Article

Building an Operational Solution Assistant System for Foreign SMEs in ROK

Hong-Danh Thai [1] and Jun-Ho Huh [1,2,*]

1 Department of Data Informatics, (National) Korea Maritime and Ocean University, Busan 49112, Korea; danhth@g.kmou.ac.kr
2 Department of Data Science, (National) Korea Maritime and Ocean University, Busan 49112, Korea
* Correspondence: 72networks@kmou.ac.kr

Abstract: Foreign Direct Investment (FDI) is an important resource that helps accelerate the development of the country's economy, add substantial funding to growth and facilitate technology transfer. Republic of Korea (ROK) is one of the world's developed countries with dynamic economy, advanced science and technology. In recent years, the Korean government has continuously formulated tax policies, policies to support the business economy and import policies to support foreign businesses in Korea. The Pangyo Valley Creative Economy Valley is being groomed as a global startup hub in Asia. Small and medium enterprises (SMEs) in foreign countries are increasingly interested and eager to seek investment opportunities in the Korean market. Nonetheless, for these companies, language barriers and cultural and institutional differences make it more difficult and time-consuming to learn about the Korean market (such as investment trends, laws, visa policies, taxes and business establishment issues in Korea, etc.). In this study, we explored the process of searching information and seeking investment opportunities and built a business consulting and support application in the first stages of starting a business in ROK to increase effectiveness and save time, which is also an innovative business practice in Use-case ROK. We designed our Virtual Assistant system that can crawl and analyze data on foreign investments in ROK from open data resource websites (data.co.kr) and used analytic and aggregation techniques to explore trends in investments of foreign enterprises. We also researched the process of searching information and seeking investment opportunities for SMEs when investing in ROK, government support policies, laws and taxes as well as a number of other related issues. We built datasets and used Natural Language Processing (NLP) together with Natural Language Understanding (NLU) algorithms to build chatbot applications. Friendly framework for new developers to add and build up the dataset of AI Assistant is built by providing input intent data function, input Entity data function, input utterance data function as well as training and test function. In addition, we built a web-app connected to the server to visualize all the results of research so that SMEs owners can easily use and look for information on investments. Based on the research results, we can make recommendations to SMEs in keeping with the changing investment trends in ROK.

Keywords: foreign direct investment; virtual assistant systems; chatbot; foreign SMEs; operational solution; innovative business practice

1. Introduction

Foreign Direct Investment (FDI) is an important resource that helps accelerate the development of the country's economy, add substantial funding to growth and facilitate technology transfer, aside from strengthening export capabilities and creating more jobs [1]. Even though a growth economy with no competitive advantage of cheap labor, ROK has a great geographic advantage (between China and Japan, the world's second and third

largest economies, respectively); it has also entered the 5G era with new industries emerging in recent years. The FDI in the Republic of Korea (ROK) has been increasing continuously especially in the service sectors where IT/ICT solutions are extensively utilized but it is notable that a number of global companies are also constantly seeking some investment opportunities in the areas where high returns can be expected by establishing their headquarters or R&D centers in the country to participate in the businesses involving cutting-edge technologies or advanced materials. According to a January 6, 2020 report by the Ministry of Industry, Trade and Resources of Korea, the total FDI registration in the year 2019 reached 23.3 billion, the second highest in history [2]. The amount of FDI is actually disbursed to US $12.8 billion, the 4th highest of all time. Compared to the record high 26.9 billion in 2018, last year's registered FDI fell 13.3%, and the capital was disbursed by 26% [2]. ROK has attracted many investments and foreign enterprises in various development industries thanks to policies seeking to support enterprises in a timely manner. Figure 1 shows foreign-invested companies in Korea.

Figure 1. Foreign-invested companies in Korea.

In the past, there have been many companies and organizations seeking investment, cooperation and development opportunities in ROK, a country of potential. According to representatives of Vietnam's foreign Ministry in the Workshop titled "Starting and registering a business in Korea" in 2019, Vietnam-ROK relations are developing well in all fields, in the context of the two countries celebrating the 10th anniversary of establishment of "Vietnam-Korea strategic cooperation partnership" (2009–2019). In terms of investment, ROK is the largest investor in Vietnam with total accumulated investment capital of US$ 62.5 billion as of the end of 2018 (accounting for 18.3%), with 7460 projects creating jobs for 70,000 employees and contributing about 30% of the total export value of Vietnam [3]. With regard to trade, ROK is Vietnam's second largest trading partner (after China) with total two-way trade turnover of US$ 65.7 billion in 2018, aiming to record VND 100 billion by 2020 [3].

As for tourism, there were nearly 3.5 million ROK tourists visiting Vietnam in 2018, whereas the number of Vietnamese tourists to ROK reached nearly 500,000, up 42.1%. People-to-people relations take place widely in all levels and fields. Currently, each country has more than 150,000 citizens studying, living and working in the other country (Vietnam's foreign ministry).

Like many countries, the ROK is striving to keep their economic growth rate to ensure their economic success in the era of the of the 4th industrial revolution that can guarantee higher employment rates and GDP levels. In this effort, the ROK government is encouraging and offering a series of startup programs by providing a substantial support to the promising young entrepreneurs, establishing a global startup hub such as Pangyo Creative Economy Valley, for example. In the context of investment opportunities in Korea, there are many small and medium enterprises (SMEs) from many different countries participating in investment; due to cultural differences, they have encountered certain challenges in the process of seeking opportunities and investments. According to the Future of Business Survey [4], some of the major challenges that may be encountered by

SMEs include increasing revenue, maintaining profitability, attracting customers, securing financing for expansion, developing new products/innovation, finding/working with suppliers, tax laws and rules, other government regulations, etc. as shown in the Figure 2.
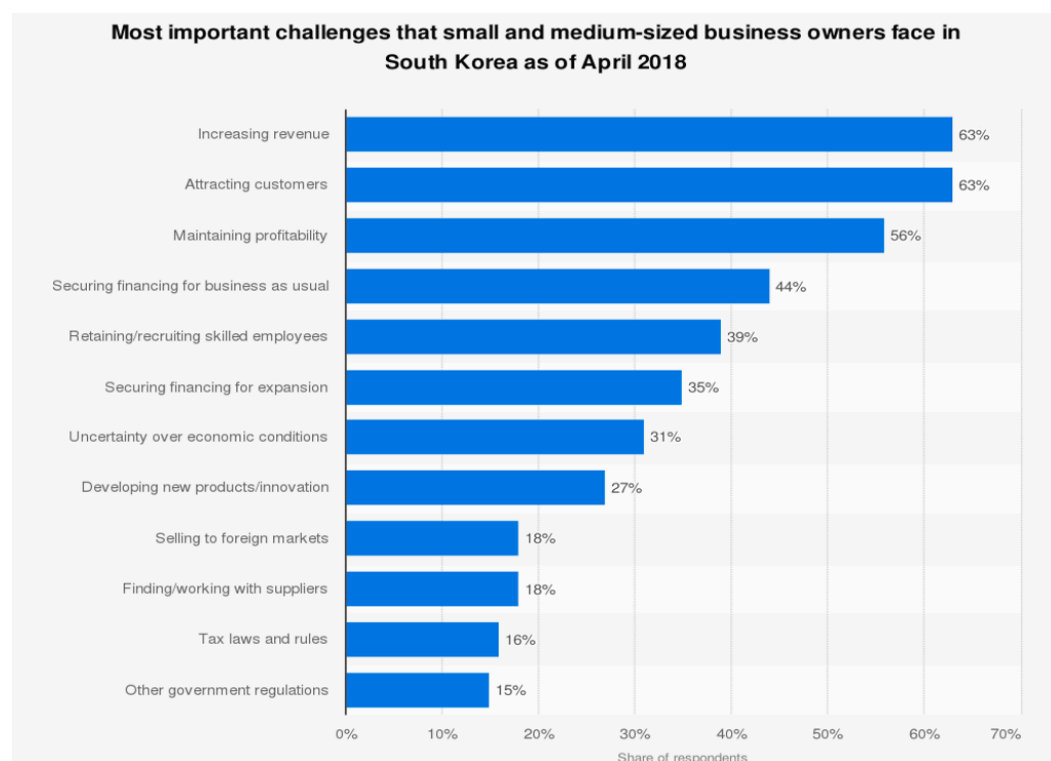

Figure 2. Most important challenges that small and medium business owners face in ROK (South Korea) as of April 2018 *(Source: Future of Business Survey; OECD; World Bank; Facebook; FactWorks)* [4].

Especially in the case of foreign business owners, they are expected to have difficulties in language particularly in seeking information related to investment issues and establishment of the company at the initial stage when entering the Korean market. Neither do they have much money to invest in building a dedicated team of legal institutions in ROK.

In this study, we focused on building an operational solution for small and medium businesses in the form of a Virtual Assistant system that can provide information on finding/working with suppliers, tax laws and rules, other government regulations, investment trends and new investment sectors in Korea, investment regions, etc. We presented methods of collecting and exploiting big data on foreign investment in Korea and methods of building data sets on investment and using Natural Language Processing (NLP) and Natural Language Understanding (NLU) technologies to build chatbot applications for the direct response of assistant systems.

The rest of this paper is organized as follows: Chapter 1 gives an introduction to investment trends of foreign companies in ROK and some difficulties of foreign SMEs particularly Vietnamese SMEs in the foundation-stage information search and in looking for investment opportunities due to language differences; Chapter 2 provides an overview of previous research or papers related to aspects of understanding the original information of foreign enterprises, process of foreign investment in Korea and techniques used to build the assistant and Chatbot; Chapter 3 introduces the methodologies for building Assistant Systems, research design, process of building a chatbot and building Operational Solutions; Chapter 4 presents the results and discussions of the research. Chapter 5, the last chapter, presents the conclusions of this research and future work to improve the results

and contribute to Innovative Business not only in Korea but also in other potential countries.

2. Methods

As for the methodology of this study, data were explored and built according to the process of searching for information and seeking investment opportunities; a business consulting and support application dealing with the first stages of starting a business in ROK was then built to increase effectiveness and reduce time, which is also an innovative business practice in Use-case ROK. We designed our operational solution as a Virtual Assistant system that can crawl and analyze data on foreign investments in ROK from open data resource websites (data.co.kr) and used analytic techniques to explore investments trends of foreign enterprises.

We also researched the process of searching for information and seeking investment opportunities for SMEs when investing in ROK, government support policies, laws and taxes as well as a number of other related issues. Then, we built data sets and Natural Language Processing (NLP) and Natural Language Understanding (NLU) algorithms to build chatbot applications. We built a web-app connected to the server to visualize all the results of research so that SMEs owners can easily use and look for information on investments. Based on the research results, we can make recommendations to SMEs according to the changing investment trends in ROK.

3. Related Research

3.1. Foreign Investors and Business Establishment in Korea

FDI could play a role in the restructuring of the industrial sector through competition in Korea [5]. Over the past decades, the Korean government has made every effort to follow the development roadmap; as technological progress and human resource development have led to an increase in the assets created by all products, the relative significance of both intra-industry and inter-industry trade and attraction of foreign direct investment (FDI) has increased [6]. The continuous increase in foreign direct investments in the Republic of Korea (ROK) are being led by China who is now on the threshold of becoming an industrialized country and some of the cash-rich Middle Eastern countries seeking lucrative investment opportunities taking advantage of the free trade systems such as WTO, FTA and the regional trade agreements [7]. Recently, such FDIs are turning their eyes to the Korean service industry or cultural industry but also seeking their business opportunities in the areas where the latest Korean technologies are being adopted or used to develop a new service system or material [8]. The World Bank regards the Republic of Korea as a country with a highly developed business environment, ranking 5th in Doing Business 2020 [9].

Factors that encourage foreign investors to invest in the Korean market can usually be classified into the following four main groups: highly educated skilled labors force/improving labor climate; excellent social infrastructure; technology and major industries; finance, tax and accounting. There have been a number of organizations established to support the creation of businesses and investments for foreign investors in Korea, such as Korea Foreign Company Association (FORCA), Invest Korea, Korea Trade-Investment Promotion Agency (KOTRA), etc. Invest KOREA's headquarters and KOTRA's 126 overseas offices, 36 of which are overseas FDI offices, are devoted to attracting foreign investment to Korea (*source: investkorea.org*). KOTRA is operating a Foreign Investor Support Center to handle investment licensing/procedures and manage investment notification and other civil affairs, provide investment consulting services (accounting, tax, legal affairs, etc.) and help foreign investors settle down in Korea (one-day secretarial and consulting services). Foreign investors can simply go to the office of KOTRA or avail themselves of Online Consulting by calling. Consulting and support for foreign businesses are

still done in traditional ways and are limited, so businesses can find information and opportunities through seminars, support programs, information in websites or business forums or get directly supported by staff at investment advisory centers. Finding such information entails a lot of time and effort for business owners. Therefore, in this research, we proposed the building of a virtual assistant that provides information on investment and business registration in Korea.

3.2. Natural Language Processing (NLP) and Natural Language Understanding (NLU) Techniques

Language is a complex system for communication between higher-order animals or thought-capable animals like humans. Natural language is not the same as artificial languages such as computer languages (C, PHP). There are currently about 7000 languages in the world. There are many ways to classify languages, with some common language classifications based on origin and characteristics. Formal Language is a set of strings built on an alphabet, bound by predefined rules or grammars. Alphabet can be a set of characters in natural language or a set of self-definition characters. The natural language model follows the rules of the Markov chain, and it was first formalized by Noam Chomsky in the 1950s and was called the "Chomsky Hierarchy Model." Chomsky proposed a series of great simplifications and abstractions for the empirical field of natural language. In particular, this approach completely ignores the meaning, with all problems related to the use of expressions like their frequency, context dependency and processing complexity [10]. These models were later used to create programming languages or applications in automated translation studies.

Since the emergence of computers until now, programmers have tried to write programs that can understand the natural language. The reason is quite clear: humans have a history of writing for thousands of years, and it would be very useful if a computer could read and understand all the data from the massive number of articles written all those years. With the development of technology and engineering, the volume of natural linguistic text worldwide has surged, carrying a large amount of knowledge, but it is increasingly difficult for people to disseminate to discover knowledge/intellect in it, particularly at any given time limit [11]. Natural Language Processing (NLP) and Natural Language Understanding (NLU) aim to do this job efficiently and accurately, just like humans do (for a limited amount of text).

In recent years, conversational UIs (chatbots, bots) have been built, researched and developed by many companies such as Facebook, Google, Amazon and Apple and can be recognized and commanded with text, voice, images, etc. [12]. In addition to the advantages of always being available to communicate and advising users 24/7, conversational UIs also help us save human resources and time and allow us to focus on other activities. In this research, we built a chatbot that builds on a given conversation database (Corpus-based chatbots). This document store can be collected using large amounts of data from user conversations via extraction methods. The system exports information (Information Retrieval) by using machine learning methods to create answers based on the context of the conversation with the user.

3.2.1. Natural Language Processing (NLP)

In the field of human language technology, Natural Language Processing (NLP) adopting a series of computational techniques for the analysis and/or representation of language-based expressions [13] plays an essential part in machine translation, question answering, text summarization, topic modeling, as well as opinion mining, IR and IE. Such an NLP has been used in this research to identify and exclude stop-words, or for the text tokenization (e.g., sentence/word tokenization). Being a basic process in NLP [14], a word string is broken into a series of tokens based on a specific protocol. While sentence tokenization separates and lists sentences in a text corpus, word tokenization breaks up the words in a sentence, based on the spaces, commas, periods and/or carriage return

characters. In these techniques, removal of high-frequency words of negligible semantic value (stop-words removal) is an essential step. Generally, word tokens can be separated by blank spaces, and sentence tokens, by stops. Removing stop words is an important step in NLP text processing. It involves filtering out high-frequency words that add little or no semantic value to a sentence.

As an offshoot of Artificial Intelligence, NLP focuses on studying the interaction between computers and natural human languages in the form of speech or text. NLP can be divided into two large, not completely independent branches: speech processing and text processing. Speech processing focuses on the research and development of algorithms and computer programs that process human language in the form of voices (audio data). Important applications of speech processing include speech recognition and speech synthesizer. Whereas speech recognition is to convert language from speech to text, synthesized speech converts language from text to speech. Text processing focuses on text data analysis. Important applications of text processing include information search and retrieval, machine translation, automatic text summary or automatic spelling check. Text processing is sometimes further divided into two smaller branches including text understanding and text birth [14]. In this research, we focused on collected and processed Text data.

Text processing consists of the following four main steps:

- Diagnostic analysis: The identification, analysis and description of the structure of a hieroglyph in a given language and other linguistic units, such as root word, verb, affix, subclass, etc. In Vietnamese language processing, two typical problems in this section are word segmentation and part-of-speech tagging.
- Parsing: Process of parsing a series of symbols in the form of natural language or computer language according to formal grammar. Formal grammar commonly used in the parsing of natural languages includes Context-free Grammar (CFG), Combinatory categorical grammar (CCG) and Secondary Grammar Dependency grammar-DG. Input to the parse is a sentence consisting of a series of words and their type tag, and output is a parse tree representing the sentence's syntactic structure.
- Semantic analysis: Process of relating semantic structure, from the phrase, clause, sentence and paragraph level to the whole article level, to their independent meanings. In other words, this is to find out the semantics of the verbal input. Semantic analysis has two levels: Semantic lexical expression, which expresses the meanings of the component words and distinguishing the meaning of words; and Component semantics, which refers to the way words combine to form broader meanings.
- Discourse analysis: Text analysis considering the relationship between language and context-of-use. Therefore, discourse analysis is performed at the paragraph level or whole text rather than just analysis at the sentence level alone.

There are many different approaches used to process natural language in NLP, and they fall roughly into four categories: symbolic, statistical, connectionist and hybrid. They have different foundations, typical techniques, differences in processing and system aspects, and robustness, flexibility and suitability for various tasks [15].

3.2.2. Natural Language Understanding (NLU)

Natural Language Understanding (NLU) can be regarded as an essential first step of NLP when dealing with understanding the semantics of a specific text [16]. However, a solution (i.e., AI algorithm, etc.) that can support NLU perfectly is yet to be developed as NLU often attempts to identify the semantics in a text where human errors are made. There are few methods that can be applied to NLU when analyzing and understanding the semantics, but they all require a specific lexicon and a parser along with grammatical rules that can be used when separating the sentences or words. Developing a content-rich lexicon with a well-defined ontology such as WordNet is not easy and often takes years of effort and human resources [17]. Again, playing a basic role for NLP, NLU attempts to

understand natural languages and represent their semantics in a form that can be interpreted by computers when NLP performs a syntactic/semantic analysis [18].

3.3. Personal Assistant Technology

A virtual assistant, which can be called a digital assistant, a voice assistant or an AI assistant, is a task-oriented programming application that recognizes human voices and executes spoken commands by the user. There are three main methods to interact with a virtual assistant through: Text, includes online chat (as messaging app or other app such as Facebook, Viber, Skype), SMS Text, e-mail; Voice, such as Amazon Alexa [19], Google Assistant, Cortana and Siri [20]; or Images Data. Voice assistants typically rely on an Automatic Voice Recognition (ASR) system across 3 levels, starting with capturing sound from the microphone and breaking it down into phonemes for processing into text. Phonology is a basic measure of user speech recognition, which gives better results than word decoding by ignoring contextual limitations. Finally, the system will model the language to find probabilistic information according to the recorded context. Basically, all three popular virtual assistants Google Assistant, Cortana and Siri [20] are different in structure but operate on the basis of neural network technology (such as networks in human brain cells) in the deep backend. Virtual assistant model is provided for free on the technology devices that these big companies release such as iPads, cell phones, laptops, iPods, smart watches and other home electronic devices as a part of added value for users [19]. In this research, we focus on exploring and setting up experiments as text interactive method and based on Chatbot technology.

Chatbot is a popular functional technique of virtual assistant and developed and used in most modern fields such as Marketing, Education, Healthcare, Customer Services [21]. Chatbot is a combination of pre-existing scenarios and self-learning in the interactive process. With the questions posed, chatbot uses Natural Language Processing systems to analyze the data, and then based on Natural Language Understanding systems and machine learning algorithms to generate different types of reactions, they will predict and respond as accurately as possible. Chatbot uses multiple systems to scan keywords within the input, then the bot initiates an action, pulls an answer with the most relevant keywords and responds to information from a database/ API, or handed over to humans. If that situation has not happened yet (not in the database), Chatbot will ignore it, but will also learn by itself to apply to future chats. There are many different concepts to build and execute a chatbot such as Artificial Intelligence Markup Language (AIML) creating natural language agents as well as using pattern recognition and pattern matching techniques available in various programming languages (for example Java, Ruby, Python, C) [22]; Latent Semantic Analysis (LSA), a technique to distribute semantics, analyze relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms [22]; Natural Language Processing (NLP) [23,24]; Natural Language Understanding (NLU) [25,26], Utterance, intent and entity are the three most important terms in building chatbot.

In our study, we will build a virtual assistant as a chatbot which can be much more interactive and personal than the rule-based chatbots. Chatbot can easily interact with users the same way humans converse and communicate in real-world situations. The conversational skills of chatbot technology empower them to deliver what the user is looking for. It can understand the context and intent of complex conversations and attempt to provide more relevant responses.

4. Design and Implementation of Building an Assistant Using NLP and NLU

4.1. Data Collection

At this stage, we focused on understanding the factors that investors may be interested in or the business development opportunities for investors when starting a business

in Korea. We also compiled documents and information of investors interested in searching the information shown in Figure 3. We collected all this data, saved it in MongoDB and prepared for the construction of the chatbot application in the next stage.

Figure 3. Information that new foreign investors in Korea.

We also collected data on foreign investment in Korea by country by year, data on new business registration by locality, and invested industries from the website "data.co.kr." These data were analyzed to find out investment trends in Korea and used for advice and suggestions to users of the system.

4.2. Proposed architecture of the Operational Solution Assistant System

The following are the major features of the Operational Solution Assistant System we built:

- Application Chatbots' direct dialogues.
- Data set about Korean business establishment procedures, laws, tax and labor policies, immigration policies and investment policies.
- Data set with updated information on competitions, government support programs in recent years.
- Data set with updated information on investment trends of newly established enterprises in Korea.
- Gather information on user questions for use in predicting user behavior and trends

This subsection presents the overall architecture of the proposed Operational Solution Assistant System based on Natural Language Understanding (NLU) techniques. The system mainly consists of three parts as shown in Figure 4: (i) Assistant interface; (ii) Python NLP function; and (iii) Python NLU function.
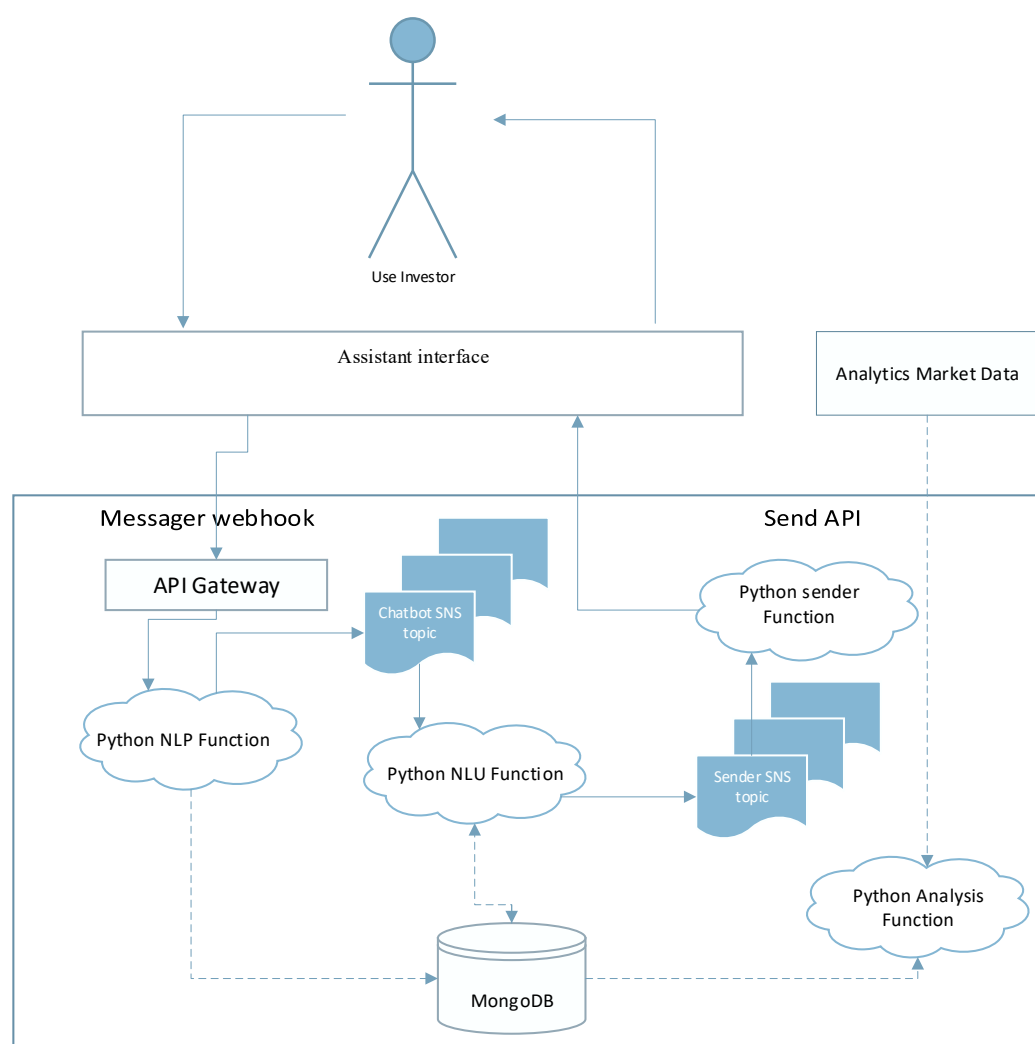
Figure 4. Proposed architecture of the Operational Solution Assistant system.

The user sends a message through the Assistant interface. The message goes to the Message box in the Assistant interface, triggers a callback to a Messenger webhook and gets sent to the API gateway. The API Gateway passes the message to the Python NLP functions. In this NLP function, the message is processed with the NLP code written in Python to remove stop words in the message and extract and organize entities of user's messages and is saved in MongoDB. After that, all the entities of the user's messages pass the Chatbot SNS topic and get sent to the Python NLU function.

The Python NLU function contains all the logic algorithm of the assistant, which will be presented in detail in the next subsection. All of the user's ideas kept in the Python NLU function are saved in MongoDB. After this function is finished, the results go to the sender SNS topic and get sent to the Python Sender function. This function sends the API of the most suitable answer to the Assistant Interface, which then shows the answer to the end user. The system also has another component called Analytics Market Data. This component is used to collect data about the realistic investment market in Korea from reports by quarter, by year or changing policy of FDI in Korea.

All data from this component are sent to the Python Analysis function and saved in MongoDB. When an assistant user asks about the trends of the market, the Python NLU function sends a request to MongoDB and collects this data to answer the assistant user.

The AI Assistant's main workflow is shown in the flowchart in Figure 5. When users send a message to the system, the AI Assistant's main workflow starts with the Define intent in the field of investment by the Python NLP functions. The system defines the

Entities belonging to the field of investment. After that, it builds the pattern of utterance and sends it to the TensorFlow training model. This model builds the chatbot using a model trained to interact with users. If these results have good confidence, this ends the session. If not, the system stores and suggests relative intents.
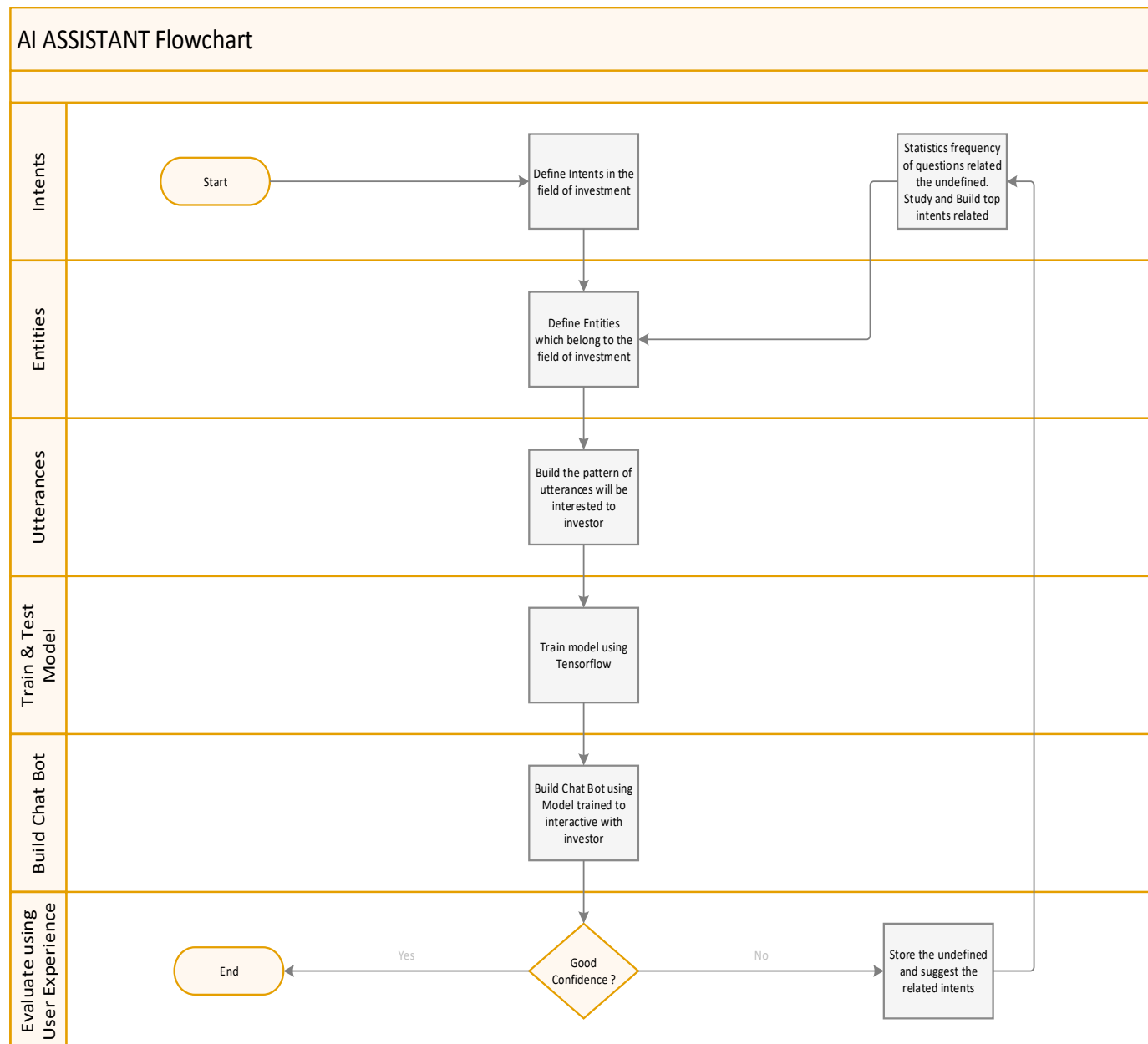
Figure 5. AI Assistant's main workflow.

The flowchart in Figure 6 presents the process of defining Intent, which is a critical factor in chatbot functionality because the chatbot's ability to parse intent is what ultimately determines the success of the interaction.
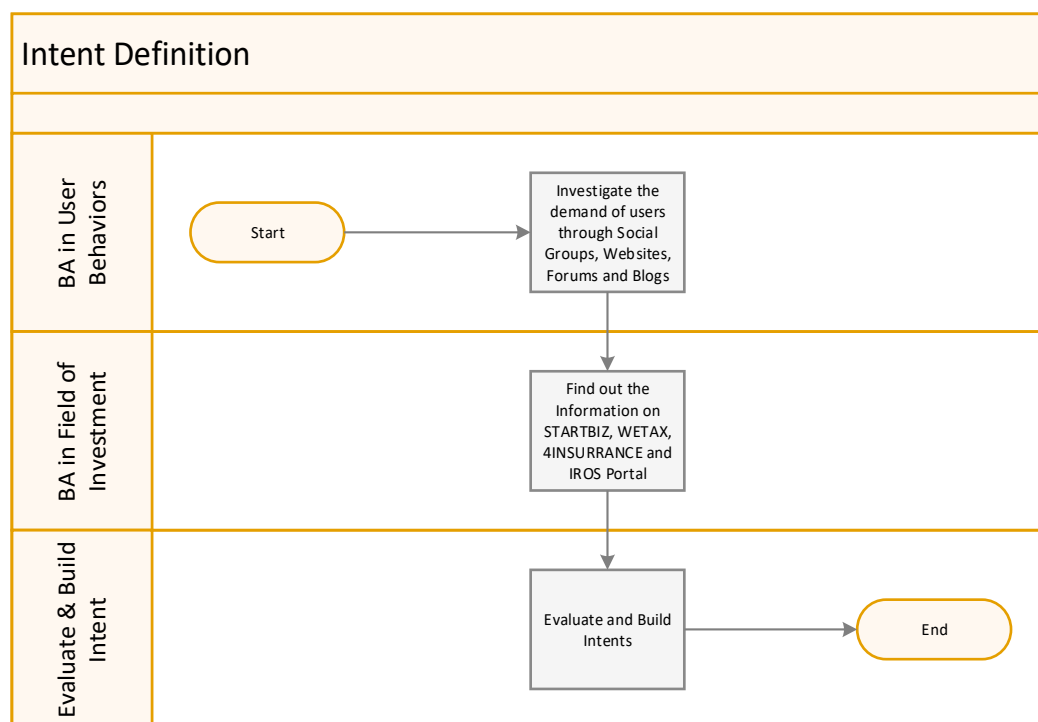
Figure 6. Process of defining Intent.

The flowchart in Figure 7 presents the process of defining Entity. Entities are knowledge repositories used by the bot to provide personalized and accurate responses. Entities can help the system extract important information from the ongoing conversation and catch important data.
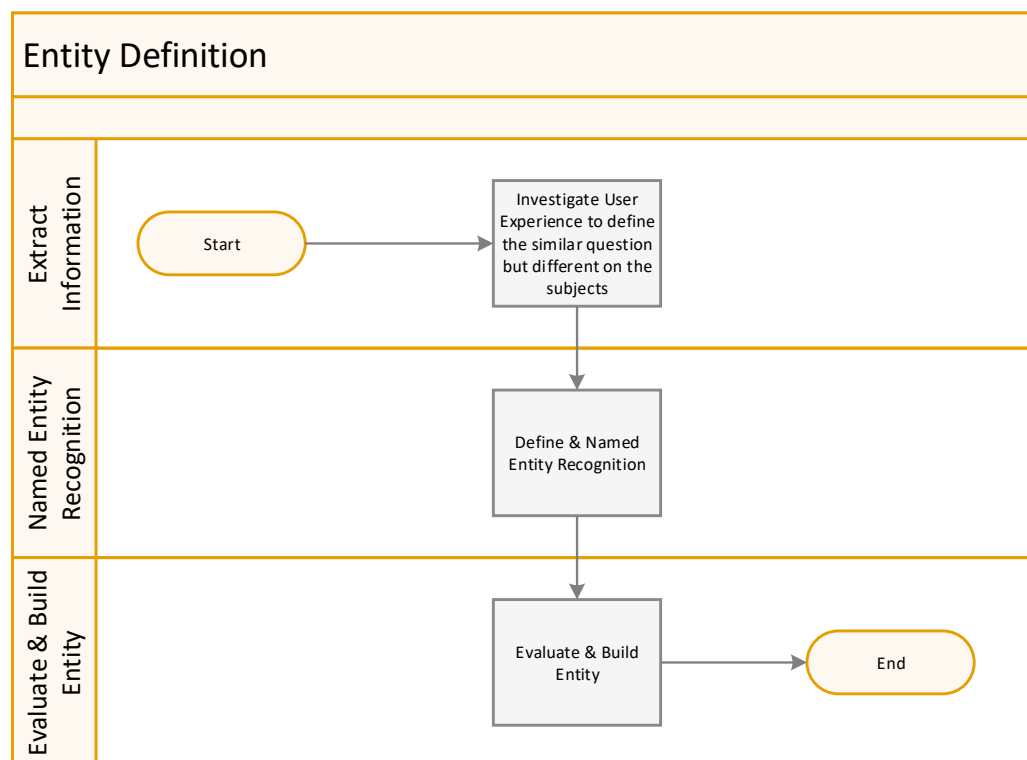
Figure 7. Process of defining Entity.

The flowchart in Figure 8 presents the process of defining Utterance. Utterances are the user input that the chatbot needs to derive intents and entities. To train any chatbot to extract intents and entities accurately from the user's dialog input, it is imperative to capture a variety of different example utterances for each and every intent.
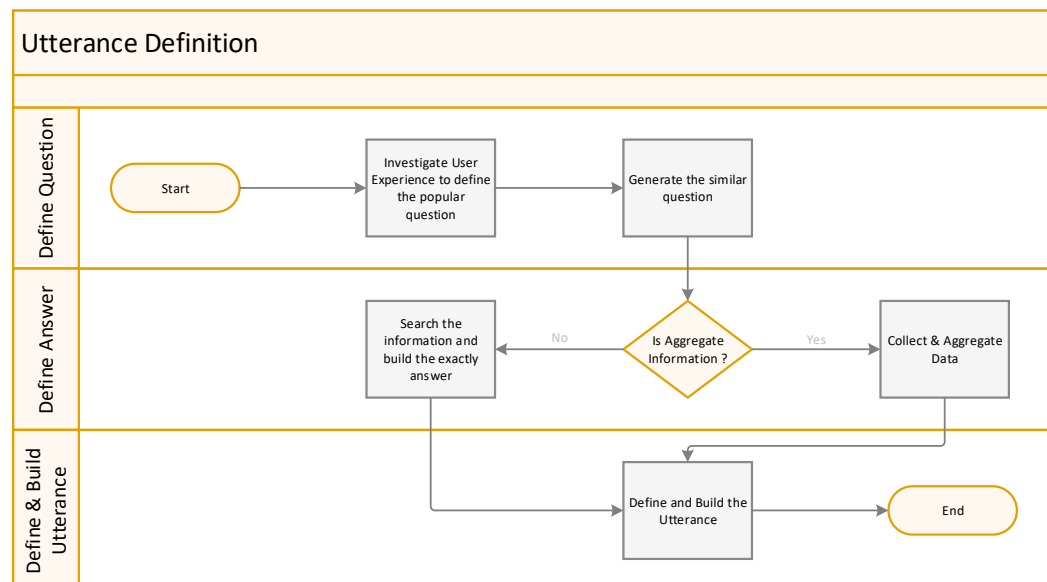
Figure 8. Process of defining Utterance.

The flowchart in Figure 9 shows the workflow of the chatbot. There are four main stages in this process: (i) Train model; (ii) User Interface; (iii) Understanding Process; and (iv) Result. The process starts in stage (i), and then AI natural language Processing and Natural Language Understanding Processing are sent to the Train and Build model. After Training, we deployed the chatbot engine in the User Interface. When the user inputs a sentence to start the conversation, this sentence is split into entities; if it has good confidence in any intent, the system will classify and print out the answer. If the confidence is not good enough, this situation will be inserted and updated in the training model. In stage (iv), after sending the output answer, if there is a similar Question, the System will send Suggest the related Information to the user for user selection. If there is none, the process ends.
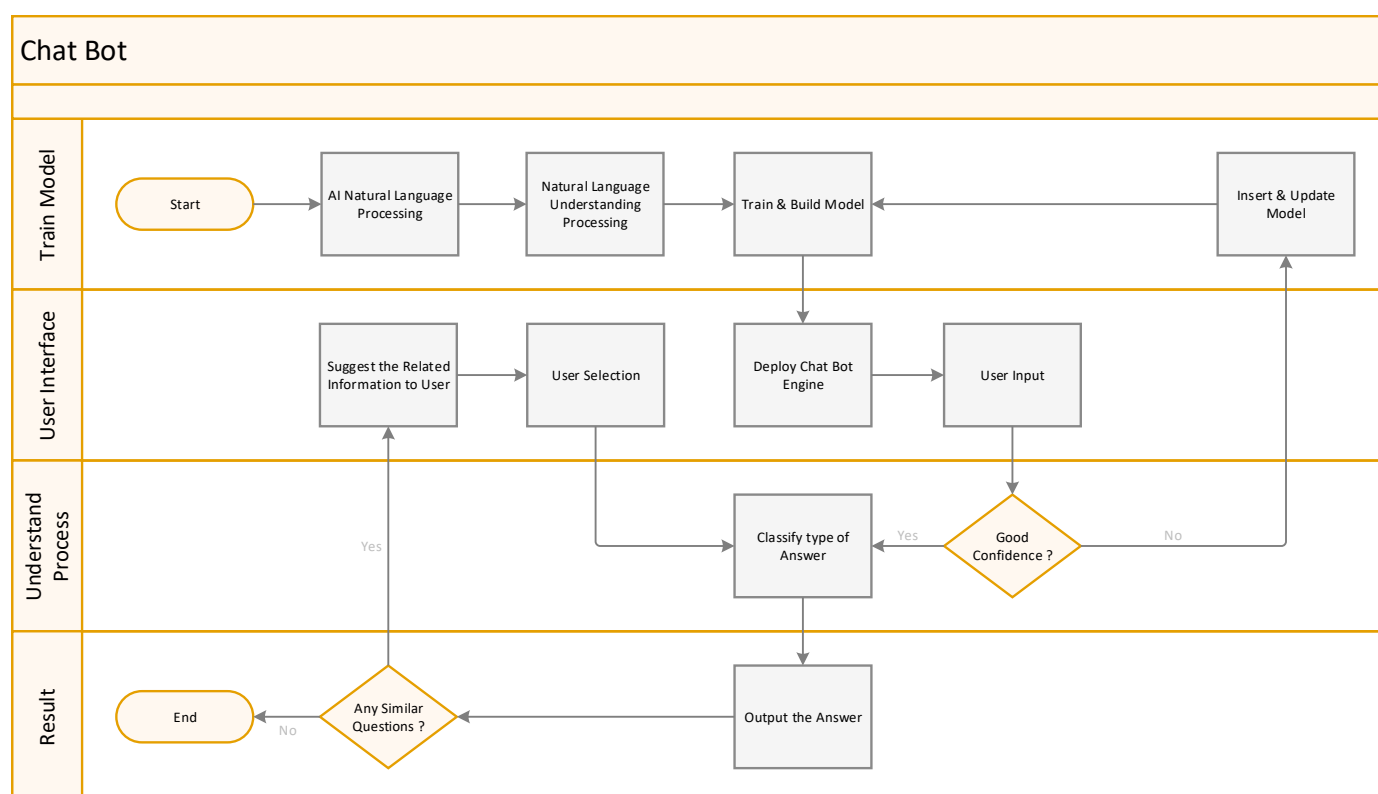
Figure 9. Flowchart of chatbot.

4.3. UML Diagram and Application

Figure 10 shows the entire UML framework of the Operational Solution Assistant system. The user can use the chatbot in MainActivity, which has the chatbot application interface. MainActivity connects with the server through the Internet network. The process starts once the user is selected, and the agreement terms for collecting and using the user's information are shown during the time of using the Application. After the user provides user_ID, the application will query and access the user database in the Server through Internet connection. This user_data contains personal data and historical data (if available) of old conversations with applications. When a user starts a conversation with our application, the application sends to the NLP and NLU functions. When the user asks the chatbot any question, the chatbot sends to the server through API. The server also connects with the NLP and NLU functions to give the output answer to the sender, which sends this message to the user.
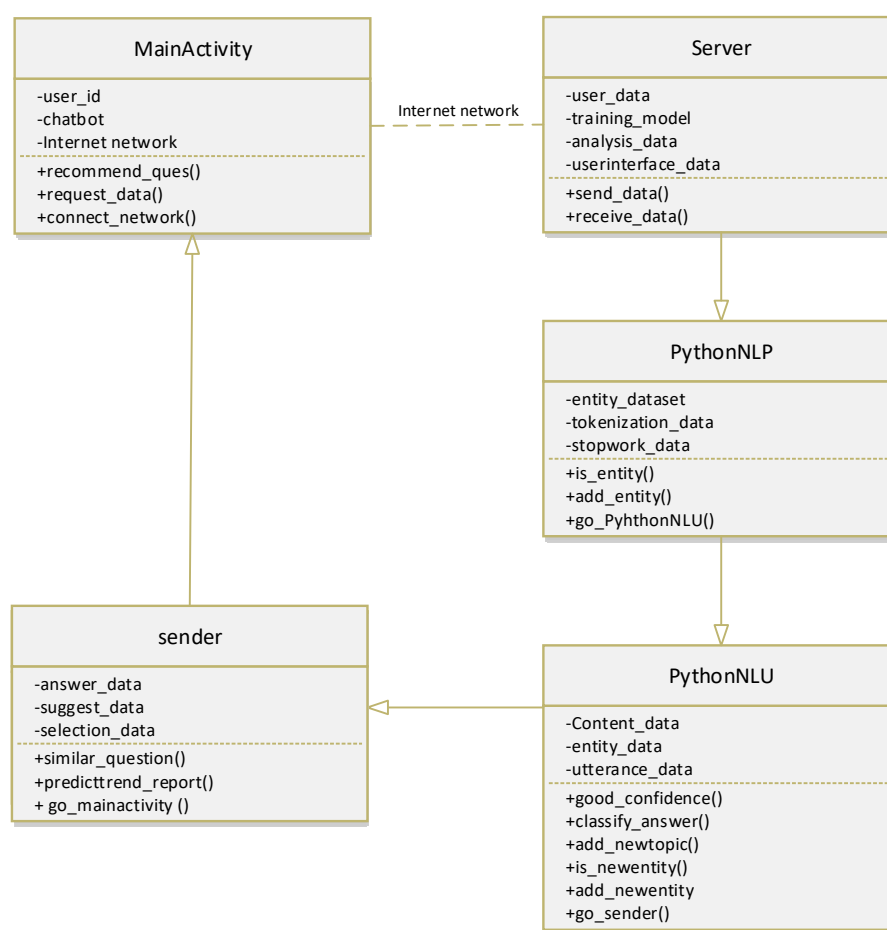
Figure 10. UML framework of the Operational Solution Assistant system.

5. Building of the Actual Application Model: Assistant Application for Foreign Companies in Korea

5.1. Building the Data Set for the System

As a result, we built a data training set that answers questions in the following main aspects: Investment in Korea: Introduction; Korea's Investment Climate; Investment Guide; Government support policies and programs, and Legal, Tax and Labor. We also built an interactive system that can enter data directly into the database and test model as shown in Figures 11–13 as below.

Figure 11. Input interface of chatbot intent data.

Figure 12. Input interface of chatbot Entity data.

Figure 13. Input interface of chatbot utterance data.

Figure 11 presents the Input interface of chatbot intent data. Developers can input the Intent code, name of intent, Synonym of intent in Korean, Intent type and Description of this intent.

Figure 12 shows the Input interface of the chatbot entity data. Developers can input Entity names, Entity value, Description of this entity and Synonym tag (the user can use many types to express a similar Entity).

Figure 13 presents the Input interface of chatbot utterance data. First, when the Developer chooses Intent code, intent name and Korean intent will appear. After that, the Developer chooses the Question's type and inputs the question with a suitable answer and the type of answer.

The Data set sample of Intent after input is presented in Figure 14, concluding the Intent code, name of intent, Synonym of intent in Korean, Intent type and Description of this intent.

Figure 14. Data set sample of Intent.

The Data set sample of Entity after inputting is presented in Figure 15, concluding Entity names, Entity value, Description of this entity and Synonym tag (the user can use many types to express a similar Entity).

Figure 15. Data set sample of Entity.

The Data set sample of Utterance after inputting is presented in Figure 16, concluding the Intent code, Intent, Intent Korean, Entity group, Question's Type, Question, Answer's Type and Answer.

Figure 16. Data set sample of Utterance.

Data is stored in MongoDB as document type which composed data as field and value pair. Value of the field can be any of data types [27], presented in Figure 17. In our research, we chose this database because of the properties of natural language. They are unstructured or semi-structured (non-structured or semi-structured) and they cannot be stored in fixed formats such as tables. In these use-case, we stored a data record field-value-pairs.

Figure 17. Intents' dataset stored in MongoDB.

In our system, after inputting Entity, Intent and Utterance, we also built functions for developer testing; the evaluation results of the chatbot are shown below Figure 18.

Figure 18. Training model and test function.

We continued with collecting investment data in ROK from public portals from the Korean government such as Korea Public Data Portal (data.go.kr (accessed on 24 July 2020)) and Busan Provincial Portal (bigdata.busan.go.kr (accessed on 24 July 2020)), as shown in the Figure 19 below. In this research, we mainly used data collected from these websites because they are accurate and are updated year by year; this is true for the investment data. The data that needs to be collected correctly will ensure the efficiency and accuracy of the system later.

Figure 19. Korea Public Data Portal (Source: data.go.kr (accessed on 24 July 2020)).

We mainly collected Invest attraction data and industry register data. Figure 20. presents the structure of specialized data of Invest attraction saved in database in our system. After processing, Invest attraction data are stored in MongoDB in the following fields: Id, Name of country, year and the amount money invest into ROK. Each data saves the information separately by ID for easier extraction and analysis.
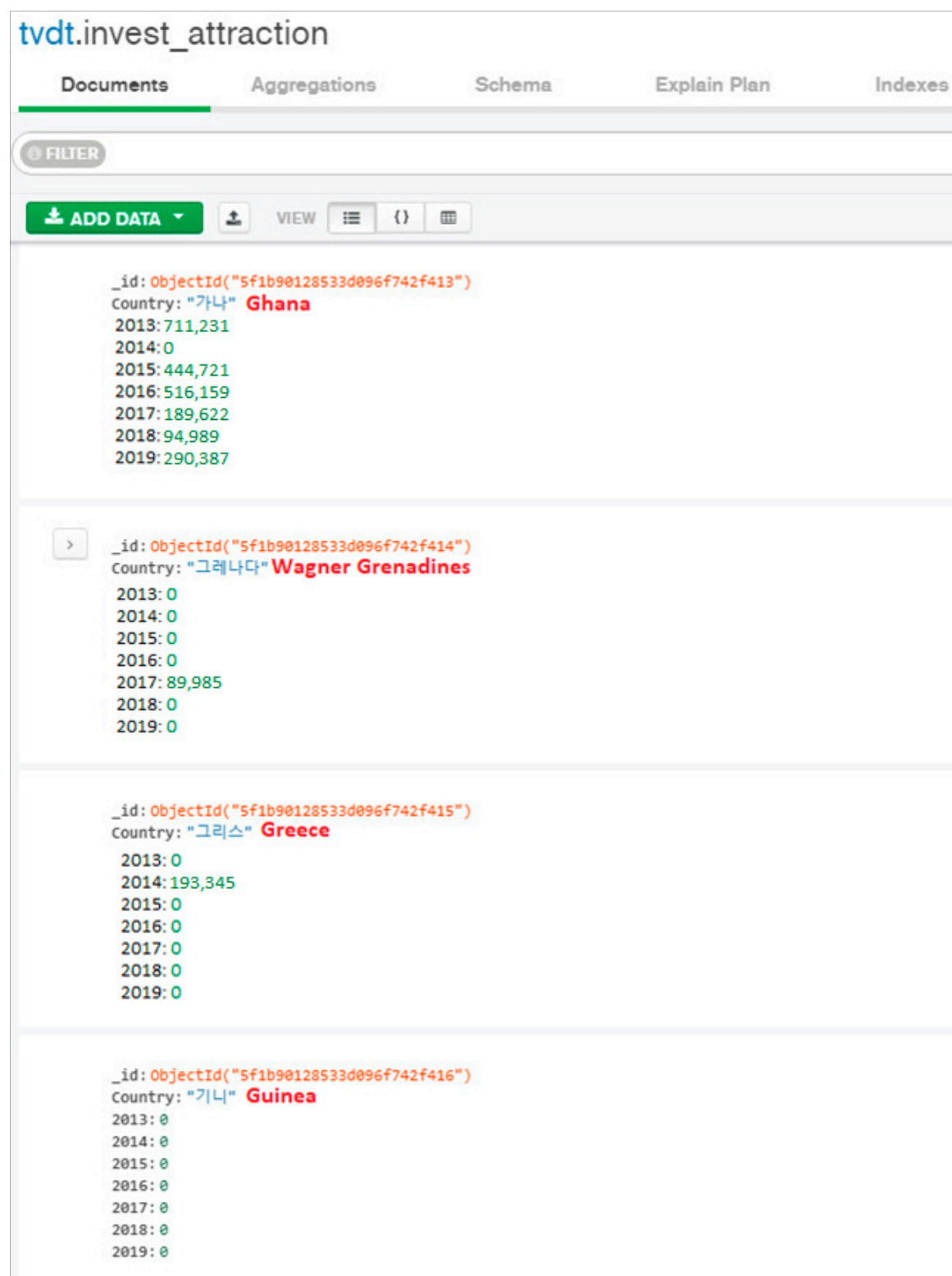
Figure 20. Databases of Invest attraction data.

Figure 21. presents the structure of specialized data of Invest attraction saved in Metadata in our system.

Figure 21. Metadata of Invest attraction data.

5.2. Data Analysis

In addition to static information such as laws and government regulations related to investing, statistical data and analysis of investment trends and business support policies were also collected and aggregated. Typically, information on the countries that invested in Korea over the years from 2013 to 2019 are shown in Figures 22a,b and 23.

Japan was the largest investor in 2013 and 2014 (2,889,075,598 KRW and 2,269,148,895 KRW, respectively), the United States in 2015 (2,354,672,216 KRW), Singapore in 2016 (2,110,752,026 KRW), Malta in 2017 (1,957,236,074 KRW), the US again in 2018 (3,803,919,878 KRW) and the Netherlands in 2019 (2,321,328,106 KRW) (Data source: www.data.go.kr) (accessed on 24 July 2020).
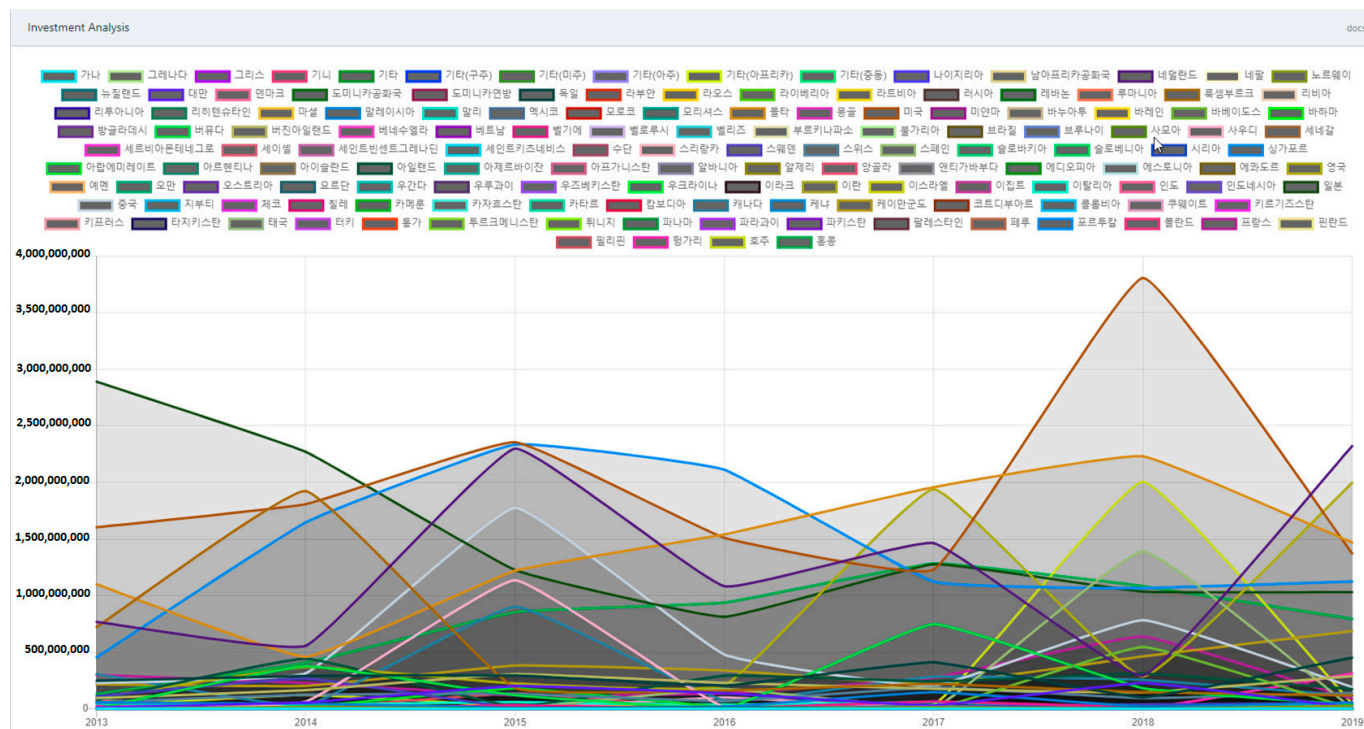
(**a**) Data analysis of investment trends

Korean	English	Korean	English	Korean	English	Korean	English	Korean	English	Korean	English	Korean	English
가나	Ghana	라오스	Laos	바하마	Bahamas	스위스	Swiss	요르단	Jordan	케이만군도	Cayman Islands
그레나다	Wagner grenadines	라이베리아	Liberia	방글라데시	Bangladesh	스페인	Spain	우간다	Uganda	코트디부아르	Ivory Coast
그리스	Greece	라트비아	Latvia	버뮤다	Bermuda	슬로바키아	Slovakia	우루과이	Uruguay	콜롬비아	Columbia
기니	Guinea	러시아	Russia	버진아일랜드	Virgin islands	슬로베니아	Slovenian	우즈베키스탄	Uzbekistan	쿠웨이트	Kuwait
기타	Etc	레바논	Lebanon	베네수엘라	Venezuela	시리아	Syria	우크라이나	Ukraine	키르기즈스탄	Kyrgyzstan
기타(구주)	Others (Europe)	루마니아	Romania	베트남	Vietnam	싱가포르	Singapore	이라크	Iraq	키프러스	Cypress
기타(미주)	Others (Americas)	룩셈부르크	Luxembourg	벨기에	Belgium	아랍에미레이트	UAE	이란	Iran	타지키스탄	Tajikistan
기타(아주)	Other (very)	리비아	Libya	벨로루시	Belarus	아르헨티나	Argentina	이스라엘	Israel	태국	Thailand
기타(아프리카)	Other (Africa)	리투아니아	Lithuania	벨리즈	Belize	아이슬란드	Iceland	이집트	Egypt	터키	Turkey
기타(중동)	Other (Middle East)	리히텐슈타인	Liechtenstein	부르키나파소	Burkina Faso	아일랜드	Ireland	이탈리아	Italy	통가	Tonga
나이지리아	Nigeria	마셜	Marshall	불가리아	Bulgaria	아제르바이잔	Azerbaijan	인도	India	투르크메니스탄	Turkmenistan
남아프리카공화국	Republic of South Africa	말레이시아	Malaysia	브라질	Brazil	아프가니스탄	Afghanistan	인도네시아	Indonesia	튀니지	Tunisia
네덜란드	Netherlands	말리	Mali	브루나이	Brunei	알바니아	Albania	일본	Japan	파나마	Panama
네팔	Nepal	멕시코	Mexico	사모아	Samoa	알제리	Algeria	중국	China	파라과이	Paraguay
노르웨이	Norway	모로코	Morocco	사우디	Saudi Arabia	앙골라	Angola	지부티	Djibouti	파키스탄	Pakistan
뉴질랜드	New Zealand	모리셔스	Mauritius	세네갈	Senegal	앤티가바부다	Antigua & Barbuda	체코	Czech Republic	팔레스타인	Palestine
대만	Taiwan	몰타	Malta	세르비아몬테네그로	Serbia and Montenegro	에디오피아	Ethiopia	칠레	Chile	페루	Peru
덴마크	Denmark	몽골	Mongolia	세이셸	Seychelles	에스토니아	Estonia	카메룬	Cameroon	포르투칼	Portugal
도미니카공화국	Dominican Republic	미국	United States of America	세인트빈센트그레나딘	Saint Vincent &Grenadine	에콰도르	Ecuador	카자흐스탄	Kazakhstan	폴란드	Poland
도미니카연방	Dominican Federation	미얀마	Myanmar	세인트키츠네비스	Saint Kitts and Nevis	영국	England	카타르	Catarrh	프랑스	France
독일	Germany	바누아투	Vanuatu	수단	Way	예멘	Yemen	캄보디아	Cambodia	핀란드	Finland
라부안	Labuan	바레인	Bahrain	스리랑카	Sri Lanka	오만	Oman	캐나다	Canada	필리핀	Philippines
호주	Australia	바베이도스	Barbados	스웨덴	Sweden	오스트리아	Austria	케냐	Kenya	헝가리	Hungary

(**b**) Country name in English

Figure 22. Data analysis of investment trends.

Country	2013	2014	2015	2016	2017	2018	2019
일본 Japan	2,889,075,598	2,269,148,895	1,225,959,585	812,817,461	1,277,065,270	1,034,715,269	1,030,575,288
미국 America	1,602,776,652	1,807,381,639	2,354,672,216	1,510,491,290	1,229,404,518	3,803,919,878	1,370,210,250
몰타 Malta	1,100,306,719	469,860,481	1,220,119,715	1,540,210,023	1,957,236,074	2,229,704,758	1,468,017,707
네덜란드 Netherlands	767,694,298	557,272,243	2,297,964,109	1,083,009,389	1,462,648,812	294,354,386	2,312,328,106
룩셈부르크 Luxembourg	722,056,735	1,922,356,951	183,635,901	162,064,033	219,519,200	146,878,416	119,521,187
싱가포르 Singpore	454,217,201	1,646,713,949	2,333,793,564	2,110,752,026	1,122,520,804	1,065,038,152	1,124,606,217
캐나다 Canada	307,323,729	29,164,936	902,078,655	75,863,100	284,082,785	256,392,235	115,384,155
프랑스 France	299,738,757	228,550,568	119,138,315	136,187,917	275,356,898	638,586,604	88,710,523
독일 Germany	248,773,456	291,037,482	295,518,789	205,046,035	410,759,132	102,829,490	452,416,391
중국 China	229,100,555	314,745,559	1,774,417,420	479,489,063	201,808,811	783,508,665	192,335,830

Figure 23. Data statistics of investment trends.

In addition, data on the most supported industry from the Korean government from 2014 to 2018 for each region were collected, with the basic analysis shown in Figures 24 and 25. A comprehensive picture of supported industries in each region from which investors can make appropriate decisions is presented.
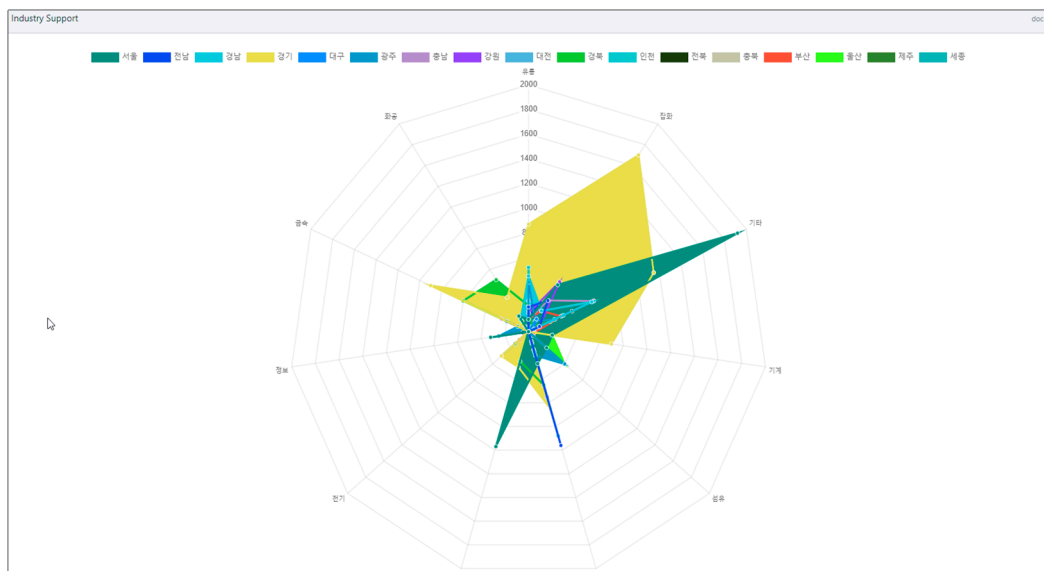
(a) Data analysis of the supported industry

서울	Seoul	종남	Chongnam	전북	Jeonbuk
전남	Jeonnam	강원	Gangwon	중북	Chungbuk
경남	Gyeongnam	대전	Daecheon	부산	Busan
대구	Daegu	경북	Gyeongbuk	울산	Ulsan
광주	Gwangju	인전	Incheon	제주	Jeju
세종	Sejong				
화공	Shoemaker	전자	Electronic	기타	Etc
금속	Gold	식료	Food	잡화	Accessories
기계	Machine	성유	Chrism	유통	Distribution
전기	Electricity				

(**b**) Country and Industry translated to English

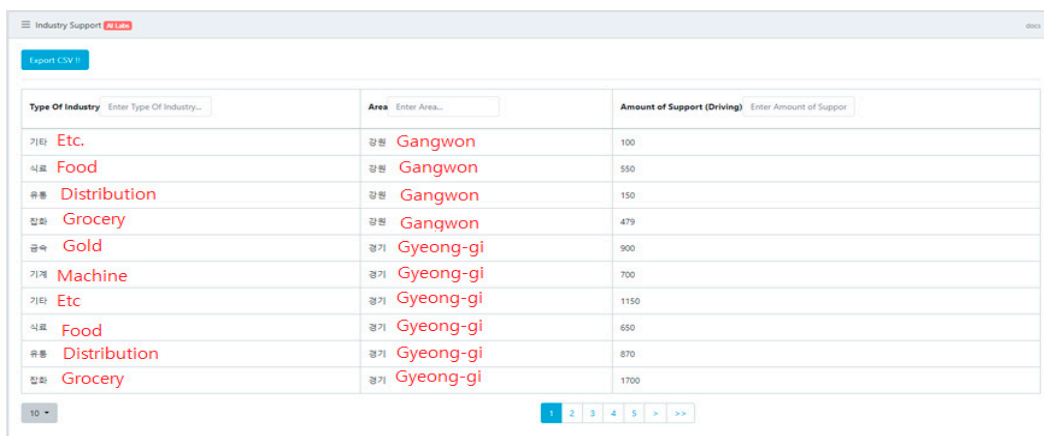**Figure 24.** Data analysis of the supported industry.

Figure 25. Data statistics of the supported industry.

Finally, in order to assist foreign businesses in seeking information on cooperation with Korean businesses, enterprise information was collected as shown in Figure 21. With the information above, the aggregation and analysis data set were built to help the Virtual Assistant learn data.

6. Result: Operational Solution Assistant System

6.1. Operational Solution Assistant: Chatbot

We built an Assistant system, a smart assistant for foreign investors looking for information such as Investment in Korea: Introduction, Korea's Investment Climate, Investment Guide, Government support policies and programs and Legal, Tax and Labor. Users receive different types of information.

As shown in Figures 26 and 27, when the user asks the Assistant in a narrative sentence, not in question form, about which type of visa allows opening companies in Korea, the Assistant can give a correct answer. The Assistant system can also interact with users naturally like a human.
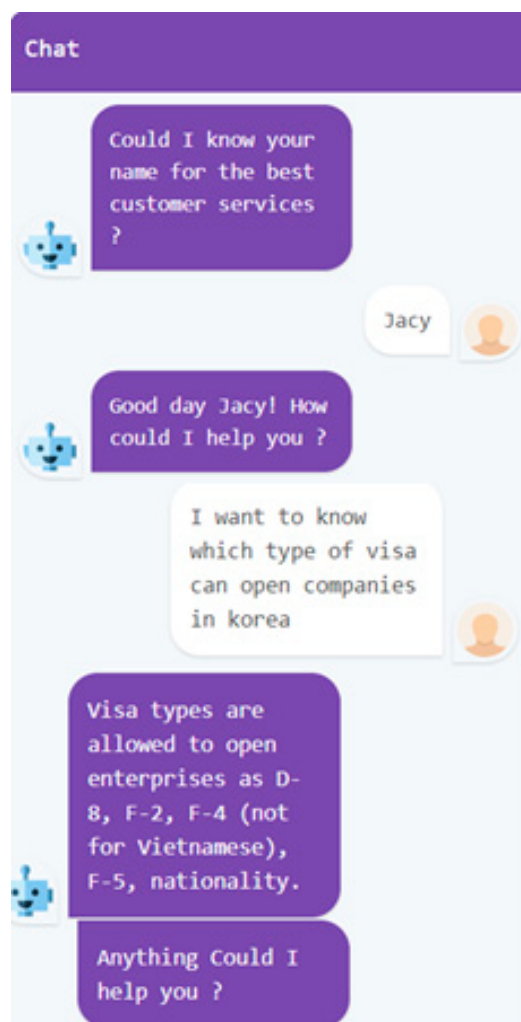
Figure 26. Assistant Application providing legal information.

Figure 27. Assistant Application answering the question on the industry most supported in Busan.

In addition, the Assistant Application can provide analysis information as shown in Figure 28. We collected data about the supported industries in regions of Korea and made some analysis, so our assistant can also provide the following analysis results.

Figure 28. Assistant Systems providing information about other companies in Korea.

As shown in Figure 29, the Assistant Application can provide information on other companies in Korea such as address and contact details.

Figure 29. Assistant Systems providing information about the country investing the most in Korea.

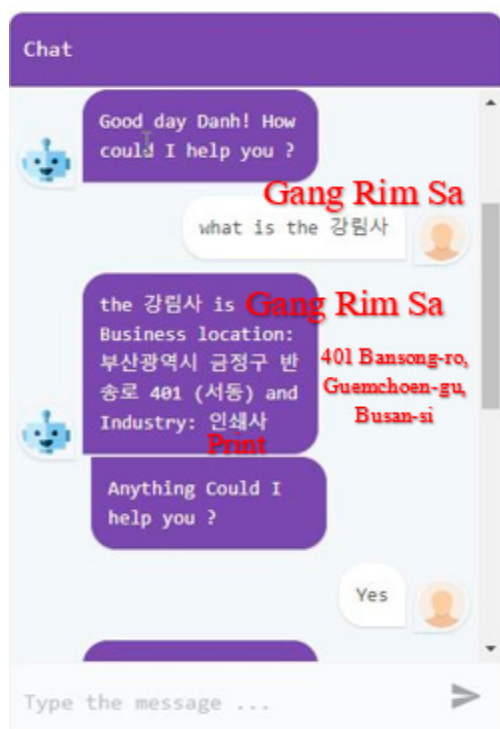Our Assistant Application also available on websites, user can get access and help to improve performance and get more conversational data recorded by responding to customers. To evaluate the performance of the application, we set up a realistic experiment with 20 colleagues and friends to test the performance of all functions in our system as well as estimate the accuracy of chatbot. First, all the functions (chatbot, visualization and developer) worked well and make a friendly environment for user to use and looking for information and seeking investment opportunities in ROK. Second, we calculate the accuracy of chatbot by the percent of correct answers or suggestion of the questions or interactive communication with users. Main status can be appeared: (a) chatbot gives correct answer, (b) chatbot gives wrong answer, (c) chatbot gives a correct answer but not understanding or missing the referred context. When testing, most of the questions chatbot can answer, the number of wrong answers is very small (accounting for 5 wrong answers out of 92 answers). Since the number of testers and database is still very small, at this stage, the accuracy of the chatbot will be high. Future. We will expand the database as well as the number of testers to come to more accurate conclusions.

6.2. Web-App Visualized Analysis Investment Data

As the result, we also build a web-app to visualize analysis investment data collected from public website in ROK. Typically, we collected data about countries that invested in ROK, business registration information for new companies, industry investment as well as the amount of investment companies over the years from 2013 to 2019 from public provider website in ROK (data.go.kr). After pre-processing and aggregating with Python language, we built tables, simple charts to visualize for user easily access and use information.

7. Conclusion and Future Work

This study used an algorithm that focuses on building an Assistant system to help foreign investors pass language barriers and cultural and institutional differences by providing tools to search for more information and seek investment opportunities in ROK. We designed our Virtual Assistant system that can crawl and analyze data on foreign investments in ROK from open data resource websites and used analytic and aggregation techniques to explore trends in investments of foreign enterprises. We built data sets about information and investment data for SMEs in ROK, concludes government support policies, laws, taxes, investment data, new established enterprises data. In this research, we

built a friendly framework for new developers to add and build up the dataset of AI Assistant by providing input intent data function, input Entity data function, input utterance data function as well as training and test function. In addition, we used cloud computing combined with programming techniques such as Python natural language processing, natural language understanding and web application to support this project in successfully building the processing core engine of a virtual assistant. In addition, we built a web-app connected to the server to visualize all the results of analyzed collected data so that SMEs owners can easily use and look for information on investments. Based on the research results, we can make recommendations to SMEs in keeping with the changing investment trends in ROK.

In our research, we have only focused on collecting and constructing investment data sets (concludes government support policies, laws, taxes, investment data, new established enterprises data) at ROK, so the data set is not yet diverse. These data just public data so they may not be complete and accurate. In order for the application to be able to develop more generator and generalize information, it is necessary to cooperate with governments and Korean and foreign enterprises to build big data collection, build dataset updated year by year as well as other useful information such as investment surveys, information on government assistance programs as well as surveys on SMEs difficulties. In this study, assess the level of approval and performance of the virtual assistant is not discussed, so it has not yet evaluated the practical applicability of the system. In the future, we will do more empirical studies in SMEs in ROK. As a next step, the big data set will have more information related to investment issues in Korea. Especially, real-time market analysis data and forecasts that foreign investors desire will be provided through cooperation with the government and domestic and foreign organizations in Korea such as KOTRA, FORCA, Invest Korea, etc. to complete the project and bring more value. Building virtual assistants friendly and make it easier for other developers who do not need strong technical knowledge to apply in other fields such as health care, social welfare, education and insurance is also the future works of our research.

References

1. Damijan, J.P.; Knell, M.; Majcen, B.; Rojec, M. The role of FDI, R&D accumulation and trade in transferring technology to transition countries: Evidence from firm panel data for eight transition countries. *Econ. Syst.* **2003**, *27*, 189–204, doi:10.1016/s0939-3625(03)00039-6.
2. Overseas Economic Research Institute Knowledge-Economy. *Foreign Direct Investment Trend Analysis 2019*; Korea Import & Export Bank; Overseas Economic Research Institute: Seoul, Korea, 2019.
3. Department Foreign Investment. *Report of Foreign Direct Investment in 2019*; Ministry of Planning and Investment of Vietnam: Hanoi, Vietnam, 2019.
4. So, W. *Important business challenges for SMEs in South Korea 2018*; Future of Business Survey: 2018. Available online: https://www.statista.com/statistics/881350/south-korea-major-challenges-facing-small-and-medium-sized-businesses/ (accessed on 10 August 2020).
5. Bishop, B. *Foreign Direct Investment in South Korea: Role of State*; Ashgate Publishing: New York, NY, USA, 1997.

6. Grossman, S.J.; Hart, O.D. One share-one vote and the market for corporate control. *J. Financ. Econ.* **1988**, *20*, 175–202, doi:10.1016/0304-405x(88)90044-x.
7. Lee, J.Y.; Kwak, J.; Kim, K.A. Subsidiary goals, learning orientations, and ownership strategies of multinational enterprises: Evidence from foreign direct investments in Korea. *Asia Pac. Bus. Rev.* **2014**, *20*, 558–577, doi:10.1080/13602381.2013.859427.
8. Bae, K.-H.; Kang, J.-K.; Kim, J.-M. Tunneling or value added? Evidence from mergers by Korean business groups. *J. Financ.* **2002**, *27*, 2695–2740.
9. Business Doing. *Comparing Business Regulation in 190 Economies*; World Bank Group: Northwest, WA, USA, 2019.
10. Jäger, G.; Rogers, J. Formal language theory: Refining the Chomsky hierarchy. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367*, 1956–1970, doi:10.1098/rstb.2012.0077.
11. Yao, M. *Does Conversation Hurt Help the Chatbot Ux*? Smashing Magazine: Freiburg, Germany, 2016.
12. Chowdhary, K.R. Natural Language Processing. In *Fundamentals of Artificial Intelligence*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2020; pp. 603–649.
13. Dan Jurafsky, J.; Martin, H.; Kehler, A.; Linden, K.V. *Nigel Ward, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*; Prentice Hall: Hoboken, NJ, USA 2000.
14. Liddy, E.D. *Natural Language Processing*; Encyclopedia of Library and Information Science: New York, NY, USA, 2001.
15. Basili, R.; Pazienza, M.T.; Velardi, P. An empirical symbolic approach to natural language processing. *Artif. Intell.* **1996**, *85*, 59–99, doi:10.1016/0004-3702(95)00116-6.
16. Webster, J.J.; Kit, C. Tokenization as the initial phase in NLP. as the initial phase in NLP. In Proceedings of the 14th conference on Computational linguistics—Volume 4, Stroudsburg, PA, USA, 23–28 August 1992; pp. 1106–1110.
17. James, A. *Natural Language Understanding*; Pearson Education: London, UK 1995.
18. Sebastian, H. Amazon Extends Alexa's Reach into Wearables. *WSJ,* 26 September 2019.
19. López, G.; Quesada, L.; Guerrero, L.A. Alexa vs. siri vs. cortana vs. Google assistant: A comparison of speech-based natural user interfaces. In Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, Los Angeles, CA, USA, 17–21 July 2017, pp. 241–250.
20. Adamopoulou, E.; Moussiades, L. An Overview of Chatbot Technology. In *Collaboration in a Hyperconnected World*; Springer Science and Business Media LLC, 2020; Vol. 584, pp. 373–383.
21. Dumais, S.T. Latent semantic analysis. *Annu. Rev. Inf. Sci. Technol.* **2005**, *38*, 188–230, doi:10.1002/aris.1440380105.
22. Marietto, M.D.G.B.; Aguiar, R.V.; Barbosa, G.D.O.; Botelho, W.T.; Pimentel, E.; Franca, R.D.S.; Da Silva, V.L. Artificial Intelligence Markup Language: A Brief Tutorial. *Int. J. Comput. Sci. Eng. Surv.* **2013**, *4*, 1–20, doi:10.5121/ijcses.2013.4301.
23. Baby, C.J.; Khan, F.A.; Swathi, J.N. Home automation using IoT and a chatbot using natural language processing. In Proceedings of the 2017 Innovations in Power and Advanced Computing Technologies (I-PACT), Vellore, India, 21–22 April 2017; pp. 1–6.
24. Shah, R.; Lahoti, S.; Lavanya, K. An intelligent chat-bot using natural language processing. *Int. J. Eng. Res.* **2017**, *6*, 281–286.
25. Ait-Mlouk, A.; Jiang, L. KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding Over Linked Data. *IEEE Access* **2020**, *8*, 149220–149230, doi:10.1109/access.2020.3016142.
26. Abdellatif, A.; Badran, K.; Costa, D.; Shihab, E. A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering. *IEEE Trans. Softw. Eng.* **2021**, doi:10.1109/tse.2021.3078384.
27. Chodorow, K.; Dirolf, M.; Mongo, D.B. *The Definitive Guide*, 1st. ed.; O'Reilly Media, Inc: Newton, MA, USA, 2010.