'Score': 0.9999924898147583, 'Text': 'harassing', 'Type': 'WEB_ACCESS_CONTROL'}, {'BeginOffset': 61, 'EndOffset': 68, 'Score': 0.9999842643737793, 'Text': 'obscene', 'Type': 'WEB_ACCESS_CONTROL'}, {'BeginOffset': 76, 'EndOffset': 88, 'Score': 0.9999842643737793, 'Text': 'pomographic', 'Type': 'WEB_ACCESS_CONTROL'}, {'BeginOffset': 225, 'EndOffset': 238, 'Score': 0.9998327493667603, 'Text': 'authorization', 'Type': 'AUTHENTICATION'}], 'File': 'NewlyCreated2.txt'}
```

**Figure 3.3.8 The output from the intent classifier model.**

To extract the label from the output, we write the following code.

```
def lambda_handler(event, context):

    s3=boto3.client('s3') #Connection to S3 service

    bucket = 'testingset' #Specify the folder that store the file
    key = 'NewOutput.txt' #Specify the filename
    response = s3.get_object(Bucket=bucket, Key=key) #Look for the file by using s3 connection and store the content into response variable

    content = response['Body'].read()
    json_object = json.loads(content)
    obj=json_object['Entities']
    entities=[]
    entities.clear()
    for person in json_object['Entities']:
        entities.append(person['Type'])

    result=intention(entities)
```

**Figure 3.3.9 The function that able to extract the classified label.**

We first get the file from s3 storage and use the JSON library to read the content of the text file. After that, we extract the parameter "Type" which stores all the labels into an array called 'entities'. After getting all the labels, the array 'entities' will be passed to function 'intention' shown as following:

```
def intention(intentList):
    #Set each of the label as a counter and initialize them as 0
    WebAccessControl=0
    Authentication=0
    SelfHealing=0
    Performance=0
    NetworkSecurity=0

    #Increment by 1 if detected the same label in the array
    for counter in intentList:
        if counter=="WEB_ACCESS_CONTROL":
            WebAccessControl+=1
        if counter=="AUTHENTICATION":
            Authentication+=1
        if counter=="PERFORMANCE":
            Authentication+=1
        if counter=="NETWORK_SECURITY":
            NetworkSecurity+=1

    intentionList=[WebAccessControl,Authentication,SelfHealing,Performance,NetworkSecurity]

    #Return the label with the most frequency
    if(max(intentionList)==WebAccessControl):
        return "WEB_ACCESS_CONTROL"
    elif(max(intentionList)==Authentication):
        return "AUTHENTICATION"
    elif(max(intentionList)==SelfHealing):
        return "SELF_HEALING"
    elif (max(intentionList)==NetworkSecurity):
        return "NetworkSecurity"
    else:
        return "Performance"
```

**Figure 3.3.10 The function that looks for the intent.**

In the functions, five labels will be first set as a counter to calculate the frequency of the label appear in a text, it will loop through the array that passes into the function and the counter will increment by one if detected the matched label found in the array. After that, all the counter will be placed in an array named '*intentionList*'. In the end, the function will return the label with the highest frequency. The function was using term frequency technique to calculate for the main intent.

Execution result: succeeded (logs)

▼ Details

The area below shows the result returned by your function execution. [Learn more](#) about returning results from your function.

"WEB\_ACCESS\_CONTROL"

Summary	
Code SHA-256	Request ID
+FjFlo1TW4txnlh/HBKAu7cbwU1d5bYyUpOlooYqqaM=	ff850fa9-84f9-431d-8ae4-5f1e0d7bdc7c
Duration	Billed duration
1520.56 ms	1600 ms
Resources configured	Max memory used
128 MB	82 MB Init Duration: 242.06 ms

**Figure 3.3.11 The function returns the intent after the computation.**

The Lambda function will return the intent of the text if the code runs successfully.



#### 4. Recommendations are given based on policy

After having the intent, we can now compute the recommendation based on the intent that the functions have computed. For each of the intent, there is a list of recommendations and those recommendations were explained why it matches with the intents. Apart from that, we also ranked these recommendations based on its effectiveness. The list of recommendations shown as follow:

##### 1. AUTHENTICATION

###### I. Password-Based Authentication

**Source:** (Oracle, n.d.), (Password Authentication Protocol (PAP), n.d.)

**Explanation:** User must provide their username and password for the system to authenticate the user identity.

###### II. Proxy Authentication

**Source:** (David, Brian, Marjorie, Anshu & Sailu, n.d.), (Jackson, n.d.)

**Explanation:** Proxy serves as access-control devices to manage the traffic in a network. Proxy authentication will not show any content until the user provides valid access-permission credentials to the proxy.

###### III. Identity Access Management (IAM)

**Source:** (Martin, Waters 2018), (What Is IAM, n.d.)

**Explanation:** IAM is a services that able to control the permission of user to the assigned resources. The services able to authenticate who has the rights to use a specific resources. The system administrator provided with the tools to trace user activities, assign user privileges, and enforce policies on a given resources.

###### IV. Third-party authentication

**Source:** (Oracle 2017), (Sella 2016)

**Explanation:** Third-party authentication allow the user to log in to an account if they have been authenticated by an external mechanism. Usually, external party schemes for login include Email, Phone or Social Media. In this authentication mechanism, the system trusts the third-party mechanism has authenticated the user correctly and so they are authenticated to the system.

**Ranking:**

1. Password-Based Authentication
2. Proxy Authentication
3. Identity Access management
4. Third-party Authentication

Across the recommendations that we have to go through, Password-based authentication was ranked as the top recommendation as it is the most common and general approach for all the authentication policies. Next, proxy authentication was ranked number 2 for the intent as proxy could effectively authenticate a user with credentials and ability to control the traffic of a network. After this, the solution that ranked number 3 would be Identity Access management (IAM) as administrator able to authenticated and authorized users' role for using a specific resource, however, it would be troublesome for setting privileges for each of the users. Last but not least, third-party authentication could be one of the effective ways to authenticate users with a third party, but one of the drawbacks is it required the user to linked their account to any third-party software. Hence, we ranked it as the 4<sup>th</sup> approach among all the recommendations.

## **2. WEB\_ACCESS\_CONTROL**

### **I. Proxy Web Control**

**Source:** (Hughes 2004), (Petters 2019)

**Explanation:** Proxy server act as a middleman between the user and the Internet, it functions as an intermediary to filter or intercept any web traffic before the web server delivers the content to the user.

### **II. Website Blocking**

**Source:** (How Does DNS Filtering Work 2019)

**Explanation:** Commonly known as DNS filtering, it is a technique of blocking access to certain websites, webpages, or IP addresses. It consists of a list of predetermined websites that can or cannot be accessed.

### **III. Access Control List**

**Source:** (Access Control List (ACL) - What are They and How to Configure Them! 2020), (Configuring IP Access Lists 2007)

**Explanation:** Access Control List is a function that used to control incoming and outgoing traffic with a set of a predefined policies. It can either restricts, block or allows the packets coming in or out from source to destination. In web access control, Access Control List can specify which services port or domain names to block on a predefined statement.

### **IV. Active Monitoring**

**Source:** (Hein 2019), (Gold 2019)

**Explanation:** Active monitoring maintain complete visibility into the network by stimulates user behavior to determine potential network performance. It provides a real-time network traffic visualization and user able to evaluate the performance.

### **V. Firewall Blocking**

**Source:** (Barlett 2018)

**Explanation:** Blocking a specific application or website through a firewall to prevent any unauthorized application runs in a host or a network.

### **VI. Keyword Filtering**

**Source:** (Beyder 2019), (Volkman & Mathew 2010)

**Explanation:** Keyword filtering acts as a layer of protection to block particular keywords and deliver the relevant content to the user. A software program will look for the predetermined list of words before the content present to the user.

#### **Ranking:**

**1. Proxy Web Control**

**2. Website Blocking**

**3. Access Control List (ACL)**

**4. Active Monitoring**

**5. Firewall Blocking**

**6. Keyword Blocking**

Among all the approaches in web access control, proxy web control ranked as top approach as it not only can control access for each user that attempts to access user activities, it can also use to trace user activities. The following approach would be website blocking as it can effectively restrict user for accessing a certain website, however, there is a drawback that administrators would have to manually define all the restricted websites. Furthermore, we ranked Access Control List (ACL) 3rd as a network packet could be blocked according to the predefined ACL but before that, the administrator would have to know the mechanism of ACL. The following approaches Active Monitoring function by stimulating a real-time network and present to the system administrator the condition of the real network. It could not effective if detecting any abnormal traffic within the network but would not take any countermeasure until the administrator does so. The last two approaches would be firewall blocking as it could block any unauthorized application coming in and reached any host within a network. The last approach we ranked is keyword blocking which will block any content that matches with the predefined list of keywords. However, there is a drawback for this approach as the administrator has to prepare a long list of keywords.

### 3. NETWORK SECURITY

#### I. Antivirus Protection

**Source:** (Comodo Cybersecurity, n.d.)

**Explanation:** A software used to protect a host from malware such as viruses, computer worms, spyware, and other threatening programs. It can be used to scan and remove viruses from your host and protect your computer.

#### II. Firewall Blocking

**Source:** (What are Network Firewalls, n.d.)

**Explanation:** A firewall that places on a network able to block or mitigate unauthorized access to a private network. Any traffic that attempts to pass through firewall must comply with the predefined firewall policies.

#### III. Intrusion Prevention System (IPS)

**Source:** (What is an Intrusion Prevention System (IPS), n.d.)

**Explanation:** A network preventive measure that continuously monitors the network to detect and prevent identified threats. It reports to the system administrator once it detects some malicious incident.

#### **IV. Active Monitoring**

**Source:** (Hein 2019), (Gold 2019)

**Explanation:** Active monitoring maintain complete visibility into the network by stimulates user behavior to determine potential network performance. It can continuously report real-time information on network availability and performance.

#### **Ranking:**

- 1. Antivirus Protection**
- 2. Firewall Blocking**
- 3. Intrusion Prevention System (IPS)**
- 4. Active Monitoring**

Among all the security features, we ranked antivirus protection as the top recommendation as it is the most effective and simple way to secure a host from cyber-attack or any other malware. After this, the installment of firewall into a network would be the second choice for preventing any unidentified network traffic flowing into the internal network. As a network firewall has its signature database for all the malware, it could identify all the known threats and prevent it from enter to the network. The following countermeasure we ranked is IPS which has the same functionality as a firewall and able to identify the threat and prevent it. However, one of the most common problems with the IPS system is the detection of false positives or false negatives, this situation happens when the system attempts to block an abnormal activity on the network and assumes it is malicious. (Urrutia, Ierace, & Bassett 2015) This situation could be troublesome for an administrator as they always have to fine-tune for the detection of alarms. Next, active monitoring could be one of the solutions for network security measures, but it could only visualize the behavior of real users on a network and let the administrator take the preventive measure according to the data computed by the system. It doesn't examine actual users and data as well as taking preventive measures towards any malicious attack.

## 4. PERFORMANCE

### I. Proxy Web Control

**Source:** (Hughes 2004), (What are Proxy Servers, n.d.)

**Explanation:** A intermediary that connects traffic to its requested destination. It can be used to detect and control the usage for the Internet for individual users. Any traffic that attempts to pass through the network will also intercept by the proxy server.

### II. Bandwidth management

**Source:** (What is Bandwidth Management, n.d)

**Explanation:** Bandwidth Management allows system administrators to control the amount of bandwidth for each of the users. It can balance the bandwidth usage and ensure that there will be no network congestion due to the high usage of a particular user.

### III. Active Monitoring

**Source:** (Hein 2019), (Gold 2019)

**Explanation:** Active monitoring maintain complete visibility into the network by stimulates user behavior to determine potential network performance. It provides real-time network traffic visualization and user able to evaluate the performance.

#### Ranking:

1. Proxy Web Control
2. Bandwidth Management
3. Active Monitoring

Among 3 of the approaches for maintaining the performance of a network, we ranked proxy web control as number one as proxy could act as an intermediary between client and server and monitor Internet usage of the user, it could also block some restricted websites before the server delivers the content to the user. Next, bandwidth management could be effectively restricting Internet usage for each user for a limited amount of quota. The mechanism could ensure every Internet user will get the allocated bandwidth. However, the inflexibility of bandwidth allocation could leak to a shortage of network bandwidth due to

limited bandwidth quota. The last approach we proposed for the system is active monitoring, it could visualize the network traffic and network administrator could pump in any testing set to test the performance of the network by observing the output. It ranked as third as it functions as just monitoring the network and assist the administrator to come out with a better network solution.

## **5. Self-Healing**

### **I. Adopt High Availability Mechanism**

**Source:** (Rouse n.d), (High Availability (HA) 2017)

**Explanation:** The ability of a system to operate continuously without failure for a long duration. In this mechanism, the system will be tested in several cases and ensure that it can sustain the failure of any network components.

### **II. Device Failover**

**Source:** (Device failover, n.d.)

**Explanation:** In device failover, if the primary unit fails, it switches to one of the subordinates automatically takes the role of the primary unit and continues to operate in the same way as the failed primary unit. To ensure a high degree of reliability, the system administrator provides failover capability in servers, systems or networks to minimize the impact of network downtime.

### **III. Avoid Single Point of Failure**

**Source:** (Rouse, n.d), (Ambrosone, 2016)

**Explanation:** Single point of failure is any software or hardware component in the network that could render the whole system down when a component fails. To ensure the system more robust, a system administrator may add redundancy in all potential single points of failure.

### **IV. Perform Regular Network Maintenance**

**Source:** (Network Maintenance, n.d.), (Welcher 2018)

**Explanation:** Procedure that takes regularly to ensure the network up and running. The procedure includes installation of hardware and software, planning for future network growth, ensure the network infrastructure comply with company policies.

## V. Load Balancing

**Source:** (What is Load Balancing, n.d.)

**Explanation:** A process of distributing incoming network traffic across multiple servers to ensure no single server bears too much load. A health check will be conducted regularly on the servers by the load balancer to ensure they can handle the requests in a healthy status. Thus, efficiency increase when ensuring no server is overloaded, which could affect the performance.

### Ranking:

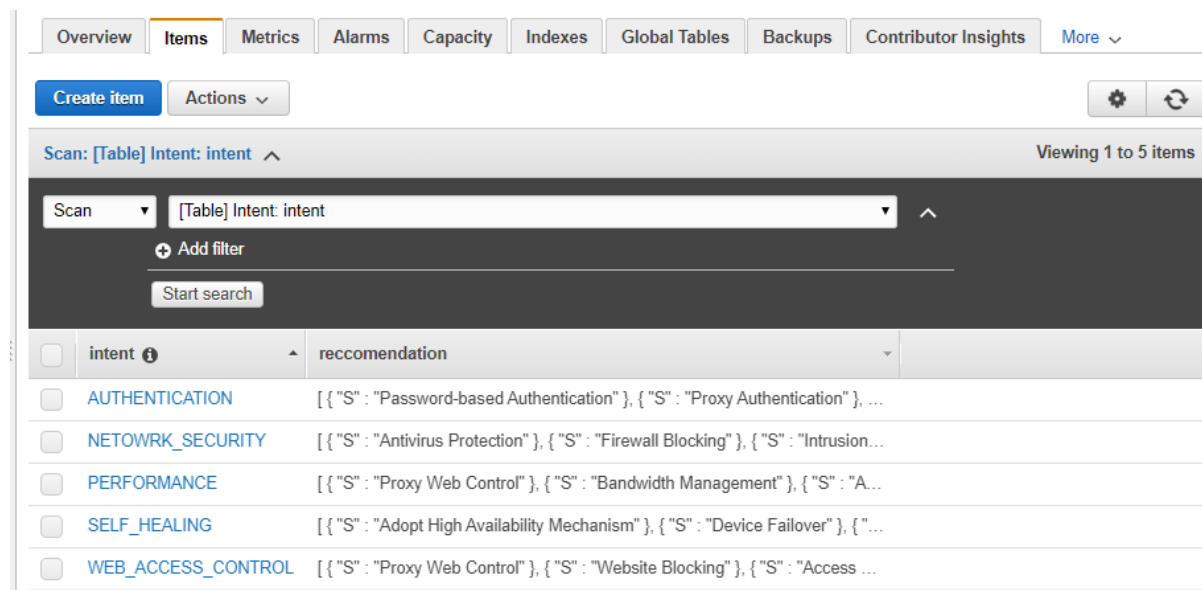
1. **Adopt High Availability Mechanism**
2. **Device Failover**
3. **Avoid Single Point of Failure**
4. **Perform Regular Network Maintenance**
5. **Load Balancing**

For Self-healing, adopt high availability could be one of the effective ways to maintain network performance as it has comprised other measures such as device failover and avoid single point of failure. High availability mechanism achieved a minimum amount of downtime by harnessing a group of clusters for device failover and make use of multiple processing points to avoid a single point of failure. The next solution the model would propose is device failover where the device operation will switch to another device with the same functionality if the primary unit down. The third proposed recommendations ‘avoid single point of failure’ has the same concept with the second approach ‘device failover’, but a single point of failure could refer to the bottleneck of the network and could stop the entire system from working if it fails. To eliminate single points of failure, each layer of the network stack must be prepared for redundancy to prevent the case when a central node down, others network nodes would be inaccessible. The next approach would be based on the human factors to perform regular network maintenance. Regular network maintenance is utmost importance in preventing any device failure and able to detect the problem early if having any issues with the network. However, as this is a human-based countermeasure, even though it can minimize the risk of network inaccessible, but it could not respond when there is a network service down. The last approach the system proposes is load balancing which could distribute traffic across multiple nodes to ensure each node won’t be handling too much loads. Besides, a load balancer can perform health checks regularly for all the devices to ensure that the device can handling tasks



in a healthy status. However, it comes with a drawback that it may result in unequal distribution when there is a small number of flows. (Celcer, Svigelj & Mohorcic 2008).

After having a list of recommendations, a table was created in AWS Dynamo DB service to store the list of recommendations. As per previous discussions, AWS Lambda functions will be the central processing point for compute the intent from a policy as well as give recommendations based on the scenario on that policy. After having the intent, the Lambda function will get the recommendations from Dynamo DB table. The recommendations in Dynamo DB table was listed according to the ranking.



intent	reccomendation
AUTHENTICATION	[{"S": "Password-based Authentication"}, {"S": "Proxy Authentication"}, ...]
NETOWRK_SECURITY	[{"S": "Antivirus Protection"}, {"S": "Firewall Blocking"}, {"S": "Intrusion..."}
PERFORMANCE	[{"S": "Proxy Web Control"}, {"S": "Bandwidth Management"}, {"S": "A..."}
SELF_HEALING	[{"S": "Adopt High Availability Mechanism"}, {"S": "Device Failover"}, {"S": "..."}]
WEB_ACCESS_CONTROL	[{"S": "Proxy Web Control"}, {"S": "Website Blocking"}, {"S": "Access ..."}]

**Figure 3.3.12 The table in Dynamo DB services**



```

Item {2}
  intent String : AUTHENTICATION
  reccomendation List [5]
    0 String : Password-based Authentication
    1 String : Proxy Authentication
    2 String : Identity Access Management (IAM)
    3 String : Third Party Authentication
    4 String : Firewall Blocking
  
```

**Figure 3.3.13 The details for attribute 'AUTHENICATION'**

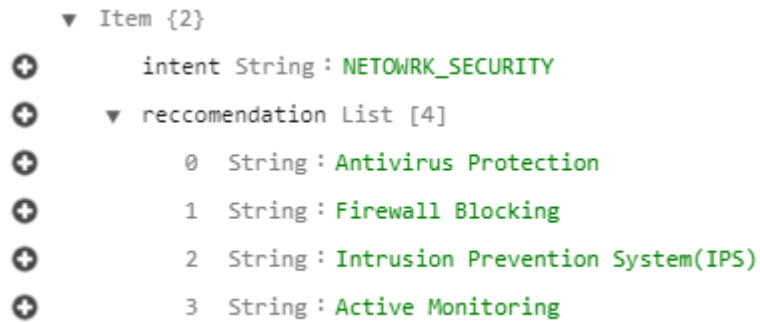


Figure 3.3.14 The details for attribute 'NETWORK\_SECURITY'

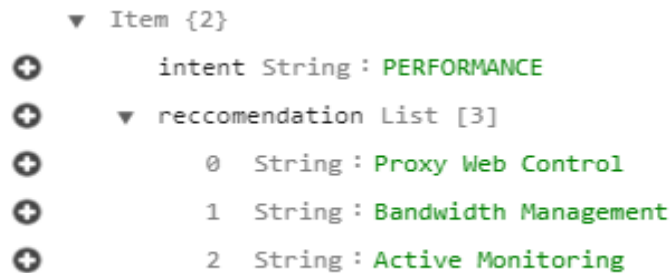


Figure 3.3.15 The details for attribute 'PERFORMANCE'

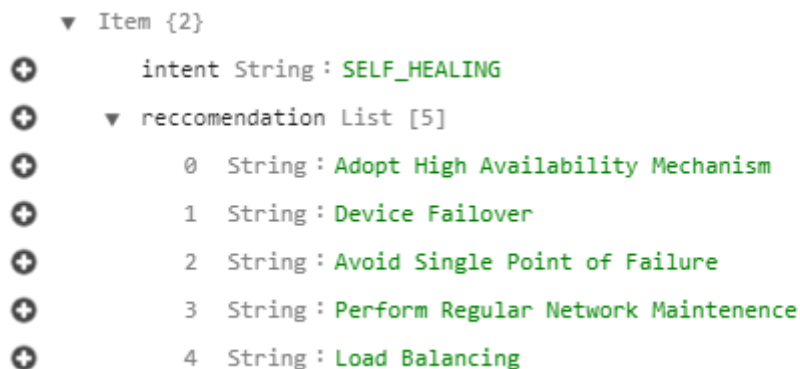


Figure 3.3.16 The details for attribute 'SELF\_HEALING'

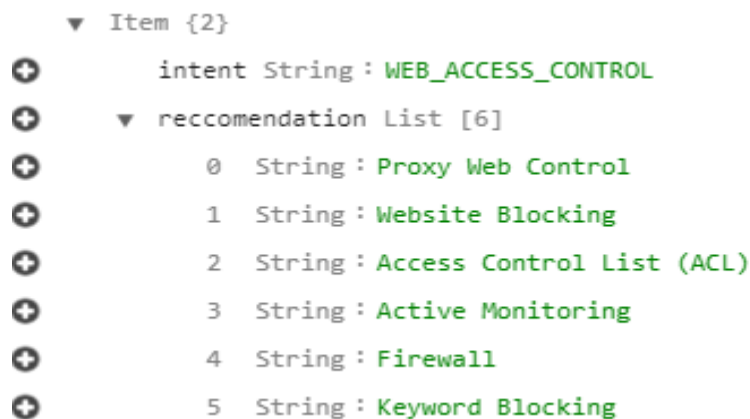


Figure 3.3.17 The details for attribute 'WEB\_ACCESS\_CONTROL'.

After having all the recommendations, we write the following code for Lambda extracts the recommendations based on intent.

```

TableName="Intent"
columnName= 'intent'

#Connection to Dynamo DB
db=boto3.resource('dynamodb')

#Look for the table and its column
table=db.Table(TableName)
response = table.get_item(Key={columnName:result})
response_obj=response['Item']
solution=response_obj['reccommendation']
print("We have a few reccommendations for your policy, you can perform the solution as follow:")
counter=1
for q in solution:
    print(str(counter) + "."+q)
    counter+=1

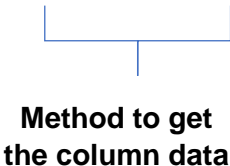
return solution;

```

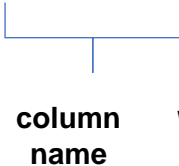
**Figure 3.3.18** The function that creates a connection to the database to extract the recommendations.

The function first creates a connection to Dynamo DB service by using the boto3 library, the connection object was stored in a variable named ‘db’. After that, we look for the table by specifying the table name with the line of code “table=db.Table(TableName)”. Next, we get the recommendations based on the result we get earlier with the following line of code:

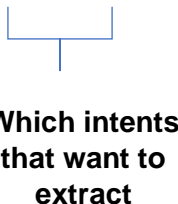
**response = table.get\_item(Key={columnName:result})**



**Method to get  
the column data**



**column  
name**



**Which intents  
that want to  
extract**

After getting the result set, we extract out the list of recommendations from the response object and print it out displayed to users.

```

START RequestId: 6cdc2ba3-6531-48e8-9530-898414e9736d Version: $LATEST
We have a few recommendations for your policy, you can perform the solution as follow:
1.Proxy Web Control
2.Website Blocking
3.Access Control List (ACL)
4.Active Monitoring
5.Firewall
6.Keyword Blocking
END RequestId: 6cdc2ba3-6531-48e8-9530-898414e9736d
REPORT RequestId: 6cdc2ba3-6531-48e8-9530-898414e9736d  Duration: 2295.41 ms    Billed Duration: 2300 ms    Memory Size: 128 MB    Max Memory Used: 78

```

**Figure 3.3.19** The output of the intent recommendation system.

## **CHAPTER 4 PRELIMINARY WORK**

Every time user input a policy, the recommendations model will analyze the text and give the recommendations based on the intent determined by the model. Different intents will have different list of recommendations. The list of recommendations has been pre-defined and we have ranked it according to its effectiveness. The following figures are the output of recommendations for each of the intents.

### **WEB\_ACCESS\_CONTROL**

```
START RequestId: 6cdc2ba3-6531-48e8-9530-898414e9736d Version: $LATEST
We have a few recommendations for your policy, you can perform the solution as follow:
1.Proxy Web Control
2.Website Blocking
3.Access Control List (ACL)
4.Active Monitoring
5.Firewall
6.Keyword Blocking
END RequestId: 6cdc2ba3-6531-48e8-9530-898414e9736d
REPORT RequestId: 6cdc2ba3-6531-48e8-9530-898414e9736d Duration: 2295.41 ms Billed Duration: 2300 ms
```

**Figure 4.1 The output for the intent ‘WEB\_ACCESS\_CONTROL’**

### **AUTHENTICATION**

```
We have a few recommendations for your policy, you can perform the solution as follow:
1.Password-based Authentication
2.Proxy Authentication
3.Identity Access Management (IAM)
4.Third Party Authentication
5.Firewall Blocking
END RequestId: 05f98071-fd94-4ae6-bc55-4e84973bba25
REPORT RequestId: 05f98071-fd94-4ae6-bc55-4e84973bba25 Duration: 2194.18 ms Billed Duration: 2200 ms
MB Init Duration: 265.94 ms
```

**Figure 4.2 The output for the intent ‘AUTHENTICATION’**

### **NETWORK\_SECURITY**

```
START RequestId: 7c94f7ab-66d7-405c-bd6b-4f7c16893197 Version: $LATEST
We have a few recommendations for your policy, you can perform the solution as follow:
1.Antivirus Protection
2.Firewall Blocking
3.Intrusion Prevention System
4.Active Monitoring
END RequestId: 7c94f7ab-66d7-405c-bd6b-4f7c16893197
REPORT RequestId: 7c94f7ab-66d7-405c-bd6b-4f7c16893197 Duration: 2239.65 ms Billed Duration: 2300 ms
Init Duration: 276.15 ms
```

**Figure 4.3 The output for the intent ‘NETWORK\_SECURITY’**

## SELF\_HEALING

```
START RequestId: 62185c79-6cfd-464c-a751-0e522fac1d0e Version: $LATEST
We have a few recommendations for your policy, you can perform the solution as follow:
1.Adopt High Availability Mechanism
2.Device Failover
3.Avoid Single Point of Failure
4.Perform Regular Network Maintenance
5.Load Balancing
END RequestId: 62185c79-6cfd-464c-a751-0e522fac1d0e
REPORT RequestId: 62185c79-6cfd-464c-a751-0e522fac1d0e Duration: 2268.10 ms Billed Duration: 2300 ms
MB Init Duration: 279.84 ms
```

**Figure 4.4 The output for the intent 'SELF\_HEALING'**

## PERFROMANCE

```
START RequestId: 375f8742-ec62-440f-bb6c-74bc718e2465 Version: $LATEST
We have a few recommendations for your policy, you can perform the solution as follow:
1.Proxy Web Control
2.Bandwidth Management
3.Active Monitoring
END RequestId: 375f8742-ec62-440f-bb6c-74bc718e2465
REPORT RequestId: 375f8742-ec62-440f-bb6c-74bc718e2465 Duration: 2259.09 ms Billed Duration: 2300 ms
Init Duration: 287.68 ms
```

**Figure 4.5 The output for the intent 'PERFORMANCE'**

## **CHAPTER 5 RESULT AND DISCUSSION**

### **5.1 Calculations of precision metric**

A comparison is made between the output of the intent recommendation system and a group of respondents who have the knowledge and background in the networking field. To get the response and result from this group of people, we have carried out a qualitative survey across 30 respondents. The criteria that we set to locate our potential respondents are at least 3 years of experience in Networking field which reflected they have sufficient knowledge in this field. Among the 30 respondents that we have selected, they came from different organizations or institutions such as UTAR, Sangfor, Jabil, and Huawei. They have been taking roles in different positions with their organization which related to networking such as network engineer, network administrator, lecturer, and project manager.

In this survey, it consists of 10 different case studies, respondents are given multiple choice for each of the use cases and the choices are similar to the recommendations system. The result will be obtained based on empirical evaluation and record in qualitative form. In the meantime, we use the same case study to feed in the recommendation model and determine the output. In the end, we will compare both results from respondents and the model and thus evaluate the model.

To evaluate our recommendation system, we used the standard *precision* metric as the baseline. Precision, in other words, also called positive predictive value which is the fraction of the relevant instances among the retrieved instances. (Bondarenko 2019) By having a precision score, we able to estimate out of the number of the recommendations that the system provided, how many of these recommendations will be matched with respondents' preferences. We will take the priority from respondents and compare the result with our recommended system. In our computation, it will work as the formula below:

$$\textbf{Precision} = \frac{\text{\textit{\# of recommend items relevant}}}{\text{\textit{\# of recommended items}}}$$

We compare the ranking of our model with the ranking from the respondents. We make an interpretation as follow, if the ranking of the recommendations of our intent recognizer model is matched with the ranking of the respondents, our model will be 100% precise. In turn,

if the ranking of our recommendation model is different from the respondents, it will be less accurate. Respondents' ranking will be based on the percentage of their responses to a question. If out of 30 respondents, 10 respondents choose A, 8 respondents choose B, 7 and 5 respondents choose C and D respectively, the ranking will be A, B, C and D. The illustrations of the calculations shown as following:

Intent Recognition Model Ranking	Respondent Ranking
A	A
C	B
B	C
D	D

$$\begin{aligned}
 \text{Precision} &= 2 \text{ out of } 4 \text{ model recommendations match with respondents ranking} \\
 &= \frac{2}{4} \\
 &= 0.5 \text{ (the precision equivalent to 50\%)}
 \end{aligned}$$

To prepare for the survey, a google form was created to collect the data from respondents. The survey form consists of 10 case studies which extract from different organization Internet usage policies and different real-life scenario. There will be two questions for each of the intent to avoid the collection of bias data.

The screenshot shows a Google Form titled "Intent Based Networking: Policy to solutions Recommendations". Below the title is a "Form description" field. The first question is: "1. Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that violate of existing applicable law." Below this question are six radio button options: "Proxy Web Control", "Website Blocking", "Access Control List (ACL)", "Active Monitoring", "Firewall", and "Keyword Blocking". The form is displayed on a light purple background with a sidebar on the right containing various icons for form editing.

**Figure 5.1.1 The survey form for respondents.**

### 5.1.1 Comparisons of results

After having a response from the 30 respondents, we make a comparison and analysis:

#### Case Study 1

1. Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that violate existing applicable law.

#### Category: Web Access Control

Ranking	Model prediction	Respondents' answer
1.	Proxy Web Control	Website Blocking
2.	Website Blocking	Proxy Web Control
3.	Access Control List (ACL)	Access Control List (ACL)
4.	Active Monitoring	Active Monitoring
5.	Firewall	Firewall
6.	Keyword Blocking	Keyword Blocking

Calculation = 4 out of 6 predictions match with respondents ranking

$$= \frac{4}{6}$$

=0.67 (equivalent to 66.67%)

#### Case Study 2

2. Users are not allowed for seeking to gain access via the internet to restricted areas of the company's computer system or another organization's or person's computer systems without authorization.

#### Category: Authentication

Ranking	Model prediction	Respondents' answer
1.	Password-Based Authentication	Identity Access Management (IAM)
2.	Proxy Authentication	Proxy Authentication
3.	Identity Access Management (IAM)	Password-Based Authentication
4.	Third-party Authentication	Third-Party Authentication
5.	Firewall Blocking	Firewall Blocking

Calculation = 3 out of 5 predictions match with respondents ranking

$$= \frac{3}{5}$$

=0.6 (equivalent to 60%)



**Case Study 3**

3. Users are prohibited to use the Internet for non-work-related bandwidth use for intensive activities such as network games and the downloading of music or video files or serving as a host for such activities.

**Category: Performance**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Proxy Web Control	Proxy Web Control
2.	Bandwidth Management	Bandwidth Management
3.	Active Monitoring	Active Monitoring

Calculation = 3 out of 3 predictions match with respondents ranking

$$= \frac{3}{3}$$

= **1 (equivalent to 100%)**

**Case Study 4**

4. Any interruption of network services can affect an organization's ability to access, collect or use data and communicate with staff, partners and customers. Interruptions put business continuity (BC) and data at risk and can result in huge customer service and public relations problems. A contingency plan for dealing with any sort of network interruption is vital to an organization's survival.

**Category: Self-Healing**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Adopt High Availability Mechanism	Adopt High Availability Mechanism
2.	Device Failover	Perform Regular Network Maintenance
3.	Avoid Single Point of Failure	Avoid Single Point of Failure
4.	Perform Regular Network Maintenance	Device Failover
5.	Load Balancing	Load Balancing

Calculation = 3 out of 5 predictions match with respondents ranking

$$= \frac{3}{5}$$

= **0.60 (equivalent to 60%)**

**Case Study 5**

5. Users are restricted or inhibit to use or enjoy the service for posting or transmitting any information or software which contains a worm, virus, or other harmful feature, or generating levels of traffic sufficient to impede others' ability to use, send, or retrieve information.

**Category: Network Security**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Antivirus Protection	Antivirus Protection
2.	Firewall Blocking	Firewall Blocking
3.	Intrusion Prevention System (IPS)	Intrusion Prevention System (IPS)
4.	Active Monitoring	Active Monitoring

Calculation = 4 out of 4 predictions match with respondents ranking

$$= \frac{4}{4}$$

**= 1 (equivalent to 100%)**

**Case Study 6**

6. A single network consists of multiple routers, nodes, or switches. One of those network components might become a single point of failure and stop working, which can trigger a cascade of failures within a single network.

**Categories: Self-Healing**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Adopt High Availability Mechanism	Adopt High Availability Mechanism
2.	Device Failover	Device Failover
3.	Avoid Single Point of Failure	Load Balancing
4.	Perform Regular Network Maintenance	Perform Regular Network Maintenance
5.	Load Balancing	Avoid Single Point of Failure

Calculation = 1 out of 3 predictions match with respondents ranking

$$= \frac{1}{3}$$

**= 0.6 (equivalent to 60%)**

**Case Study 7**

7. During any use of the computer or internet, you must not visit internet sites or download any files that contain indecent, obscene, pornographic, hateful or other objectionable materials.

**Category: Web\_Access\_Control**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Proxy Web Control	Website Blocking
2.	Website Blocking	Proxy Web Control
3.	Access Control List (ACL)	Access Control List (ACL)
4.	Active Monitoring	Active Monitoring
5.	Firewall	Firewall
6.	Keyword Blocking	Keyword Blocking

Calculation = 4 out of 6 predictions match with respondents ranking

$$= \frac{4}{6}$$

**= 0.67(equivalent to 67%)**

**Case Study 8**

8. Users are not allowed to access to company information without authorization. This includes unauthorized reading of confidential data, unauthorized access of personnel file information, and accessing information that is not needed for the proper execution of job functions. All user login must be authenticated by username and password.

**Category: Authentication**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Password-Based Authentication	Password-Based Authentication
2.	Proxy Authentication	Proxy Authentication
3.	Identity Access Management (IAM)	Identity Access Management (IAM)
4.	Third-party Authentication	Third-Party Authentication
5.	Firewall Blocking	Firewall Blocking

Calculation = 5 out of 5 predictions match with respondents ranking

$$= \frac{5}{5}$$

**= 1 (equivalent to 100%)**

**Case Study 9**

9. Network services can get disrupted after a cyber-attack, whose aim is to prevent the organization from delivering its services, forcing it to shut down.

**Category: Network Security**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Antivirus Protection	Intrusion Prevention System (IPS)
2.	Firewall Blocking	Firewall Blocking
3.	Intrusion Prevention System (IPS)	Antivirus Protection
4.	Active Monitoring	Active Monitoring

Calculation = 2 out of 4 predictions match with respondents ranking

$$= \frac{2}{4}$$

= **0.5 (equivalent to 50%)**

**Case Study 10**

10. Users are monitored for bandwidth usage. The limit for office computers is typically one gigabyte of data per day (upload and download), although exceptions can be made based on specific needs.

**Category: Performance**

<b>Ranking</b>	<b>Model prediction</b>	<b>Respondents' answer</b>
1.	Proxy Web Control	Bandwidth management
2.	Bandwidth Management	Proxy Web Control
3.	Active Monitoring	Active Monitoring

Calculation = 1 out of 3 predictions match with respondents ranking

$$= \frac{1}{3}$$

= **0.33 (equivalent to 33.33%)**

After having all the precision scores, we calculate the average precision score of our intent recognition model.

**Average precision =**

$$= \frac{0.67+0.6+1+0.6+1+0.6+0.67+1+0.5+0.33}{10}$$

$$= \frac{6.97}{10}$$

$$= 0.697 \text{ (equivalent to 69.7\%)}$$

The precision score for the recommendation system has an average score of 69.7%. It can interpret that every use case given to the recommendation system will be 69.7% of the chance match with users' ranking. In other words, if given 10 policy to a user, 69% of the recommendations will match their intentions. However, it has a 30.3% of the possibility that it could not meet the requirement of the user. Hence, 69.7% of the precision score can interpret as the system can only achieve more than half of the user requirements.

## **5.2 Calculation of Personalization Metric**

To have more study on our recommendations system, we have calculated another metric which is *personalization*. Personalization is a metric that used access if a model recommends different things to different users. High personalization score meaning the model delivers a personalized experience to each user. (Claire 2018) To calculate it, we perform 3 testing with the following use case: *"Users should consider their Internet activities as periodically as non-work related activities are not allowed during working hours."* After 3 testing, the results returned with a list of recommendations which include Proxy Web Control, Website Blocking, Access Control List (ACL), Active Monitoring, Firewall, and Keyword Blocking. We use the result illustrated with the equation below:

$$\text{Results} = \{$$

$$[\text{'Proxy Web Control'}, \text{'Website Blocking'}, \text{'Access Control List (ACL)'}, \text{'Active Monitoring'}, \text{'Firewall'}, \text{'Keyword Blocking'}],$$

$$[\text{'Proxy Web Control'}, \text{'Website Blocking'}, \text{'Access Control List (ACL)'}, \text{'Active Monitoring'}, \text{'Firewall'}, \text{'Keyword Blocking'}],$$

$$[\text{'Proxy Web Control'}, \text{'Website Blocking'}, \text{'Access Control List (ACL)'}, \text{'Active Monitoring'}, \text{'Firewall'}, \text{'Keyword Blocking'}]$$

$$\}$$

After that, we turn it into a metric that represented in binary indicator ('1' represent the item recommended to a user while '0' represent the item that was not recommended), as 3 times of the testing give the same result, we illustrated the result in a matrix form shown as below:

	<i>Proxy Web Control</i>	<i>Website Blocking</i>	<i>Access Control List (ACL)</i>	<i>Active Monitoring</i>	<i>Firewall</i>	<i>Keyword Blocking</i>
<b>Testing 1</b>	1	1	1	1	1	1
<b>Testing 2</b>	1	1	1	1	1	1
<b>Testing 3</b>	1	1	1	1	1	1

Next, we compute with the matrix by using cosine similarity metric, cosine similarity is used to determine how different the metrics are regardless of its size. (Selva Prabhakaran 2018). To calculate it, we use the following formula and write python code to calculate similarity matrix.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

**Figure 5.2.1** The formula that used to calculate similarity matrix.

The formula was written in the python code below:

```
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine
import numpy as np

m = np.array([
    [ 1, 1, 1, 1, 1, 1],
    [ 1, 1, 1, 1, 1, 1],
    [ 1, 1, 1, 1, 1, 1]
])

d = m.T @ m

norm = (m * m).sum(0, keepdims=True) ** .5

similar=d / norm / norm.T
```

**Figure 5.2.2** The python code that used to calculate similarity matrix.

*m* = The metrics that represent the result of the 3 testing

*d* = The numerator of the equation

*norm* = parameter  $\|\mathbf{A}\|$  in denominator

*similar* = cosine similarity matrix

In the end, we get the following similarity matrix. The matrix was separated by an upper triangle and a lower triangle to make a comparison between them.

```
array([[1., 1., 1., 1., 1., 1.],
       [1., 1., 1., 1., 1., 1.],
       [1., 1., 1., 1., 1., 1.],
       [1., 1., 1., 1., 1., 1.],
       [1., 1., 1., 1., 1., 1.],
       [1., 1., 1., 1., 1., 1.]])
```

**Figure 5.2.3 The result of the similarity metric.**

From the comparison shown above, there is no difference between the value in the upper triangle and the lower triangle (label with blue triangle in figure). We can conclude that the recommendation system will return the same result every time the user inputs the same use case. Hence, there was no personalization for a different user that use the recommendation system. The recommendation system will return the same result based on what keywords it has detected in the use case and perform direct mapping to get the recommendations for the user.

### **5.3 Calculations of Coverage Metric**

Coverage is the metric used to calculate the percentage of the recommendations of the system that could give based on what it has inside. (Claire 2018) To calculate it, we use the following formula:

$$\text{Coverage} = \frac{n}{N} \times 100$$

$$N = \text{Total recommendations in the system}$$

$$n = \text{Recommendations used}$$

To compute, we first define the variable N and n. There are 23 recommendations in total for all the intents. Hence, this is the value for variable N. For variable n, we define it with the value of 23 as well. This is because every time after the recommendation system computes for intent, it retrieved all the recommendations consist in the list and there are no leftover recommendations. After having both variables N and n, we can make the following calculation:

$$\begin{aligned} \text{Coverage} &= \frac{23}{23} \times 100 \\ &= 100\% \end{aligned}$$

From the result, we get 100% of the coverage in the recommendations system, this percentage can interpret that the system able to utilize all the recommendations in the system which is a good result for evaluating a recommendation system.

#### **5.4 Finding and Implications**

In a nutshell, we can conclude that the recommendation system has a precision score of 69.7%, no personalization for user and 100% coverage of the system. 69.7% of precision score interprets that the system able to meet 69.7% of user preference in networking countermeasure which is a moderate performance as it could only meet more than half of user requirement. Besides, the system does not customize for personalization as there are fixed predefined recommendations for every intent. Last but not least, 100% of the coverage represents that the recommendation system able to utilize all the recommendations in the system and propose to the users.

The recommendation system could get a good result which aligned with user preferences when it able to detect the specific keywords in an input policy. However, if it could not locate the keywords in the input policy or there are multiple intents in a policy, the system might interpret the intent wrongly and end up suggest with an inappropriate solution. The intent recognizer may not detect the keywords if it does not consist in the datasets.

In a nutshell, further improvement in the model is needed as the precision of the recommendation system still not good enough to handle general-purpose policies. Personalization for different users of the system will be considered as well in the future as it was able to cater to different needs for different users



## **CHAPTER 6 CONCLUSION**

### **6.1 What we have achieved?**

This report proposes an intent recognition model that can turn policy into a list of recommendations. The model was built on top of API in Amazon Web Services (AWS). User can process their input and output within a single platform. The user will first input a policy or use case to the recognition model, after that the intent recognition model will have its computation and the output of the system is a list of recommendations that arranged based on weightage.

With the use of Amazon Comprehend, we able to utilize Natural Language Processing techniques to analyze a text and categorize different keywords into a specific label which able to determine the intents of an input policy. The input and output for training a classifier model in Amazon Comprehend rely on Amazon Simple Storage Service (S3) to store the object. After having the output of the model, we write a function in Amazon Lambda for getting the output from S3 storage and determine the intents of a policy. The Lambda functions will then extract out the recommendations from Dynamo DB services. In our system architecture, Dynamo DB storage used to map the intent with its recommendations. At the end of the result, the user will get a list of recommendations that aligned with the system ranking.

Throughout the project development, we making use of semantic analysis to extract the entities from an unstructured text and categorized them into different labels. By using term frequency, we able to compute and determine the intent of a text. Next, to look for recommendations, we use data mapping technique to create a connection between the source and target tables or attributes. The collected data will be transformed into a format that used for the operational and analytical process.

The recommendation model might tend to miss out some keywords if the keywords is not included in the training dataset. Hence, the recommendation model might interpret wrongly for the intents and giving out solutions that is not suitable. Nevertheless, the recommendations given by the model might not suit to every user's preference and sometimes it might not be suitable for some use cases as there is no user personalization for the recommendation model.

## 6.2 Future Direction

The work could be further enhanced by building a larger dataset for the recognition model able to interpret more of the keywords and labels in a text. Hence, the accuracy and chance of the recognition model to identify intent from a text will be much higher. Besides, the recommendations list could be further enhanced by knowing the real-time scenario and give the best and most suitable recommendations to the user. To increase the precision and personalization of the recommendation system, dynamic matching of keywords and labels may be used.

We might consider one use case probably will have two intents, hence a more dynamic design of the recommendation is needed. Besides, to make sure that the recommendations given will not overlap with each other's, decision making or any other technology could be included in the recommendation model as well.

## BIBLIOGRAPHY

*Acceptable Use Policy-Prohibited Activities* 2019. Available from:

<<https://www.gallaudet.edu/gallaudet-technology-services/technology-policy/prohibited-activities>>.[25 March 2020].

Afkhamizadeh, M., Avakov, A., & Takapoui, R. 2013. *Automated Recommendation Systems Collaborative Filtering Through Reinforcement Learning*. Stanford University

*A Holistic Approach of Advanced Threat Mitigation*, 2019. Available from:

<<https://www.sangfor.com/source/news-online-webinars/1304.html>> [5 March 2020].

*Amazon Comprehend*, n.d. Available from: <<https://aws.amazon.com/comprehend/>>

*Amazon DynamoDB*, n.d. Available from: <<https://aws.amazon.com/dynamodb/>>

*Amazon S3*, n.d. aws. Available from: <<https://aws.amazon.com/s3/>>

Amar, C, Amrita, A, Atharva, K, Dewang, G, Lakshay, K, Levi, P, Rahil, G & Sapna, G  
2019. *VIVONET: VISUALLY-REPRESENTED, INTENTBASED, VOICE-ASSISTED NETWORKING*. University of Colorado Boulder, USA. [25 March 2020].

Ambrosone, M 2016, *How to Eliminate One of IT's Biggest Threats: Single Point of Failure (SPOF)*. Available from: <<https://www.vircom.com/blog/how-to-eliminate-one-of-its-biggest-threats-single-point-of-failure-spo/>>. [10 March 2020].

*Authorization*, n.d. Available from:

<<https://www.webopedia.com/TERM/A/authorization.html>>. [3 March 2020].

*Authentication, Authorization, and Accounting Configuration Guide, Cisco IOS Release 15SY*, 2019. Available from:<[https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/sec\\_usr\\_aaa/configuration/15-sy/sec\\_usr\\_aaa-15-sy-book/sec-cfg-authentifcn.html](https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/sec_usr_aaa/configuration/15-sy/sec_usr_aaa-15-sy-book/sec-cfg-authentifcn.html)>. [1 March 2020].

Agrawal, A, J & Kakde, O, J 2013. *Semantic Analysis of Natural Language Queries Using Domain Ontology for Information Access from Database*. I.J. Intelligent Systems and Applications, 12.

*Avoiding a Trojan Virus:Keeping the gates Closed*, n.d. Available from :

<<https://www.kaspersky.com/resource-center/preemptive-safety/avoiding-a-trojan-virus>>. [3 March 2020].

- AWS Lambda, n.d. Available from: <<https://aws.amazon.com/lambda/>>. [29 march 2020].
- Barlett, M 2018, *Block or Unblock Programs in Windows Defender Firewall*.  
Available from: <<https://www.technipages.com/blockunblock-programs-in-windows-firewall>>[10 March 2020]
- Bandwidth Monitor 3.4 build 757 released!*, n.d. Available from :  
<<http://www.bwmonitor.com/>>. [5 March 2020].
- Beyder, A, 2019, *Blocked out: the dangers of overzealous keyword blocking*. Available from:  
<<https://digitalcontentnext.org/blog/2019/06/26/blocked-out-the-dangers-of-overzealous-keyword-blocking/>> [10 March 2020].
- Bondarenko, K 2019. *Precision and recall in recommender systems. And some metrics stuff*.  
Available from: <<https://medium.com/@bond.kirill.alexandrovich/precision-and-recall-in-recommender-systems-and-some-metrics-stuff-ca2ad385c5f8>>.[10 April 2020].
- Boyd N 2018. *What is Intent-Based Networking?* Available from:  
<<https://www.sdxcentral.com/networking/sdn/intent-based/definitions/what-is-intent-based-networking/>>. [21 March 2020].
- Butler, B. 2017. *Cisco has jumped into intent-based networking market*.  
Available from: <<https://www.networkworld.com/article/3202699/what-is-intent-based-networking.html>>. [1 March 2019].
- Celcer, Tine, Svigelj, Ales & Mohorcic, Mihael. 2008. *Network Architectures Exploiting Multiple HAP Constellations for Load Balancing*. 7.
- Charalambides, M 2005, *Policy conflict analysis for quality of service management. Sixth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'05)*. pp. 99-108.
- Chapter 4 : Capturing Live Network Data*, n.d. Available from:  
<[https://www.wireshark.org/docs/wsug\\_html\\_chunked/ChapterCapture.html](https://www.wireshark.org/docs/wsug_html_chunked/ChapterCapture.html)>. [5 March 2020]
- Claire, L 2018. *Evaluation Metrics for Recommender Systems*.  
Available from: <<https://towardsdatascience.com/evaluation-metrics-for-recommender-systems-df56c6611093>>. [26 March 2020]

- Creating a bash completion script* 2018. Available from:  
<<https://iridakos.com/programming/2018/03/01/bash-programmable-completion-tutorial>>. [13 April 2020].
- Credentials Processes in Windows Authentication*, 2016.  
Available from: <<https://docs.microsoft.com/en-us/windows-server/security/windows-authentication/credentials-processes-in-windows-authentication>>. [23 March 2020].
- Comodo Cybersecurity, n.d. *Define Antivirus*. Available  
from :<<https://antivirus.comodo.com/security/define-antivirus.html>>. [10 March 2020].
- Danelle, A 2013, *Monitoring Your Unknown Network Traffic*. Available at:  
<<https://www.securityweek.com/monitoring-your-unknown-network-traffic>> [5 March 2020].
- David.G, Brian.T, Marjorie.S, Anshu.A & Sailu.R, n.d. *Proxy Authentication*.  
Available from: <<https://www.oreilly.com/library/view/http-the-definitive/1565925092/ch06s07.html>>. [24 March 2020].
- Davide, S, Daniele, M, Mattia, G, Ilario, F, Antonio, C 2018. *ONOS Intent Monitor and Reroute service: enabling plug&play routing logic*. IEEE International Conference on Network Softwarization (NetSoft 2018) - Technical Sessions. [26 March 2020].
- Deep, M & Karthik, R 2018, *Multipath Routing*. Network Routing(Second edition), pp396-422. [27 March 2020].
- Device failover*, n.d. Available from: <[https://help.fortinet.com/fos50hlp/54/Content/FortiOS/fortigate-high-availability-52/HA\\_failoverDevice.htm](https://help.fortinet.com/fos50hlp/54/Content/FortiOS/fortigate-high-availability-52/HA_failoverDevice.htm)>. [10 March 2020].
- Desot, T, Portet, F and Vacher, M 2019. *Towards End-to-End spoken intent recognition in smart home*. 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). Timisoara, Romania, pp. 1-8.
- Dmitriy, S, Wang, Y, Christian, F, Kumar, A, Liu, B, Bengio, Y 2018. *Towards end-to-end spoken language understanding*. Cornell University.
- Employee Computer Use Policies*, n.d. Available  
from:<<http://www.canseeyou.com/computer-use-policies/>>. [3 March 2020].

Expert System Team 2017. *Semantic analysis: When you really want to understand meaning in text*. Available from: <<https://expertsystem.com/natural-language-process-semantic-analysis-definition/>>. [26 March 2020].

Failover 2017. Available from : <<https://www.techopedia.com/definition/1202/failover>>. [8 March 2020].

Foltz, P 1996. *Latent Semantic Analysis for Text-Based Research*. Behavior Research Methods. 28. 197-202. 10.3758/BF03204765. [26 March 2020].

Fruhlinger, J 2020, *What is a cyber-attack? recent examples show disturbing trends*. Available from: <<https://www.csoonline.com/article/3237324/what-is-a-cyber-attack-recent-examples-show-disturbing-trends.html>>. [4 March 2020].

George, N 2019, *Upload vs. download speed: what's the difference?*. Available from: <<https://www.allconnect.com/blog/difference-between-download-upload-internet-speeds>>.[5 March 2020].

Gold, K 2019, *The role of active and passive monitoring in virtual networks*. EXFO. Available from: <<https://www.exfo.com/en/resources/blog/active-passive-network-monitoring/>>.[10 March 2020].

Hanumanthappa. R. S. M 2016. *Techniques of Semantic Analysis for Natural Language Processing – A Detailed Survey*. M S Ramiah Institute of Technology, Bangalore.

Hoffman, C 2014. *Use Tab Completion to Type Commands Faster on Any Operating System*. Available from: <<https://www.howtogeek.com/195207/use-tab-completion-to-type-commands-faster-on-any-operating-system/>>.[13 April 2020].

Hein, D 2019, *Active Monitoring and Passive Monitoring: What's the Difference?* Available from: <<https://solutionsreview.com/network-monitoring/active-monitoring-and-passive-monitoring-whats-the-difference/>>.[10 March 2020].

High Availability (HA), 2017. Available from: <<https://www.techopedia.com/definition/1021/high-availability-ha>> [10 March 2020].

Hoffman,C 2019, *How to the Monitor the Bandwidth and Data Usage of Individual Devices on Your Network*. Available from: <<https://www.howtogeek.com/222740/how-to-the-monitor-the-bandwidth-and-data-usage-of-individual-devices-on-your-network/>> [6 March 2020].

*How Does DNS Filtering Work?* 2019. Available from:

<<https://www.spamtitan.com/web-filtering/how-does-dns-filtering-work/>> [24 March 2020].

*How To Prevent unauthorized Computer Access* 2019. Available from :

<<https://www.computerhope.com/issues/ch000464.htm>>. [4 March 2020]

*How To Prevent unauthorized Computer Access* 2018. Available at :

<<https://www.completecontroller.com/how-to-prevent-unauthorized-computer-access/>>. [4 March 2020].

Hughes, J 2004, *Proxy appliances control Web access..* Available

from : <<https://www.networkworld.com/article/2332826/proxy-appliances-control-web-access.html>>. [10 March 2020]

Hussain, I 2005, *Understanding High Availability of IP and MPLS Networks.*

Available from: <<https://www.ciscopress.com/articles/article.asp?p=361409&seqNum=4>> [5 March 2020].

*Identity Authentication – Are They Who They Say They Are?* 2018. Available from:

<<https://www.trulioo.com/blog/identity-authentication/>>. [23 March 2020].

*Intent-based Networking explained* 2018. Available from:

<<https://www.noction.com/blog/intent-based-networking-ibn-explained>>. [8 April 2020].

*Internet And Email Access Policy* 2015. Available from:

<[https://www.hwca.com/app/uploads/2015/02/Internet\\_and\\_Email\\_Access\\_Policy.pdf](https://www.hwca.com/app/uploads/2015/02/Internet_and_Email_Access_Policy.pdf)>. [3 March 2020].

*Internet Access Control*, n.d. Available from: <<https://www.webtitan.com/internet-access-control/>>. [3 March 2020].

*Introduction to Business Continuity*, n.d. Available from :

<<https://www.thebci.org/knowledge/introduction-to-business-continuity.html>>. [5 March 2020]

Jackson, G, S, n.d. *Proxy Authentication Types*. Available from:

<<https://itstillworks.com/proxy-authentication-types-3269.html>>. [28 March 2020].

- Jose, N & Somani, A, K 2003. *Connection rerouting/network reconfiguration*. Fourth International Workshop on Design of Reliable Communication Networks. Proceedings., Banff, Alberta, Canada, 2003, pp. 23-30.
- Kinght, G 2014, *Download Speed vs Upload Speed: What's the Difference?* Available from: <<https://www.bandwidthplace.com/download-speed-vs-upload-speed-whats-the-difference-article/>>. [5 March 2020].
- Koehrsen, W 2018. *Building a Recommendation System Using Neural Network Embeddings*. Available from: <<https://towardsdatascience.com/building-a-recommendation-system-using-neural-network-embeddings-1ef92e5c80c9>>. [14 April 2020].
- Liu B & Ian L 2016. *Attention-based recurrent neural network models for joint intent detection and slot filling*. Carnegie Mellon University. pp. 685-689.
- Lupu, E, C & Sloman, M 1999. *Conflicts in policy-based distributed systems management*. in *IEEE Transactions on Software Engineering*, vol. 25, no. 6, pp. 852-869, Nov.-Dec.
- Malicious Network Traffic*, n.d. Available from: <[https://www.limestonenetworks.com/support/knowledge-center/1/78/malicious\\_network\\_traffic.html](https://www.limestonenetworks.com/support/knowledge-center/1/78/malicious_network_traffic.html)> .[4 March 2020].
- Malicious Traffic: Understanding What Does and Doesn't Belong on Your Network*, n.d. Available from: <<https://logrhythm.com/webcasts/malicious-traffic-understanding-what-does-and-doesnt-belong-on-your-network/>>.[26 March 2020].
- Martin.J.A, Waters.J.K 2018. *What is IAM? Identity and access management explained*. Available from: <<https://www.csoonline.com/article/2120384/what-is-iam-identity-and-access-management-explained.html>> .[27 March 2020].
- Malware*, n.d, Available from : <<https://www.malwarebytes.com/malware/>> . [24 March 2020].
- Meet Django* 2018. The Web framework for perfectionists with deadlines — Django. Available from: <<https://www.djangoproject.com/>>. [23 March 2020].
- Natural Language Processing (NLP)*, n.d. Available from: <<https://www.scm.tees.ac.uk/isg/aia/nlp/NLP-overview.pdf>>.[25 March 2020].
- Network Maintaince*, n.d. Available from: <<https://networklessons.com/cisco/ccie-routing-switching-written/network-maintenance>>. [10 March 2020].



O'Boyle, B 2019. *What is Siri and how does Siri work?* Available from: <<https://www.pocket-lint.com/apps/news/apple/112346-what-is-siri-apple-s-personal-voice-assistant-explained>>.[20 March 2020].

Oracle 2017. *Oracle® Secure Global Desktop Administration Guide for Release 5.3. 2.6 Third-Party Authentication*. Available from: <[https://docs.oracle.com/cd/E65412\\_01/E65658/html/third-party-auth.html](https://docs.oracle.com/cd/E65412_01/E65658/html/third-party-auth.html)> .[26 March 2020].

Oracle, n.d.. *Oracle® Fusion Middleware Part 12*. Available from: <[https://docs.oracle.com/cd/E65459\\_01/dev.1112/e65461/content/authn\\_ws\\_user.html](https://docs.oracle.com/cd/E65459_01/dev.1112/e65461/content/authn_ws_user.html)>.[26 March 2020].

*Password Authentication Protocol (PAP)*, n.d. techopedia. Available from: <<https://www.techopedia.com/definition/4043/password-authentication-protocol-pap>>.[23 March 2020].

Petters, J 2019, *What is a Proxy Server and How Does it Work?* Available from: <<https://www.varonis.com/blog/what-is-a-proxy-server/>>. [10 March 2020]

Petters, J 2018. *What is a Domain Controller, When is it needed+Set Up*. Available from: <<https://www.varonis.com/blog/domain-controller/>>.[23 March 2020].

Poremba, S, M 2017. *Network Access Control: Restricting and Monitoring Access to Your Network and Data*. Available from: <<https://www.esecurityplanet.com/network-security/network-access-control.html>>. [23 March 2020].

Prescient 2017, *How to Prevent 7 Common Causes of Network Failures*. Available from: <<https://www.prescientsolutions.com/blog/prevent-7-common-causes-network-failures/>>. [23 March 2020]

Redd, M, V & Hanumanthappa, M 2014. *Semantical and Syntactical Analysis of NLP*//, *IJCSIT*, Vol. 5 (3), 3236 –3238. [26 March 2020].

Reed, J 2019. *How to Create an Effective Network Disaster Recovery Plan*. Available from: <<https://www.nakivo.com/blog/create-effective-network-disaster-recovery-plan/>>

- Rossi,B 2015, *Six network security checks to mitigate the risk of data security breaches*. Available from: <<https://www.information-age.com/six-network-security-checks-mitigate-risk-data-security-breaches-123459554/>>. [5 March 2020]
- Rouse, M 2018. *authentication*. Available from: <<https://searchsecurity.techtarget.com/definition/authentication>>.[ 12 March 2020].
- Rouse, M., n.d., *high availability (HA)*. Available from : <<https://searchdatacenter.techtarget.com/definition/high-availability>>. [10 March 2020]
- Rouse, M 2006. *authorization*. Available from: <<https://searchsoftwarequality.techtarget.com/definition/authorization>>. [2 March 2020].
- Rouse, M 2018. *access control*. Available from: <<https://searchsecurity.techtarget.com/definition/access-control>>. [5 March 2020].
- Rouse, M, n.d., *high availability (HA)*. Available from: <<https://searchdatacenter.techtarget.com/definition/high-availability>>.[5 March 2020].
- Sangfor Internet Access Management*, n.d. Available from: <<https://www.sangfor.com/product/sxf-network-security-iam.html>>. [3 March 2020].
- Sample Internet Policy*, n.d. Available from: <<https://www.connectingup.org/learn/articles/sample-internet-policy>>
- Sample internet usage policy*, n.d. Available from: <<https://www.gfi.com/pages/sample-internet-usage-policy>>. [4 March 2020].
- Security Authentication vs. Authorization | What's the Difference?*, 2018. Available from: <<https://swoopnow.com/security-authentication-vs-authorization/>>.[1 March 2020].
- Sella, R 2016. *No more username/passwords: Just use a 3rd party for authentication*. *medium*. Available from: <<https://medium.com/@sellarafaeli/no-more-username-passwords-just-use-a-3rd-party-for-authentication-59b12db092a4>>.[23 March 2020].
- Selva,P. 2018, *Cosine Similarity – Understanding the math and how it works (with python codes)*. Available from : <<https://www.machinelearningplus.com/nlp/cosine-similarity/>>.[30 March 2020].

- Singh, M, K, Rishi, O, P, Awasthi, S, Srivastava, P & Wadwa, S 2020. *Classification and Comparison of Web Recommendation Systems used in Online Business*. International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 2020, pp. 471-480.
- Statt, N 2019. *AI voice assistants reinforce harmful gender stereotypes, new UN report says*. Available from: <<https://www.theverge.com/2019/5/21/18634322/amazon-alexa-apple-siri-female-voice-assistants-harmful-gender-stereotypes-new-study>>. [21 March 2020]
- Stanford Computer and Network Usage Policy 2014. Available from: <<https://adminguide.stanford.edu/chapter-6/subchapter-2/policy-6-2-1>>[3 March 2020].
- Taylor, H 2020, *What Are Cyber Threats And What To Do About Them*. Available from : <<https://preyproject.com/blog/en/what-are-cyber-threats-how-they-affect-you-what-to-do-about-them/>> .[4 March 2020].
- Technical Steering Team (TST), n.d. *Open Networking Operating System (ONOS)*. Available from: <<https://www.opennetworking.org/onos/>>. [28 March 2020].
- THATAGENCY 2020. *WHAT IS KEYWORD MAPPING AND WHY IS IT IMPORTANT?*. Available from : <<https://blog.thatagency.com/what-is-keyword-mapping>> [1 March 2020].
- Thushan Ganegedara 2018. *Intuitive Guide to Latent Dirichlet Allocation*. towardsdatascience. Available from: <<https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158>>.[26 march 2020].
- Tobias, G, M 2019. *What Is Failover?* Available from: <<https://www.datto.com/library/what-is-failover>>. [26 march 2020].
- Tunggal, A, T 2020, *What is a Cyber Attack?* Available from: <<https://www.upguard.com/blog/cyber-attack>>. [4 March 2020].
- Urrutia,C, Ierace, N, & Bassett, R, 2005. *Intrusion Prevention Systems*. Ubiquity
- Volkman & Mathew J 2010, *Filtering & Blocking Software*. Available from: <<https://wiki.uiowa.edu/pages/viewpage.action?pageId=41879146>> [10 March 2020].

- Welcher, P 2018. *Network Maintenance Best Practices*. Available from: <<https://www.networkcomputing.com/networking/network-maintenance-best-practices>>. [10 March 2020].
- What's the Difference in Intentions vs Goals?* 2020. Available from: <<https://contentmentquesting.com/whats-the-difference-in-intentions-vs-goals/>>. [14 April 2020].
- What Are the Most Common Cyber Attacks?* n.d. Available from: <<https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html>>. [28 March 2020].
- What are malware, viruses, Spyware, and cookies, and what differentiates them?* ,n.d, Available from: <<https://community.broadcom.com/symantecenterprise/communities/communityhome/librarydocuments/viewdocument?DocumentKey=f4d70db5-0f14-4f87-9087-98c4978ddaba&CommunityKey=1ecf5f55-9545-44d6-b0f4-4e4a7f5f5e68&tab=librarydocuments>>. [4 March 2020].
- What are Network Firewalls*, n.d. Available from: <<https://www.barracuda.com/glossary/network-firewall>> [10 March 2020]
- What are Proxy Servers*. Available from: <<https://www.paloaltonetworks.com/cyberpedia/what-is-a-proxy-server>> [10 March 2020].
- What is an Intrusion Prevention System (IPS)?*, n.d, Available from: <<https://www.forcepoint.com/cyber-edu/intrusion-prevention-system-ips>> [10 March 2020].
- What Is Alexa?* n.d. Available from: <<https://developer.amazon.com/en-GB/alexa>>. [24 March 2020].
- What is a denial of service attack (DoS)?*, n.d. Available from: <<https://www.paloaltonetworks.com/cyberpedia/what-is-a-denial-of-service-attack-dos>> [3 March 2020].
- What is Bandwidth Management?*, n.d., Available from: <<https://www.manageengine.com/products/netflow/bandwidth-management.html>>[10 March 2020].

*What is Business Continuity?*, 2017. Available from:

<<https://www.inap.com/blog/business-continuity/>>. [8 march 2020]

*What Is Load balancing?* n.d. Available from:

<<https://avinetworks.com/what-is-load-balancing/>>. [5 March 2020].

*What is Trojan Virus?* n.d. Available from :

<<https://www.kaspersky.com/resource-center/threats/trojans>>. [3 March 2020].

*What Is IAM*, n.d. *AWS Identity and Access Management User Guide*. Available from:

<<https://docs.aws.amazon.com/IAM/latest/UserGuide/introduction.html>> . [5 March 2020].

*WS-Security UsernameToken authentication*. n.d.. Available from:

<[https://docs.oracle.com/cd/E65459\\_01/dev.1112/e65461/content/authn\\_ws\\_user.html](https://docs.oracle.com/cd/E65459_01/dev.1112/e65461/content/authn_ws_user.html)>. [1 March 2020].

Y. Han, J. Li, D. Hoang, J. Yoo & J. Hong 2016. *An intent-based network virtualization platform for SDN*, IEEE 12th International Conference on Network and Service Management (CNSM), pp. 353-358, 2016.

Young,C.S, 2016. *Authentication Credentials*. ScienceDirect. Available from:

<<https://www.sciencedirect.com/topics/computer-science/authentication-credential>>.[3 March 2020].

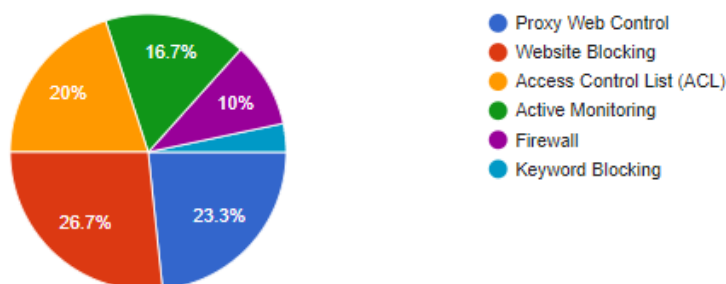
Zhang, J & Mostafa, J 2002. *Information Retrieval by Semantic Analysis and Visualization of the Concept Space of D-Lib Magazine*. University of North Carolina, Chapel Hill and Laboratory for Applied Informatics Research Indiana University. [26 March 2020].

## APPENDICES

### The statistics for the 10 use cases in Section 4.1.1

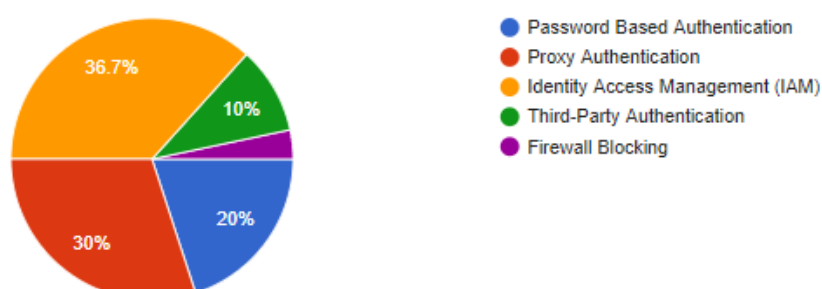
1. Users must not send, view or download fraudulent, harassing, obscene (i.e., pornographic), threatening, or other messages or material that violate of existing applicable law.

30 responses



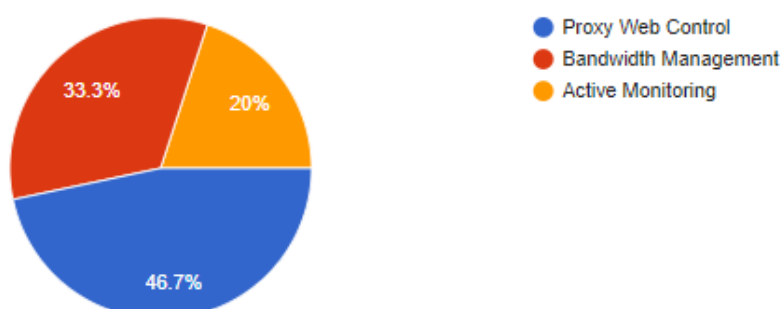
2. Users are not allowed for seeking to gain access via the internet to restricted areas of the company's computer system or another organization's or person's computer systems without authorization.

30 responses



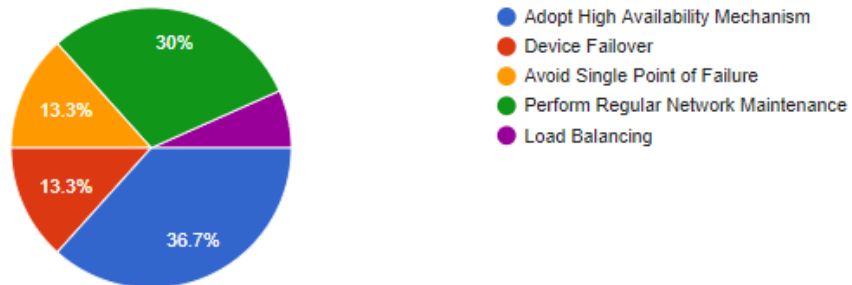
3. Users are prohibited to use Internet for non-work-related bandwidth use for intensive activities such as network games and the downloading of music or video files or serving as a host for such activities.

30 responses



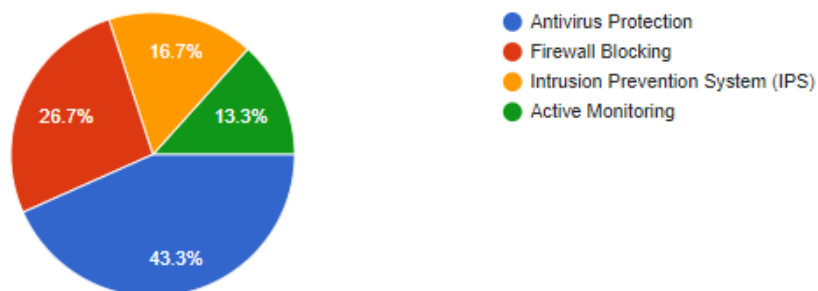
4. Any interruption of network services can affect an organization's ability to access, collect or use data and communicate with staff, partners and customers. Interruptions put business continuity (BC) and data at risk and can result in huge customer service and public relations problems. A contingency plan for dealing with any sort of network interruption is vital to an organization's survival.

30 responses



5. User are restricted or inhibit to use or enjoy the service for posting or transmitting any information or software which contains a worm, virus, or other harmful feature, or generating levels of traffic sufficient to impede others' ability to use, send, or retrieve information.

30 responses



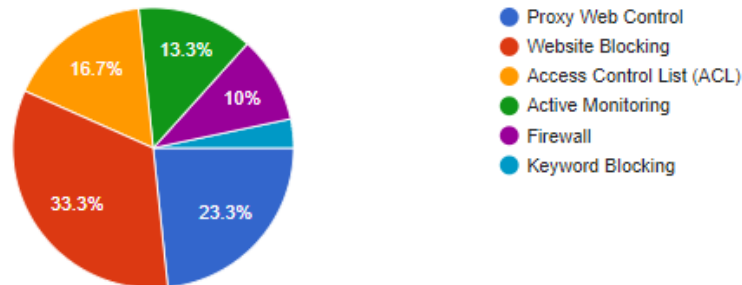
6. A single network consists of multiple routers, nodes, or switches. One of those network components might become single point of failure and stop working, which can trigger a cascade of failures within a single network.

30 responses



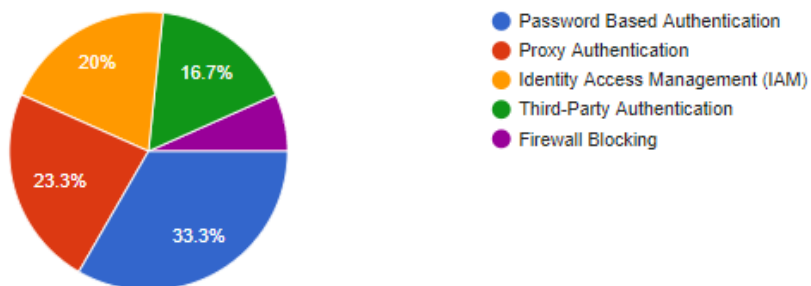
7. During any use of the computer or internet, you must not visit internet sites or download any files that contain indecent, obscene, pornographic, hateful or other objectionable materials.

30 responses



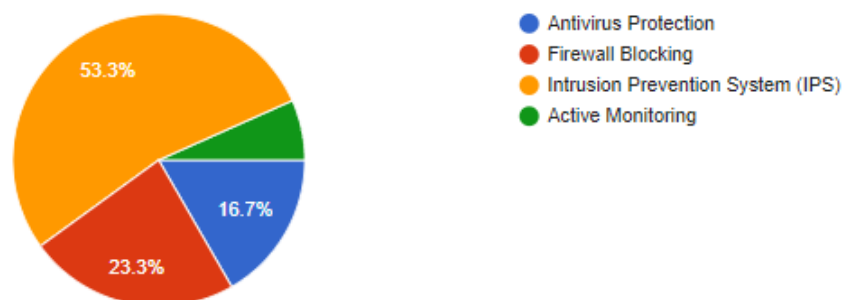
8. User are not allowed access company information without authorization. This includes unauthorized reading of confidential data, unauthorized access of personnel file information, and accessing information that is not needed for the proper execution of job functions. All user login must be authenticated by username and password.

30 responses



9. Network services can get disrupted after a cyber-attack, whose aim is to prevent the organization from delivering its services, forcing it to shut down.

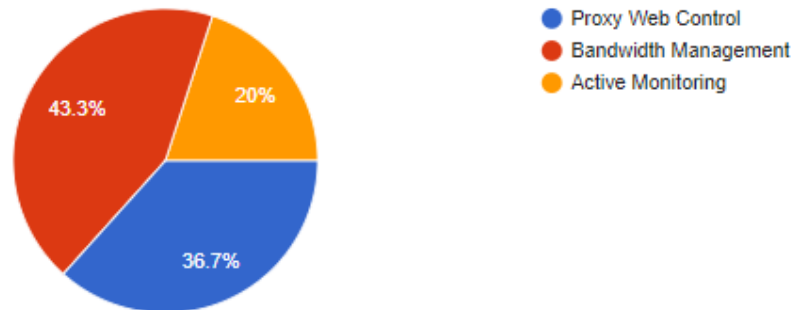
30 responses

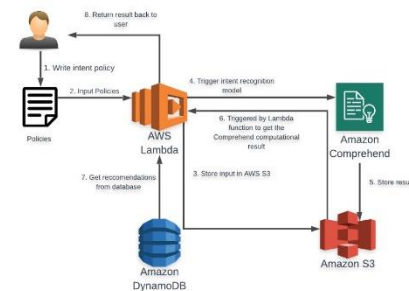




10. Users are monitored for bandwidth usage. The limit for office computers is typically one gigabyte of data per day (upload and download), although exceptions can be made based on specific needs.

30 responses





- 1.Proxy Web Control
- 2.Website Blocking
- 3.Access Control List (ACL)
- 4.Active Monitoring
- 5.Firewall
- 6.Keyword Blocking

The work could be further enhanced by building a larger dataset in order for the recognition model able to interpret more of the keywords and label in a text. Hence, the accuracy for the recognition model to understand a text will be much higher. Besides, the recommendations list could be further enhanced by knowing the real time scenario and give the best and most suitable recommendations to the user. To increase the precision of the recommendation system, dynamic matching of keywords and labels may be used.

## PLAGIARISM CHECK RESULT

feedback studio
Low Jun Sheng
FYP2 Report

Match Overview

14%

14

Submitted to Universiti ...  
Student Paper

2%

2

hal.archives-ouvertes.fr  
Internet Source

2%

3

Davide Sanvito, Daniele...  
Publication

1%

4

airconline.com  
Internet Source

1%

5

ijarce.com  
Internet Source

1%

6

searchdisasterrecovery...  
Internet Source

<1%

7

www.nakivo.com  
Internet Source

<1%

CHAPTER 1 INTRODUCTION

1.1 Problem Statement

Network solution architecting is important to identify a set of tools to address certain networking goals and security requirements. Among a vast of solutions and providers, these options become a selection dilemma for DevOps to choose the right implementation considering the cost of deployment and interoperability issues. The workload for network administrator is getting bigger as the configurations of the network nowadays are getting more complex. With the introduction of Intent Based Networking System, it allows administrator to have an automated network management to achieve what they needed. Network administrator only need to specify their intention on network and IBNS will automate at scale automatically. (Butler 2017) Traditionally, if administrator would like to perform actions on network, they would need to have a background or knowledge in networking to control the network devices through Command Line Interface. Apart from that, the configurations of the whole network

Page: 1 of 79

Word Count: 17006

Text-only Report

High Resolution

## Turnitin Originality Report

Processed on: 16-Apr-2020 00:40 +08  
 ID: 1298044982  
 Word Count: 17006  
 Submitted: 2

FYP2 Report By Low Jun Sheng

Document Viewer

**Similarity Index**

14%

**Similarity by Source**

Internet Sources: 9%

Publications: 5%

Student Papers: 10%

[include quoted](#)
[include bibliography](#)
[excluding matches < 8 words](#)

mode: [quickview \(classic\) report](#)

[print](#)
[download](#)

<p>2% match (student papers from 14-Aug-2019)</p> <p>Class: FYPI2019-May</p> <p>Assignment: FYPI-Report</p> <p>Paper ID: <b>1159985618</b></p>
<p>2% match (Internet from 02-Dec-2019)</p> <p><a href="https://hal.archives-ouvertes.fr/hal-02316743/file/2019_SPED_Desot_final%20%282%29.pdf">https://hal.archives-ouvertes.fr/hal-02316743/file/2019_SPED_Desot_final%20%282%29.pdf</a></p>
<p>1% match (publications)</p> <p><a href="#">Davide Sanvito, Daniele Moro, Mattia Gulli, Ilario Filippini, Antonio Capone, Andrea Campanella, "ONOS Intent Monitor and Reroute service: enabling plug&amp;play routing logic", 2018 4th IEEE Conference on Network Softwareization and Workshops (NetSoft), 2018</a></p>
<p>1% match (Internet from 11-Apr-2019)</p> <p><a href="http://airconline.com">http://airconline.com</a></p>
<p>&lt;1% match (Internet from 27-Jan-2017)</p> <p><a href="http://ijarce.com">http://ijarce.com</a></p>
<p>&lt;1% match (Internet from 22-Mar-2020)</p> <p><a href="https://searchdisasterrecovery.techtarget.com/definition/Network-disaster-recovery-plan">https://searchdisasterrecovery.techtarget.com/definition/Network-disaster-recovery-plan</a></p>
<p>&lt;1% match (Internet from 11-Oct-2019)</p> <p><a href="https://www.nakivo.com/blog/create-effective-network-disaster-recovery-plan/">https://www.nakivo.com/blog/create-effective-network-disaster-recovery-plan/</a></p>
<p>&lt;1% match (publications)</p> <p><a href="#">Mahesh Kumar Singh, Om Prakash Rishi, Shashank Awasthi, Arun Pratap Srivastava, Sumit Wadhwa, "Classification and Comparison of Web Recommendation Systems used in Online Business", 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2020</a></p>

<b>Universiti Tunku Abdul Rahman</b>			
<b>Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)</b>			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1 of 1



**FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY**

<b>Full Name(s) of Candidate(s)</b>	<b>Low Jun Sheng</b>
<b>ID Number(s)</b>	<b>16ACB02227</b>
<b>Programme / Course</b>	<b>Bachelor of Information Technology (Hons) Communications and Networking</b>
<b>Title of Final Year Project</b>	<b>Intent-Based Networking : Policy to Solutions Recommendations</b>

<b>Similarity</b>	<b>Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)</b>
<b>Overall similarity index: <u>14</u> %</b>  <b>Similarity by source</b> Internet Sources: <u>9</u> % Publications: <u>5</u> % Student Papers: <u>10</u> %	
<b>Number of individual sources listed of more than 3% similarity: <u>0</u></b>	
<b>Parameters of originality required and limits approved by UTAR are as Follows:</b> (i) Overall similarity index is 20% and below, and (ii) Matching of individual sources listed must be less than 3% each, and (iii) Matching texts in continuous block must not exceed 8 words <i>Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.</i>	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

***Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.***

Signature of Supervisor

Name: Aun Yichiet

Date: 23/4/2020

Signature of Co-Supervisor

Name: \_\_\_\_\_

Date: \_\_\_\_\_



# UNIVERSITI TUNKU ABDUL RAHMAN



## FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

### CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	16ACB02227
Student Name	Low Jun Sheng
Supervisor Name	Aun Yichiet

TICK (✓)	DOCUMENT ITEMS
	Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
/	Front Cover
/	Signed Report Status Declaration Form
/	Title Page
/	Signed form of the Declaration of Originality
/	Acknowledgement
/	Abstract
/	Table of Contents
/	List of Figures (if applicable)
/	List of Tables (if applicable)
/	List of Symbols (if applicable)
/	List of Abbreviations (if applicable)
/	Chapters / Content
/	Bibliography (or References)
/	All references in bibliography are cited in the thesis, especially in the chapter of literature review
/	Appendices (if applicable)
/	Poster
/	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

\*Include this form (checklist) in the thesis (Bind together as the last page)

<p>I, the author, have checked and confirmed all the items listed in the table are included in my report.</p>  <p>_____ (Signature of Student) Date: 21/4/2020</p>	<p>Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction.</p>  <p>_____ (Signature of Supervisor) Date: 23/4/2020</p>
---	--

# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

Trimester, Year: Y3S3	Study week no.: Week2
Student Name & ID: Low Jun Sheng 16ACB02227	
Supervisor: Dr. Aun Yichiet	
Project Title: Intent-Based Networking : Policy to Solutions Recommendations	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- i) Revised the report in project 1.
- ii) Study for any extension of the work has been done.

## 2. WORK TO BE DONE

- i) Rebuild the recognition model as the model in project1 is not suitable for giving recommendation
- ii) Separate the project into two pipelines, the first pipeline is to look for the intent and the second pipeline is used to give recommendations based on intent.

## 3. PROBLEMS ENCOUNTERED

- i) Collect dataset is time consuming
- ii) Have to complete intent extraction within two weeks

## 4. SELF EVALUATION OF THE PROGRESS

Work could be finish according to timeline by figuring out the method



Supervisor's signature



Student's signature

# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

Trimester, Year: Y3S3	Study week no.: Week 4
Student Name & ID: Low Jun Sheng 16ACB02227	
Supervisor: Dr. Aun Yichiet	
Project Title: Intent-Based Networking : Policy to Solutions Recommendations	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- i) The intent recommendation model is built and is able to extract intent from a text.
- ii) The flow of system is planned.

## 2. WORK TO BE DONE

- i) Implement the flow of the system according to what has planned.
- ii) Studied any other ways to accomplish it.

## 3. PROBLEMS ENCOUNTERED

- i) Encounter some technical issue as some code could not be run

## 4. SELF EVALUATION OF THE PROGRESS

The progress has to be hurry up to make sure everything could meet the timeline.



Supervisor's signature



Student's signature



# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

<b>Trimester, Year: Y3S3</b>	<b>Study week no.: Week 6</b>
<b>Student Name &amp; ID: Low Jun Sheng 16ACB02227</b>	
<b>Supervisor: Dr. Aun Yichiet</b>	
<b>Project Title: Intent-Based Networking : Policy to Solutions Recommendations</b>	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- i) The prototype of whole system is built, waiting to be evaluated.
- ii) The technique used in for the recommendation system is term frequency and data mapping.

## 2. WORK TO BE DONE

- i) Doing evaluation of the recommendation system by conducting empirical evaluation.
- ii) Start writing the FYP 2 report (mainly focus on Chapter 1 & 2).

## 3. PROBLEMS ENCOUNTERED

- i) Data collection for evaluating the system is not easy.

## 4. SELF EVALUATION OF THE PROGRESS

Trying to fit the recommendations system to the real-world scenario, time consuming for thinking out the solutions.



Supervisor's signature



Student's signature

# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

<b>Trimester, Year: Y3S3</b>	<b>Study week no.: Week 8</b>
<b>Student Name &amp; ID: Low Jun Sheng 16ACB02227</b>	
<b>Supervisor: Dr. Aun Yichiet</b>	
<b>Project Title: Intent-Based Networking : Policy to Solutions Recommendations</b>	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- i) Complete the basic part in FYP report for Chapter 1 and Chapter 2
- ii) Still working on the evaluation of the recommendation system.

## 2. WORK TO BE DONE

- i) Enhance the report by adding the criteria requested by supervisor.
- ii) Ensure all the dataset that used to train the recognition model came from authentic source and have to cite for each source.

## 3. PROBLEMS ENCOUNTERED

- i) Time management as there are other assignments need to complete as well.

## 4. SELF EVALUATION OF THE PROGRESS

Trying to be discipline in time management



Supervisor's signature



Student's signature

# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

<b>Trimester, Year: Y3S3</b>	<b>Study week no.: Week 10</b>
<b>Student Name &amp; ID: Low Jun Sheng 16ACB02227</b>	
<b>Supervisor: Dr. Aun Yichiet</b>	
<b>Project Title: Intent-Based Networking : Policy to Solutions Recommendations</b>	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- i) The evaluation of the system is done, working on the report writing

## 2. WORK TO BE DONE

- i) Enhance the report until it able to meet the expectation from the supervisor.

## 3. PROBLEMS ENCOUNTERED

Due to the outbreak of Covid-19, it caused a big trouble in communicating with supervisor.

## 4. SELF EVALUATION OF THE PROGRESS

Try to complete my part regardless of what happened in the surrounding



\_\_\_\_\_  
Supervisor's signature



\_\_\_\_\_  
Student's signature

# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

<b>Trimester, Year: Y3S3</b>	<b>Study week no.: Week 12</b>
<b>Student Name &amp; ID: Low Jun Sheng 16ACB02227</b>	
<b>Supervisor: Dr. Aun Yichiet</b>	
<b>Project Title: Intent-Based Networking : Policy to Solutions Recommendations</b>	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- i) The report was complete, waiting for the checking result from supervisor

## 2. WORK TO BE DONE

- i) Follow the instruction from supervisor to improve the fyp2 report

## 3. PROBLEMS ENCOUNTERED

Due to the outbreak of Covid-19, it caused a big trouble in communicating with supervisor.

## 4. SELF EVALUATION OF THE PROGRESS

Try to complete the full report within the given timeline



Supervisor's signature



Student's signature

# FINAL YEAR PROJECT WEEKLY REPORT

(Project I / Project II)

<b>Trimester, Year: Y3S3</b>	<b>Study week no.: Week 13</b>
<b>Student Name &amp; ID: Low Jun Sheng 16ACB02227</b>	
<b>Supervisor: Dr. Aun Yichiet</b>	
<b>Project Title: Intent-Based Networking : Policy to Solutions Recommendations</b>	

## 1. WORK DONE

[Please write the details of the work done in the last fortnight.]

- 1) The report was modified according to the expectation of the supervisor

## 2. WORK TO BE DONE

- 1) Compile the report and go for Turnitin check.
- 2) Make sure the fyp format comply with the guidelines.

## 3. PROBLEMS ENCOUNTERED

No problem encounters so far

## 4. SELF EVALUATION OF THE PROGRESS

Satisfied with my work could meet the expectation from supervisor



\_\_\_\_\_  
Supervisor's signature



\_\_\_\_\_  
Student's signature