

# The Neural Network Conversation Model enables the Commonly Asked Student Query Agents

Nittaya Muangnak<sup>1</sup>, Natakorn Thasnas<sup>2</sup>, Thapani Hengsanunkul<sup>3</sup>, Jakkarin Yotapakdee<sup>4</sup>

Department of Computer Science and Information Technology<sup>1,3,4</sup>

Department of Electrical and Computer Engineering<sup>2</sup>

Faculty of Science and Engineering, Kasetsart University  
Sakon Nakhon, Thailand

**Abstract**—One of the challenges in academic counselling is to provide an automated service system for students. There several query questions asking the faculty staffs about related-academic services each semester. Offered the communication interface more convenience, the novel approach based on neural network model is introduced to investigate the automated conversational agent. The pre-defined dialogue sentences were collected manually from the student query questions and used as the training dataset. The questions have been varied and grouped by topic-categorizing queried from the registration help desk of the department. Artificial intelligence and machine learning have contributed each other to build the conversational agent so-call KUSE-ChatBOT plugged and used in the modern messenger application, LINE. The system is also included the dialogue back-end management system to use in further deep learning model updating. Tensorflow, the machine learning development platform originated by Google, was performed and obtained the learning model using Python development kits. The LINE Messaging APIs is then contributed as the user interface where users could have FAQs' conversation via the LINE application. The KUSE-ChatBOT is outperformed and efficient by providing automated consultation to the students precisely with the accuracy rate over 75 percent. The system could assist the staffs to be able to lessen the workload of answering the same question repeatedly and give response to the student timely.

**Keywords**—Automated conversational agent; chatbot; natural language processing; FAQs' bot; artificial neural network; artificial intelligence; machine learning

## I. INTRODUCTION

Smart conversational agents so-called CHATBOTs have become popular among people in everyday life for specific purposes. Chatbot [1] is a computer program which is developed through artificial intelligence techniques to play a role in automatically responding to conversations as requested by users via text messages or voice commands. “talkbot”, “bot”, “IM bot” or “interactive agent” are the synonyms for the chatbot. There have been well-known chatbots used in everyday life, Google Assistant and Siri. These chatbots can help to manage schedules, report weather, including providing various assistance. It is mostly used in online business applications in the provision of intelligent assistants to interact with humans automatically. There exist several related applications investigating the CHATBOT.

ERASMUS [2] is a chatbot on Facebook, which is used to answer questions related to the college information. This

chatbot helps users to find the needed information by clicking and typing questions. ERASMUS is based on AI methodology. Users can search for activities related to the college through the system. They do not need to go to the college website to ask for any further details.

MoSHCA [3] is an intelligent chatbot which helps patient-doctor interaction and to support the self-management of chronic diseases for the patients themselves. MoSHCA medical and well-being decision support through embedded software in mobile devices by utilizing specific sensors and data from customized information systems. Besides, there are several modern conversational agents based voice-activated system found popular among end-users with noticeable features [4], for example, Apple Siri, Simsimi, Microsoft Cortana, Google Now, Facebook Messenger [5], and Amazon Alexa.

However, the above-mentioned chatbots are suitable for the general-purpose assistant. The faculty of Science and Engineering, Kasetsart University, Sakon Nakhon, Thailand, has provided educational services to students each year. Students inquiring varying information each day, for example, general information of the faculty, enrollment, application for scholarships, and student activity credits. The faculty provides channels for the students to contact both online and offline, typically contacting in person at the help desk and asking via Facebook fan page. There was still trouble causing the staff unable to answer the question instantly. Therefore, the idea in the development of automatic conversational agents via the application Line, the new developing service, is introduced.

Several approaches were combined with utilizing conversational agents. A text-based conversation is commonly employed to the classical pattern matching criteria. This simple technique could be produced erroneous replied sentences. Early artificial intelligence-based method referred to rule-based approach is also allowed more alternative response for user inputs. This approach unlikely remains constraint when there is not a diversity of sentence patterns. Solely simple sentence patterns could be recognized into the bot brain and returned minor human-like response [6][7]. Dialogue initiation has the major play a role in leading the agents becomes a human-like conversation as well. Some of the early text-based chatbots found that the grammatical structure has been required to improve the context understanding. Even though modern chatbots employ

generative knowledge-based techniques and produce more precise conversation response, machine-based learning technique is also challenging breakthrough those mentioned limitations. Conversational agents in the traditional language, Thai language, in academic-proposed especially an automated service in information retrieval, are also limited to in-proposed organizations. Therefore, the novel approaches based on machine learning are introduced to investigate natural language processing to build the automated conversational agent. <reviewer 1 response: list reasons for choosing proposed methods to address the problems including emphasize the limitations of the existing methods.

The importance of this article is to introduce the proposed methods in applying neural network based method to develop an automated conversational agent for Frequently Asked Questions (FAQs) in Registration Division Help Desk. The modern developing service to allow user interface connection is the LINE Messaging APIs. An intent identification method is also investigated to stretch out to the meaning of the input sentence. The objective of the study is to provide automated consultation for students promptly and more convenient and faster operation. The additional information for students to prepare before contacting the office help desk is included in the proposed answer. It also allows users to give feedback based on the satisfaction reply by the Chatbot. The frequently asked questions collected from staff and the social media were used as the training data collection. Natural language processing (NLP) and natural language understanding (NLU) principles are integrated with machine learning techniques to develop the chatbot model. Moreover, history conversation is also required to be kept for further chatbot brain reconstruction.

The rest of this article is arranged as follows. This section provides background, chatbot, problem statement, and purpose of the study. Section 2 includes literature reviews on automated conversational agents applied in several existing works and related background. The proposed methods in the automated conversational agent for FAQs in the registration division is describes Section 3. Section 4 describes the data collection as well as evaluation schemes used in the experiment. Corresponding experimental results including discussions of the study are stated in Section 5. Section 6, which is the last section, summarizes the whole study and provides for the future work, including other perspectives.

## II. RELATED WORKS

The relevant studies on conversational agent development approaches are presented in this section.

Amongst the current techniques, the methodology of developing chatbot is how to train the bot interacting with users simultaneously. Rule-based method is the conventional methodology to construct a chatbot model by detecting the keyword from a user question then match with answer from rule engine database. It has to be the exact keyword found in the sentence pattern from the rule engines. To allow users to communicate naturally, high level chatbot based upon artificial intelligence (AI) is introduced to simulate conversational agents as similar as chat with the real humans. It provides an answer to users more genuinely and

understandable on the sentence patterns and meanings. The chatbot model's performance relies on training and testing dataset which is the data collection of prior questions and answers. There exist some related works studying to employ the conversational agent approaches in applications.

The very first chatbot known as ALICE [8], stands for Artificial Linguistic Internet Computer Entity, is a conventional chatbot based on Alan Turing's test in 1950. Alice introduced the Artificial Intelligence Markup Language (AIML) and Extensible Markup Language (XML) to design the stimulus-response Q&A system of the chatbot.

Tantibaphanrak [9] introduced the tutor supporting agent for the classroom. The dialogue collected from students and instructors has been used to train the tutor agent model based on machine learning and NLP. The dialogue collected from the students and the instructor were used as the rule-based training dataset. The current state of the conversation has been annotated and linked to the intent. Once the intent was obtained using the prediction model, the answer was selected from the next sentence consecutively. This approach indicates that the conversational agent of the tutoring system is not too diverse. The predefined penalty matching with the input sentence has been limited by the intents to fit the model. Better the prediction model accuracy using comparison-based methods, Naïve Bayes and decision tree, the dialogue patterns from different tutors could be varied. The accuracy of the prediction model was performed maximum up to 76.5 percent.

Tangkathach [10] presented the learning portal system for the chatbot. The relevant contents obtained Label the English grammar elements with Semi-CRFs from the website were extracted and collected to build conversation dialogue. In creating the dialogue conversation, the domain knowledge was used to differentiate the sentences patterns in English. A set of word groups that describe additional meanings in the main sentence was obtained including a dictionary from WordNet's vocabulary database and irregular verb variation datasets. In the experiment, the questions producing answered accurately were found in the sentence's datasets. The construction of sentence patterns satisfied most of the input sentences. However, there is still a sentence ambiguity found during the chatbot could not provide the correct answer or without interaction by 31.07 percent and 4.00 percent, respectively. The performance measurement has reported representing correctly 64.93% out of the total number of questions.

Another alternative application deploying the conversational agent for Pakistani fashion brands was represented by Nazir et al. [11]. This study is based on the uses of ontology-driven information retrieval. Non-related questions were filtered and revised regarding the dialogue brand information and the ontology scope. The revised question sentences were processed to obtain the corpus before constructing the semantic ontology framework. The global cardinality constraints on properties were also defined in-out relationship to retrieve the replied messages and information. This study mainly performs depending upon the ontology structure without cultivating on the principles of natural language processing and natural language understanding.

Various researchers have attempted to manipulate the conversational agents in-purposed organization assistance services. Li et al. [12] introduced a persona-based approach to improve the conversational agents to be able to respond naturally and feel like having a conversation with the individual characteristic. Oh et al. [13] applied emotional recognition based analysis and multi-modal approach to improve the communication that is suitable for mental health counselling for patients through the conversational service platform. A survey on the recent literature by Laranjo et al. [14] cited 14 conversational agents without constrained NLP qualification from 17 articles. The survey was mainly focused on the characteristics, current applications, and performance measures of chatbots. Most of the articles mentioned the useful of chatbots to assist human querying self-treatment recommendation and prior symptom diagnosis before getting harmful. The performance measurements of the reviewed studies were mainly human expert evaluation. In particular, rule-based, frame-based and agent-based approaches used in finite-state dialogue management systems and their combination were proposed in [15], [16], and [17]. Recent advanced and desirable techniques in machine learning-based methods and a drawn-out motivation in neural network model to the conversational agent development having more complexity and high efficiency were suggested in [4], [18], [19], [20], [21], and [22].

Academic simultaneously, there exist researches have been cooperated the help desk conversational agents in academic-purposed services. Hijjawi et al. [23] introduced the text-based conversational agent in Arabic language so-called ArabChat to assist the students' information queries. The ArabChat natural language generation is based on the hybrid approaches of rule based and pattern matching technique. Subramaniam et al. [24] have introduced a cognitive multi-bot conversational framework for technical support, generally referred to COBOT. Knowledge-based and information retrieval engine have been considered to contribute the conversational agent construction. The COBOT efficiency depends on the knowledge graph structure. The wrong answer could be yielded when there was more than one intent path found in queried words. One of the most relevant studies is proposed by Latorre-Navarro and Harris [25] that also applied an artificial intelligent technique to develop the natural language conversation system in the university department. The approach has provided web-based chatbot for giving academic advice to the bachelor students. Knowledge-based and expert system have been involved to obtain the conversational agent. The natural language understanding was determined to include the grammar rules or part of speech in the stage of natural language generation. This application is useful for students to give more information especially in customed-study plan. However, the authors lead to the future work that statistical learning technique could be considered to the natural language processing.

### III. PROPOSED METHODS

This section presents the proposed methods utilized in this study including dataset acquisition and preparation, intent identification method using neural network model and one-hot encoding. The proposed methods are divided into three parts

(Fig. 1 illustrates the conceptual framework of the proposed methods) which are: first is the administration management system for the KUSE-Chatbot providing the FAQs' dataset for further model training process. The conversational agent machine learning (ML) model is also investigated to carry out the automatic intent identification system. Finally, the ML model would be integrated with the LINE Messaging APIs to implement application programming interface for users via LINE application.

#### A. FAQs and Personalized Data Preparation for the Chatbot

The frequently asked questions and answers (FAQs) regarding the services related to students within the Faculty of Science and Engineering were collected and used as the dataset in this project. The dialogue was gathered from two main solutions. First solution is the dialogue acquired face-to-face at the faculty help desk serving by the faculty staffs. The second channel is taken from inquiries via social media, including the private groups of the students via Facebook and LINE. The dialogue has been considered to classify into the relevant inquiry cases. Each matched question and answer were checked the grammar correction according to the Thai language structure by the experts. All three officers are assigned to be the language experts. All questions were sent to the 5 different students for generating five more sentences per each sentence. Each question offered to the model training process supposed to get five different sentence patterns but remained the same semantic. The preprocessed dataset was stored in Spreadsheet format with four columns: intents, question patterns, corresponding answers, and primary answers.

To evaluate the proposed methods, hand-drawn ground truth (GT) intents were annotated from human experts. Each language expert was asked to hand-drawn classify into the relevant inquiry cases or intents. The inter-observer between different experts was determined by finding the ratio of the correct answers that at least two out of three intents are intersected. The inter-observer variability is totally 1.00 for the question collections due to the leading terms in each intent were easily to classify.

The 100 sets of questions and answers from 14 categories are labeled as the intents of the conversation dialogues. The questionnaires are circulated to the students for entering different five sentences for each intent. The collected data has been manipulated by the MySQL database management system.

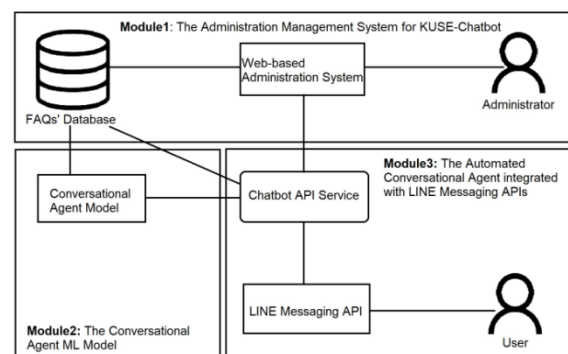


Fig. 1. The Illustration of the Conceptual Framework of the Proposed Methods.

The data preparation for the chatbot model training process is based on the intents and its following five various sentence patterns. The examples of the training data are shown in Table I which is for the intent “การแบ่งชำระค่าเทอมหรือชำระล่าช้า (a tuition fee loan or late tuition fee payment)”.

### B. One-Hot Encoding based Natural Language Preprocessing for the Training Dataset

The collected dataset has been processed through the natural language processing and binary encoding before transferring into the training process. The natural language processing steps implemented in this study are described sequentially as follows.

1) *Word segmentation* [26] is the most significantly pre-processing of natural language processing (NLP) for non-boundary delimiters language such as Thai. A challenging of the problem is the ambiguity of the boundary to decide the proper morpheme that may result in misinterpretations.

In this regard, researchers proposed several algorithms for Thai’s word segmentation. The traditional algorithm namely Longest Matching was proposed by [27]. It considered the word by scanning the characters in the sentence from left to right direction, then compared the word to a dictionary until it found the longest word. However, the algorithm still has the problem called word ambiguity. This problem can be caused when the first longest word was determined as the meaningful word at the first time, while the next longest word was not considered.

To cope with this problem, [28] proposed the algorithm called Maximal Matching. This algorithm considered only all possible words in the sentence from the dictionary to determine the boundary of a word. However, all correct words cannot be achieved, and the unknown word can occur because it depended only on the words in the dictionary.

In addition, Hengsanankun [29] work introducing the string matching and the fast updating algorithms was proposed to cope with the unknown word problem. The algorithm also considered the speed of adding new words to the dictionary and the memory consumption. The experimental results show that their method improved the accuracy of word segmentation compared to the traditional algorithms. Moreover, the speed of the positioning to add a word and memory consumption of the dictionary are significantly improved. The algorithm of Word Corpus Building was next proposed to add unknown words into the corpus. This paper applied the Longest Matching algorithm to segmenting the word. The Stop-word List Removal algorithm was next implemented using the Thai Stop-word corpus from PythaiNLP to remove all insignificant words in sentence [30].

For example, the result of sentence “จะต้องย้ายทะเบียนบ้านมาเลขที่ มหาวิทยาลัยหรือไม่ (Is it necessary to change the resident registration to the university’s post address?)” can be tabulated in Table II to verify the corpus by adding the new words.

2) *Stop-word list removal*: In terms of the meaningless of insignificant words, the Stop-word List Removal method removed all insignificant words (e.g. preposition, conjunction) in the sentence [31] that aims to reduce the processing time in which the meaning of that sentence does not change.

3) *Bag of Words (BOW)* is one of the most famous document representation methods for NLP [32]. This method operated on word analysis which aims to word categorization. In this regard, a document is represented as a bag of the terms appearing in it, and another term is assumed to be an independent term.

4) *Term Frequency-Inverse Document Frequency (TF-IDF)* is the term weighting of the word [33]. It is weighting of the important word which is used as behalf of the sentence. The method considered of word frequency appeared in the documents; if the words frequently appeared in many documents meaning that those words were unable to be the document’s behalf. In general, those words are also called the stop word, e.g. ‘a’, ‘and’, and ‘the’.

5) *One-hot Encoding* was proposed by [34]. The method determines each word in the sentence to binary; only the interesting word is encoded to ‘1’ while other unmatched words are ‘0’, thus the results of this method consist only of ‘0’ or ‘1’ values. For example, sentence “I am a teacher” is encoded as [0, 0, 0, 1] meaning that the interesting word is only ‘teacher’.

TABLE I. THE EXAMPLE OF THE TRAINING DATASET

Q_ID	IntentID	Sentence Patterns
1	1	ขอสอบถามว่าค่าเทอมนี้เราสามารถแบ่งจ่ายหรือจ่ายช้ากว่ากำหนดได้รึเปล่านั้น (We would like to ask whether the tuition fee can be divided or paid later than scheduled.)
2	1	สามารถแบ่งจ่ายค่าเทอมหรือจ่ายทีหลังได้ไหมครับ (Can we do tuition fees payment by installment or pay later?)
3	1	ขอสอบถามว่าเราแบ่งจ่ายหรือจ่ายค่าเทอมช้ากว่ากำหนดได้ไหม (May I ask if we can split or pay the tuition fees later than scheduled?)
4	1	ค่าเทอมสามารถแบ่งจ่ายหรือจ่ายช้ากว่ากำหนดได้ไหม (Can tuition fees be split for payment or late?)
5	1	ถ้าจะขอทำการผ่อนผันค่าเทอมกับทางมหาวิทยาลัยจะได้ไหมคะ (Would it be possible to request for a tuition fee waiver with the university?)

TABLE II. THE EXAMPLE OF THE TRAINING DATASET

Processing Step	Results
Example sentence	จะต้องย้ายทะเบียนบ้านมาเลขที่ มหาวิทยาลัยหรือไม่ (Is it necessary to change the resident registration to the university’s post address?)
Word segmentation	จะ   ต้อง   ย้าย   ทะเบียนบ้าน   มา   เลขที่   มหาวิทยาลัย   หรือไม่ (Is   it   necessary   to   change   the   resident registration   to   the   university   post address   ?)
Stop-word List removal	ย้าย   ทะเบียนบ้าน   เลขที่   มหาวิทยาลัย (change   resident registration   university   post address)



The parameters referred to the intent identification framework in this study were represented the network dimensions. Once the weighted input variables to minimize the bias has been activated to obtain the classified intent. The dimensional reduction in hidden layers of the convolutional neural network was performed to yield the best-fit parameters (dimensions) using the simple rectified linear unit (ReLU) activation function. Also, Severyn and Moschitti [35] used the CNN network to model ideal sentence representations of questions and answers. The additional features have been proposed to embed the relational information form that is given by matching words between the question and answer pair. The network tuned these parameters to produce comparable results to state-of-the-art approaches. The conversational agent picked a parameter-based action which involves predicting the next sentence of a sequence at each time step.

While the methods of the convolutional learning show promising results, the natural language post-processing such as the term frequency-inverse document frequency (TF-IDF) is still required to sharply define the pilot matching question and answer (described in the next sub-section). <<the response on the influence parameters>>

#### D. Automated Conversational Agent Application using LINE Messaging API

LINE Messaging API was integrated with the Chatbot model to implement the automated conversational application via the LINE application. Accordingly, a question asking by a user sent via LINE could be preprocessed by natural language processing and used as the input variables to the model to identify the intent. LINE Message Management Will rely on an intermediary working to send messages from the LINE application, called the Web Hook. Webhook is the developer interface of the LINE application that would be functioned to send and receive messages through LINE bots. It operates to process the instruction sets on the server side. Fig. 4 shows the automated conversational system integrated with the LINE messaging API.

The answer identification algorithm was designed to provide the respond through the automated conversational system. The algorithm is shown in the pseudocode in Fig. 5.

To obtain the corresponding answers, the sub-intent from the intent could be considered when users continued entering the next question to the Chatbot. The previous conversation was recorded in the context memory to specify the corresponding sub-intent. The term frequency-inverse document frequency (TF-IDF) value was calculated to identify how the sentence had the most common words. If there are not matching words found, the corresponding answers would be obtained by the conversational agent model. This process will be described next.

The input question was proceeded based on the natural language techniques in Section 3(B) to generate a binary code. The binary code was brought into the process of identifying the intent by the conversational agent model. The corresponding answer would be verified using the TF-IDF of the following sub-intents. The most correlated pattern in the sub-intent would produce the maximum weight. However, the

corresponding answer might be selected from the main intent when the sub-intent mismatched. For example, all preprocessed words were turned to the term frequency value where they were found in the sub-intents {D1, D2, D3, D4, D5}. Tables IV and V show the process of sub-intent matching and obtaining the corresponding answer.

where the input question is “ต้องย้ายทะเบียนบ้าน ทำไม (I need to move the resident registration, why?)”. The alternative questions or sub-intents are shown below as:

D1: ทำไมต้องย้ายทะเบียนบ้าน

(Why do have to transfer the resident registration?)

D2: ไม่ย้ายทะเบียนบ้านได้หรือไม่

(Can we do not transfer the resident registration?)

D3: ต้องย้ายทะเบียนบ้านด้วยสาเหตุใด

(What is the reason why do we have to transfer the resident registration?)

D4: เหตุใดจึงต้องย้ายทะเบียนบ้าน

(Why do have to move to a resident registration?)

D5: จะต้องย้ายทะเบียนบ้านมาเลขที่มหาวิทยาลัยหรือไม่

(Is it necessary to change the resident registration to the university’s post address?).

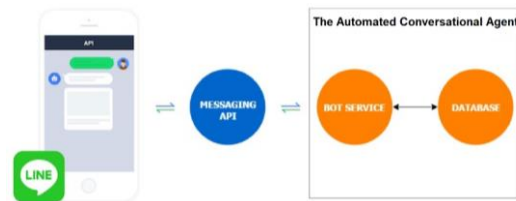


Fig. 4. The Parameters of the Intent Identification Model.

#### Input:

*IQ*: An input question

*CA*: A collection of the corresponding answers

*CM*: A collection of the context memory

*CT*: A collection of the intent

*BcV*: the collection of binary code vector {*b<sub>cv1</sub>*, *b<sub>cv2</sub>*, ..., *b<sub>cv<sub>|bcv|</sub></sub>*}

**Output:** *OCA*: a corresponding answer

#### Corresponding answer identification Algorithm:

*BcV* = NL\_Preprocessing(*IQ*)

#Determine the correlated intent from the previous question from the context memory

*CorrPreInt* = Find\_CorrelatedPreviousIntent(*BcV*, *CM*)

if *CorrPreInt* is true

*OCA* = Find\_CorrespondAnswer(*BcV*, *CA*)

else

*CT* = Find\_CorrelatedIntent(*BcV*)

    if *CT* is in sub\_intent

*CM* = Add\_Intent(*CM*, *CT*) #Replace the intent into the current context memory

    else

*OCA* = Find\_CorrespondingAnswer(*CT*, *CA*)

    endif

endif

return *OCA*

Fig. 5. The Answer Identification Algorithm.

TABLE IV. TERM FREQUENCY FOUND IN SUB-INTENTS

	ต้อง (need)	ย้าย (move)	ทะเบียนบ้าน (resident registration)	ทำไม (why)
D1	1 (0.25)	1 (0.25)	1 (0.25)	1 (0.25)
D2		1 (0.25)	1 (0.25)	
D3	1 (0.25)	1 (0.25)	1 (0.25)	
D4	1 (0.25)	1 (0.25)	1 (0.25)	
D5	1 (0.25)	1 (0.25)	1 (0.25)	

TABLE V. THE EXAMPLE OF MULTI-LAYERS CONVERSATION USING CONTEXT MEMORY

Context Memory	Multi-layer Conversation
#1 Question	จะย้ายกลับได้เมื่อไหร่ครับ (When will I be able to move back?)
#1 Corresponding Answer	ขอภัยไม่เข้าใจคำถามค่ะ (Sorry, I do not understand the question.)
#2 Question	ทำไมต้องย้ายทะเบียนบ้านครับ (Why do I need to move the resident registration?)
#2 Corresponding Answer	เพื่อให้เป็นไปตามพระราชบัญญัติทะเบียนราษฎร พ.ศ. 2535 ค่ะ (To be agreed with the Thailand Civil Registration Act in 1992)
Context Memory	ย้ายทะเบียนบ้าน (transfer the resident registration)
#3 Question	จะย้ายกลับได้เมื่อไหร่ครับ (When do I can reinstall the resident registration back?)
#3 Corresponding Answer	หลังจากที่นิสิตสำเร็จการศึกษาแล้ว หรือนิสิตศึกษาอยู่ชั้นปีที่ 4 ภาคการศึกษาปีที่ 2 ค่ะ (Later, after students have graduated or students have been studying in the 2 <sup>nd</sup> semester of the 4 <sup>th</sup> year)

The decimals inside a pair of parentheses were indicated the term frequency (IF) of the found words in the document over the total found words. For example, the word "need" in the D1 document, whose frequency is 1 divided by the total number of words is 4, will be 1/4 equal to 0.25. Next, the values of inverse document frequency (IDF) were produced. The IDF values are calculated according to (2),

$$IDF = \log(t_d/d_w) \quad (2)$$

where  $t_d$  is the total number of documents and  $d_w$  is the number of documents appears in found words. The IDF values of the word "need", "move", "resident registration", and "why" are 0.097, 0, 0, and 0.699, respectively. Besides, the term frequency - inverse document frequency (TF-IDF) values of each word found in the documents were yield by TF and IDF multiplying and shown in Table VI.

TABLE VI. TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY

	ต้อง (need)	ย้าย (move)	ทะเบียนบ้าน (resident registration)	ทำไม (why)	Total TF-IDF
D1	0.02425	0	0	0.17475	0.19900
D2		0	0		0
D3	0.02425	0	0		0.02425
D4	0.02425	0	0		0.02425
D5	0.02425	0	0		0.02425

Consequently, the selected document having the maximum of TF-IDF summation was determined as the output result which is the corresponding intent. According to the Table VI, the maximum value is 0.199, which is approximate 79% from the average value of 0.05435. Therefore, the answer of document, D1 is chosen as the answer.

### E. The FAQs Management System for Automated Conversational Agent Application

This subsection gives the administration management system for managing the frequent asked questions obtained from users using the Chatbot via LINE. The system consists of three main entities which are intent, question patterns of each intent, and information categories. Fig. 6 demonstrates the entity relationship. The FAQs information collected from users would be used for further updating precisely of the conversational agent ML model. Fig. 7 manifests the application user interface of the FAQs management system.

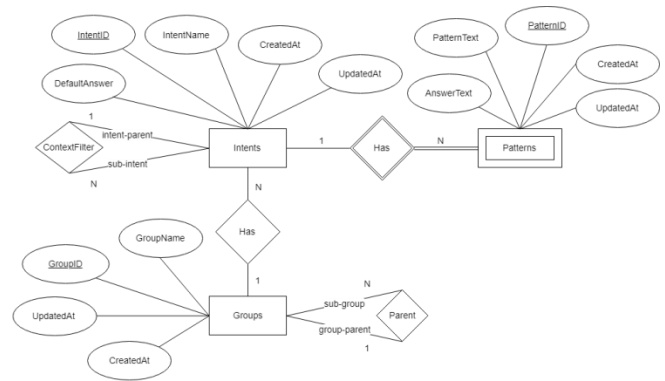


Fig. 6. The Entity Relationship of the FAQs Management System.

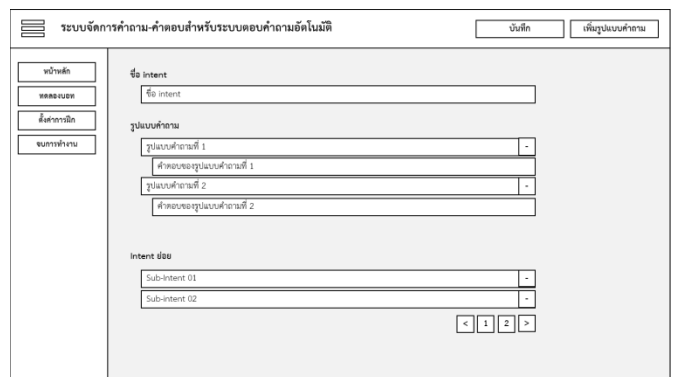


Fig. 7. User Interface for the FAQs Management System.

## IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section provides the experimental results including the discussions of the study. Besides, the experimental setting as well as the evaluation schemes used in the experiments are also described. The study findings on the conversational agent ML model, the automated conversational agent application via LINE, and the evaluation by experts are mentioned as follows.

### A. Experimental Setup and Evaluation Schemes

The collections of the frequent asked questions in Thai language for 75 questions from two main categories were used in the experiments. The data collection was collected from the

FAQs website and the students in Kasetsart University Chalermphrakiat Sakonnakhon Province Campus (KU.CSC). Some questions were collected from the FAQs regarding the services related to students within the Faculty of Science and Engineering. Each question was created with five different sentence patterns and yield overall 375 questions in total. The questions are related to enrollment, academic report, graduation, examination, tuition fees, services from university libraries, dormitory information, scholarships, and general information. Each question was revised by the faculty staffs who work under the registration division. The FAQs dataset has been processed through Thai natural language processing and binary encryption. The output binary code table was introduced to the conversational agent learning process. Fig. 8 shows the example of the model input data.

The binary code carried out from the two main NLP steps, which are 1) word segmentation and stop word removal and 2) bag of words collection and one-hot encoding was delivered to the data modeling process. To create the machine learning model, the Tensorflow development kit was implemented and described next.

The conversational agent ML model has been relied on the machine learning development kit developed by Google, the Tensorflow and the Keras library package. The ML development kit has been developed for modeling with an artificial neural network algorithm. Three main learning layers were defined to determine the relevant parameters. The input layer parameters were determined by the size of the neural network equal to the size of the dictionary. The activation function used to reduce the overfitting in the input and hidden layers is 'ReLU' activation function. The output layer was activated by 'Softmax' function by the same size of the neural network in the input layer. After configuring the parameters, the data obtained from binary coding were used to train the model using the 'model.fit' function. Table VII shows the relevant set of parameters for data modeling parameters.

The model validation was investigated the model accuracy by considering the learning rate (epochs) and the shuffle pattern before training. The learning rate (epochs) was defined by seven methods of the learning cycles which are 10, 50, 100, 150, 200, 250, 300 times. The patterns of the training dataset were shuffled to get the best-fit model results.

There are two evaluation schemes used for this study. First is the evaluation scheme for the conversational agent model. Three experts provided the ground truth for evaluation which are the corresponding answer to the intent. The performance measurement of the KUSE Chatbot was evaluated using three standard schemes: precision, recall, and F-measure. The first one reveals the correctness while the second one reflects accuracy of the obtained solution. The third one yields the efficiency of the word segmentation. These three measurements can be calculated as in,

$$F = \frac{2 \times P \times R}{P + R} \quad (3)$$

where P (precision) is the accuracy of the answer to the question compared with the accuracy of the answers from

experts. R (recall) is the correctness of the answer to the question compared to the correctness of the answer from the database.

### B. Model Selection and the Testing Performance of the Automated Conversational Agent

The model testing has turned the highest efficiency value of 0.80. One (t1\_250) having the highest accuracy out of 9 models which is 72 percent of accuracy has been considered.

From the data modeling implementation using the tools Tensorflow and Keras, the most effective model is training with WordNet, model t1\_250 and shuffling training data in descending. Fig. 9 demonstrates the learning errors and performance graphs during training process.

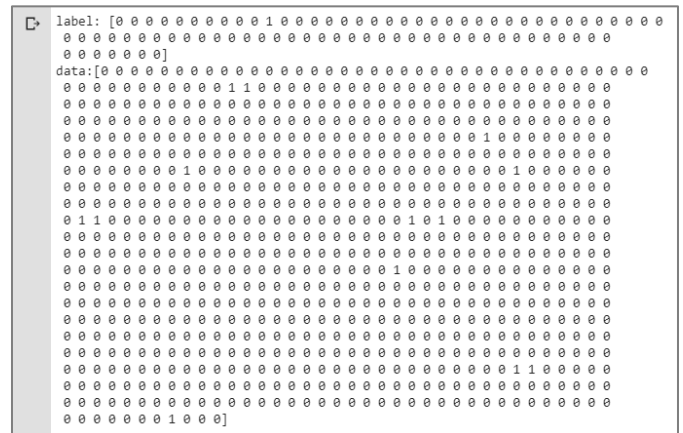


Fig. 8. The Example of Binary Code Table as the Model Input Data.

TABLE VII. THE PARAMETER SETTING FOR THE DATA MODELING

1	model = tf.keras.models.Sequential()
2	model.add(tf.keras.layers.Dense( <b>dataset size</b> , input_shape=( <b>dataset size</b> )))
3	model.add(tf.keras.layers.Activation('relu'))
4	model.add(tf.keras.layers.Dropout(0.1))
5	model.add(tf.keras.layers.Dense(32, activation='relu'))
6	model.add(tf.keras.layers.Dense( <b>size of intents</b> ))
7	model.add(tf.keras.layers.Activation('softmax'))
8	model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['categorical_accuracy'])
9	model.fit(train_data, train_labels, epochs=250, batch_size=10, verbose=1, validation_split=0.2)

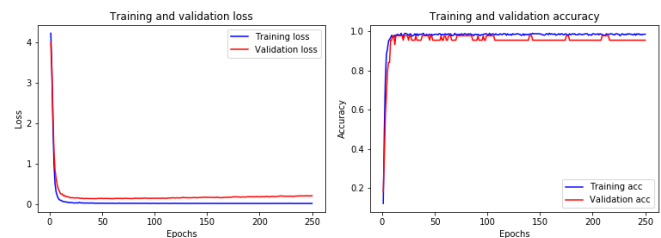


Fig. 9. The Learning Tolerances and Efficiency Graphs during Training of the t1\_250 Model with Shuffle Data Pattern in Descending.



Meanwhile, the performance measurement was performed by asking five users devoting. Five questions related to the intents per each in total of 25 questions were used to test the KUSE-Chatbot performance. The F-Measure is 0.76 when using the automated FAQs and the system could be able to answer correctly 19 questions.

### C. The Automated Conversational Agent with the Line Messaging API

The automated conversational agent model has been implemented with the LINE messaging API to provide the system framework for users. The system could operate 24-7 hours via LINE. In order to connect to the Line Messaging API, the system has to automatically have internet access and have a domain host registered. The automated conversational agent integrated with the Line Messaging API is as shown in Fig. 10 by the example of having conversations with the KUSE-Chatbot.

### D. The FAQs Management System for Administrators

Administrators can manage question-answer information used for further system model training updated. The question-answer management system was developed using web application technology. It was designed to serve the operations as follows: user login, intent data management, question-answer information management, importing questions and answers for system model training, and user privacy. Fig. 11 shows the operation of the web-based management system.

### E. Discussion

This section brings about the discussion on the study findings. Getting started by the dialogue dataset, the correlation between the vary of the input sentences affect the accuracy. The study provided different five patterns in each sentence from different students, so the sentence patterns were diversity. The parameters referred to the intent identification framework were prominent as performing the best-fit parameters (dimensions) using just the simple rectified linear unit (ReLU) activation function. The conversational agent based neural network model enabled a parameter-based optimization for the intent classification.

While the methods of convolutional learning show promising results in [35], they require careful handling of the action and state space, which can restrict the models' expressive capacity and learning capability. Keep in mind that standalone CNN based models aim for their expressive ability and natural language modelling capability as same as our based approach. The natural language post-processing was also required to yield the pilot matching question and answer to make strongly identification.

However, the natural language processing tools for the Thai language found some limitations. For example, the search for part of speech found that it was not possible to correctly identify every word. Finding synonyms using WordNet, misspellings can still search for synonyms. The dictionary is quite large; therefore, it takes quite a long time to train the model. The question sentence patterns are too close due to the interobserver variability from the same class students. It is necessary to use sentence patterns for a variety of key language questions.

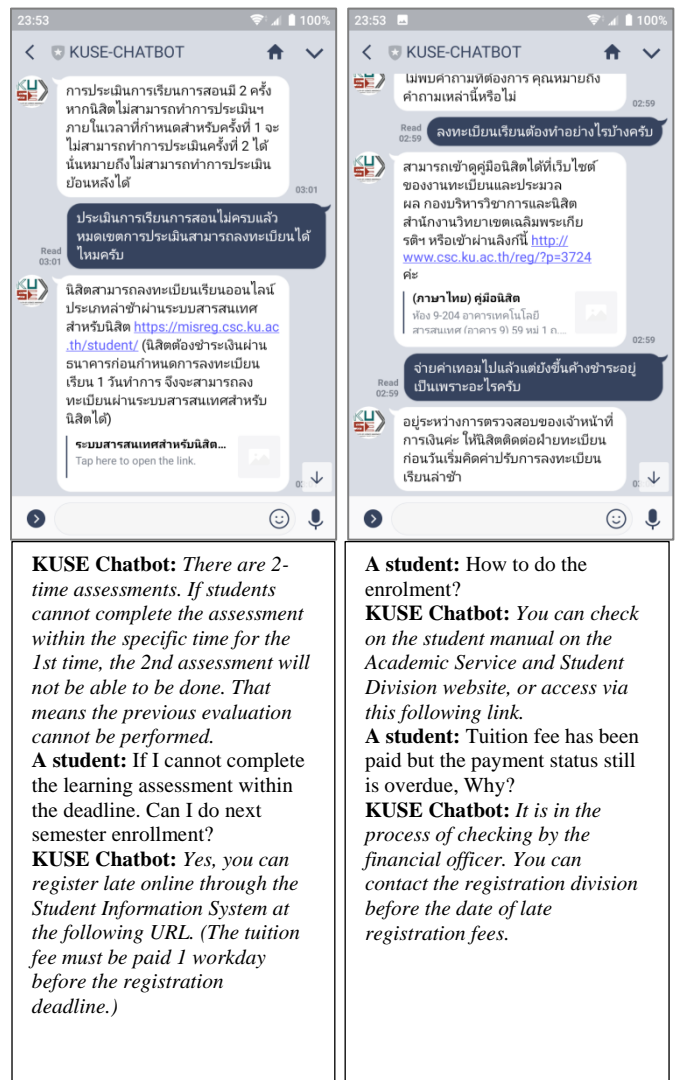


Fig. 10. The Example of having Conversations with the KUSE-Chatbot.

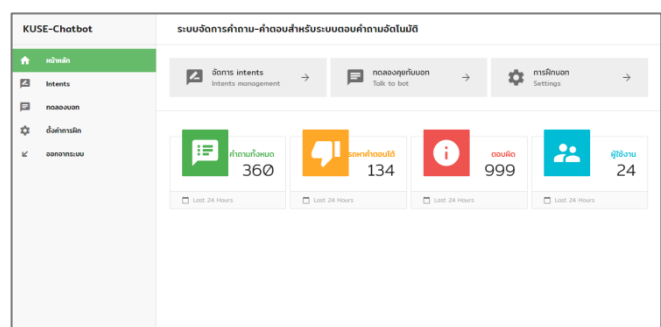


Fig. 11. The Main Page of the FAQs Management System after Signing-in.

## V. CONCLUSION

The summarization of the research findings is provided in this section. The additional perspectives relating to this article is also discussed as follows.

The automated conversational agent by using the neural network model and one-hot encoding could summarize the

results of operations and suggestions to whom interested in with the following content.

The study has been involved with 3 components, the conversational agent ML model, the automated conversational agent integrated with the Line Messaging API, and the FAQs management system for managing questions and answers for system administrators. To create the Chatbot model, the collected question-answer datasets would be used to train the machine learning model to be able to answer user-entered questions through the LINE application. Users or students can inquire about basic information related to them. The system was developed using the Python development kit in conjunction with the machine learning development kit by Google, Tensorflow, which has an efficient system that can provide automated advice to users by the accuracy of 76 percent. The faculty of Science and Engineering from Kasetsart University Chalermphrakiat Sakonnakhon Province Campus has been performed to apply the KUSE-Chatbot.

## VI. FUTURE WORK

There are still parts that need to be improved. The guidelines for further development are as follows.

- Asking for Thai language experts to help examine the question sentence format.
- Checking misspelled words or allowing correcting incorrect words in order to make the correct filtering of synonyms derived from WordNet more accurate.
- Developing to be able to learn from articles or documents instead of using question formats.
- Developing user interface in addition to applications via LINE, such as through the Faculty of Science and Engineering website or could be used via Facebook services.

## ACKNOWLEDGMENT

The authors would highly acknowledge to the Faculty of Science and Engineering, Kasetsart University Chalermphrakiat Sakonnakhon Province Campus, Sakon Nakhon, Thailand for funding support including facilities and data collections.

## REFERENCES

- [1] T. Lalwani, S. Bhalotia, A. Pal, S. Bisen, and V. Rathod, "Implementation of a Chat Bot System using AI and NLP," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 6, no. 3, pp. 26–30, 2018, doi: 10.21276/ijrest.2018.6.3.2.
- [2] J. Thakkar, P. Raut, Y. Doshi, and K. Parekh, "Erasmus – AI Chatbot," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 498–502, Oct. 2018.
- [3] A. Hommersom et al., "MoSHCA - my mobile and smart health care assistant," in 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013), Oct. 2013, pp. 188–192, doi: 10.1109/HealthCom.2013.6720664.
- [4] S. Chandel, Y. Yuying, G. Yujie, A. Razaque, and G. Yang, "Chatbot: Efficient and Utility-Based Platform," in *Intelligent Computing*, Cham, 2019, pp. 109–122, doi: 10.1007/978-3-030-01174-1\_9.
- [5] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," *Comput. Educ.*, vol. 151, p. 103862, Jul. 2020, doi: 10.1016/j.compedu.2020.103862.
- [6] K. Ramesh, S. Ravishankaran, A. Joshi, and K. Chandrasekaran, "A Survey of Design Techniques for Conversational Agents," in *Information, Communication and Computing Technology*, Singapore, 2017, pp. 336–350, doi: 10.1007/978-981-10-6544-6\_31.
- [7] S. Hussain, O. Ameri Sianaki, and N. Ababneh, "A Survey on Conversational Agents/Chatbots Classification and Design Techniques," in *Web, Artificial Intelligence and Network Applications*, Cham, 2019, pp. 946–956, doi: 10.1007/978-3-030-15035-8\_93.
- [8] R. Wallace, "Chatbot A.L.I.C.E., A.L.I.C.E. AI Foundation | Virtual Assistant A.L.I.C.E. | Virtual agent A.L.I.C.E. | Chat bot A.L.I.C.E. | Conversational agent A.L.I.C.E. | (4510)," [chatbots.org. https://www.chatbots.org/chatbot/a.l.i.c.e/](https://www.chatbots.org/chatbot/a.l.i.c.e/) (accessed Apr. 17, 2020).
- [9] R. Tantipantarak, "Human Tutor Imitation in Tutorial Dialogue System by Teaching-Log," Thammasat University, Thailand, 2010.
- [10] C. Tangkathach, "Generating Conversation Dialogues for Chatbots using Website Contents," Thesis, Chulalongkorn University, 2010.
- [11] A. Nazir, M. Y. Khan, T. Ahmed, S. I. Jami, and S. Wasi, "A Novel Approach for Ontology-Driven Information Retrieving Chatbot for Fashion Brands," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 10, no. 9, 2019, doi: 10.14569/IJACSA.2019.0100972.
- [12] J. Li, M. Galley, C. Brockett, G. P. Spithourakis, J. Gao, and B. Dolan, "A Persona-Based Neural Conversation Model," *ArXiv160306155 Cs*, Jun. 2016, Accessed: Apr. 19, 2020. [Online]. Available: <http://arxiv.org/abs/1603.06155>.
- [13] K.-J. Oh, D. Lee, B. Ko, and H.-J. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," in 2017 18th IEEE International Conference on Mobile Data Management (MDM), May 2017, pp. 371–375, doi: 10.1109/MDM.2017.64.
- [14] L. Laranjo et al., "Conversational agents in healthcare: a systematic review," *J. Am. Med. Inform. Assoc.*, vol. 25, no. 9, pp. 1248–1258, Sep. 2018, doi: 10.1093/jamia/ocy072.
- [15] R. López-Cózar, Z. Callejas, G. Espejo, and D. Griol, "Enhancement of Conversational Agents By Means of Multimodal Interaction," 2011, doi: 10.4018/978-1-60960-617-6.ch010.
- [16] "Spoken dialogue technology: enabling the conversational user interface: ACM Computing Surveys: Vol 34, No 1." <https://dl.acm.org/doi/10.1145/505282.505285> (accessed Apr. 19, 2020).
- [17] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial," *JMIR Ment. Health*, vol. 4, no. 2, p. e19, Jun. 2017, doi: 10.2196/mental.7785.
- [18] N. M. Radziwill and M. C. Benton, "Evaluating Quality of Chatbots and Intelligent Conversational Agents," *ArXiv170404579 Cs*, Apr. 2017, Accessed: Apr. 19, 2020. [Online]. Available: <http://arxiv.org/abs/1704.04579>.
- [19] B. Juang and S. Furui, "Automatic recognition and understanding of spoken language - a first step toward natural human-machine communication," *Proc. IEEE*, 2000, doi: 10.1109/5.880077.
- [20] S. Young, M. Gašić, B. Thomson, and J. D. Williams, "POMDP-Based Statistical Spoken Dialog Systems: A Review," *Proc. IEEE*, vol. 101, no. 5, pp. 1160–1179, May 2013, doi: 10.1109/JPROC.2012.2225812.
- [21] S. Arsovski, S. H. Wong, and A. D. Cheok, "Open-Domain Neural Conversational Agents: The Step Towards Artificial General Intelligence," *Int. J. Adv. Comput. Sci. Appl. Ijacsca*, vol. 9, no. 6, 2018, doi: 10.14569/IJACSA.2018.090654.
- [22] G. Mesnil et al., "Using recurrent neural networks for slot filling in spoken language understanding." IEEE Press, Mar. 01, 2015.
- [23] M. Hijjawi, Z. Bandar, and K. Crockett, "A General Evaluation Framework for Text Based Conversational Agent," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, 2016, Accessed: Apr. 12, 2020. [Online]. Available: <https://e-space.mmu.ac.uk/605759/>.
- [24] S. Subramaniam, P. Aggarwal, G. B. Dasgupta, and A. Paradkar, "COBOTS - A Cognitive Multi-Bot Conversational Framework for Technical Support," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, Stockholm, Sweden, Jul. 2018, pp. 597–604, Accessed: Apr. 12, 2020. [Online].

- [25] E. Latorre-Navarro and J. Harris, "An Intelligent Natural Language Conversational System for Academic Advising," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 1, pp. 110–119, 2015.
- [26] N. Chaicharoen, "Computerized Integrated Word Segmentation and Part-of-Speech Tagging of Thai," Thesis, Chulalongkorn University, 2001.
- [27] Y. Poowarawan and V. Imarom, "Dictionary-based Thai Syllable Separation," presented at the 9th Electrical Engineering Conference, 1986, Accessed: Apr. 10, 2020. [Online]. Available: <https://www.semanticscholar.org/paper/Dictionary-based-Thai-Syllable-Separation-Poowarawan/d15dca32d2bd685c12fb9a5cdb5bbdf6ccc366b2>.
- [28] V. Sornlertlamvanich, "Word segmentation for Thai in machine translation system," *Mach. Transl. NECTEC*, pp. 50–55, 1993.
- [29] T. Hengsanankun and P. Seresangtakul, "Thai Text Segmentation using Fast and Compact Updating Algorithm," presented at the The 2nd Joint Conference on Computer Science and Software Engineering, Nov. 2005, pp. 144–150.
- [30] PyThaiNLP, "Thai natural Language processing in Python," 2018. <http://github.com/PyThaiNLP/pythainlp> (accessed Aug. 15, 2018).
- [31] N. Chirawichitchai, P. Sanguansat, and P. Meesad, "A comparative Study on Term Weight Techniques for Thai Document Categorization," *J. Sci. Ladkrabang*, vol. 19, no. 1, pp. 26–40, 2010.
- [32] W. Pu, N. Liu, S. Yan, J. Yan, K. Xie, and Z. Chen, "Local Word Bag Model for Text Categorization," *Seventh IEEE Int. Conf. Data Min. ICDM 2007*, pp. 625–630, 2007, doi: 10.1109/ICDM.2007.69.
- [33] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.
- [34] L. Buitinck et al., "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, Sep. 2013, pp. 108–122, Accessed: Apr. 10, 2020. [Online]. Available: <http://arxiv.org/abs/1309.0238>.
- [35] A. Severyn and A. Moschitti, "Modeling Relational Information in Question-Answer Pairs with Convolutional Neural Networks," *ArXiv160401178 Cs*, Apr. 2016, Accessed: Apr. 20, 2020. [Online]. Available: <http://arxiv.org/abs/1604.01178>.