

Development of computer-based learning system for learning behavior analytics

Kanyalag Phodong¹, Thepchai Supnithi², Rachada Kongkachandra³

¹Department of Computer Science, Faculty of Science and Technology, Thammasat University, Pathumthani, Thailand

²Language and Semantic Technology Laboratory, Intelligent Informatics Research Unit, National Electronic and Computer Technology, Pathumthani, Thailand

³Data Science and Innovation Program, College of Interdisciplinary Studies, Thammasat University, Pathumthani, Thailand

Article Info

Article history:

Received Apr 5, 2021

Revised Nov 23, 2021

Accepted Nov 28, 2021

Keywords:

Computer-based learning

Learning analytics

Natural language processing

Self-regulated learning

ABSTRACT

This paper aims to analyze the learning behavior of Thai learners by using a computer-based learning system for English writing. Three main objectives were set: the development of a computer-based learning system, automatic behavior data collection, and learning behavior analytics. Firstly, the system is developed under a multidisciplinary idea that is designed to integrate two concepts between the self-regulated learning model and components of natural language processing. The integration design encourages self-learning in the digital learning environment and supports appropriate English writing by the provided component selection. Second, the system automatically collects the writing behavior of a group of Thai learners. The data collected are necessary input for the process of learning analytics. Third, the writing behaviors data were analyzed to find the learning behavioral patterns of the learners. For learning analytics, behavior sequential analysis was used to analyze the learning logs from the system. The 31 undergraduate students are participated to record writing behaviors via the system. The learning patterns in relation to grammatical skills were compared between three groups: basic, intermediate, and advanced levels. The learning behavior patterns of the three groups are different that use for reflecting learners and improving the learning materials or curriculum.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rachada Kongkachandra

Data Science and Innovation Program, College of Interdisciplinary Studies, Thammasat University

Rangsit Center, Phahon Yothin, Klong Luang, Pathumthani 12121, Thailand

Email: krachada@staff.tu.ac.th

1. INTRODUCTION

The English language is considered as essential for Thai people and is therefore a fundamental part of the education system. Thai learners often experience difficulties in studying English as a foreign language (EFL), in reading, speaking and especially writing [1]. Most language teaching in Thailand is a one-size-fits-all that is unable to clearly identify the weaknesses of each learner. Personalized learning for the English language is one possible solution. This aims to analyze individual learning behavior in order to identify each learner's strengths and weaknesses.

Computer technology increases learning behavior analytics for personalized learning in terms of the storage and speed of analytics processing. Learning behavior analytics using computer-based technology is quicker and cheaper than human analysis. Although computer technology supports data storage and faster processing, language learning requires an underpinning pedagogy to foster self-learning for personalized learning

analytics. The self-regulated learning model is an essential model to get positive outcomes in learning, such as encouraging learners' skills to shape their own learning [2] and supporting lifelong autonomous learning [3].

There is various researches present model to encourage for learning of foreign language. The self-regulated learning model is an efficient factor to improve the learning performance [4]. The model is applied to foreign language learning. Incorporating the concepts of the self-regulated learning model to the foreign language learning that supports the development of autonomous learners [3]. The self-regulated learning model facilitates learners lead to higher efficiency in language skills such as comprehension of writing [5]. The model consists of three main phases: forethought, performance and self-reflection [6], [7]. The self-regulated learning model encourages interaction between person, behavior, and environmental factors to increase effective learning [8]. Computer technology is being extensively used in the education field [4], [9]. A computer-based learning system is a tool of computer technology that can be used for encouraging interactive behavior between personal and learning environments. A computer-based learning system could support a better learning experience that learners could engage the interactions with learning tasks [10]. In addition, the computer-based learning system supports automatic data collection for recording learning behavior while using the digital system. The system can also automatically collect learning behavior data to analyze the pattern of learning behaviors.

Natural language processing (NLP) aims to make the computer able to understand the language through computer processing. There are six levels of language processing: morphological, lexical analysis, syntactic analysis, semantic analysis, pragmatics, and discourse [11]. These processes are applied to develop many NLP tools such as word segmentation, lexical analysis and parsing [12]. Moreover, applying natural language processing is an effective tool for enhancing the education field. NLP can improve the learning ability of the student in case of student fails to understand the context due to the barrier of language. NLP and digital technology are combined to improve a computer-assisted teaching system [13]. Mathew *et al.* [14] provide the application of NLP techniques for an assistant tool to support teachers get insights about each student's learning progress. Therefore, a computer-based learning system could integrate the methods of NLP to assist Thai EFL learners in their understanding of language structure and encourage learners' improvement in English writing, in particular.

Learning analytics in digital learning environments is an integration of two research fields, which are those of education and computer technology. Learning analytics is an important issue in education. Learning analytics is the analysis of 'learning logs' and education data for improving learning outcomes, learning designs, and learning environments [15]. On the other hand, computer technologies have become popular for communications and learning. Technologies are convenience to access through portable devices such as smartphones, tablets, and laptops. Therefore, the integration of these two fields can help improving education.

This paper aims to acquire the learning behavior by using the provided computer-based learning system for composing the English sentences. The system is designed by incorporating concepts of the self-regulated learning model and components of NLP. Self-regulated learning encourages learners to set goals, as well as monitoring their behaviors and reflect writing performance to learners. The NLP learning environment encourages action between learner and system to compose the target sentences. Furthermore, all behaviors are automatically recorded for use in the learning analytics process. The results of learning analytics are useful to demonstrate learning performance and to support the improvement of learning materials. This paper is structured as follows: Section 2 explains background information and related works about the model of self-regulated learning, components of NLP, and learning analytics in foreign language learning. Section 3 describes the computer-based learning system. Section 4 describes the experimental design. Section 5 describes the experimental results. Section 6 provides a discussion. Finally, section 7 gives conclusions.

2. BACKGROUND

2.1. Self-regulated learning model

The self-regulated learning model is a conceptual framework of interaction between person, behavior and environment in a learning context and comprises three main phases: forethought, performance and self-reflection [16], as illustrated in Figure 1.

- a. Forethought phase: Learners set goals and learning plans. The learners plan how to reach them in the learning strategies activation process.
- b. Performance phase: Learners control themselves while executing the task and they monitor their progress in completing the task.

- c. Self-reflection phase: Learners evaluate their satisfaction in performing the task, making attributions for their achievement or failure. These attributions generate self-reactions that positively or negatively influence learners.

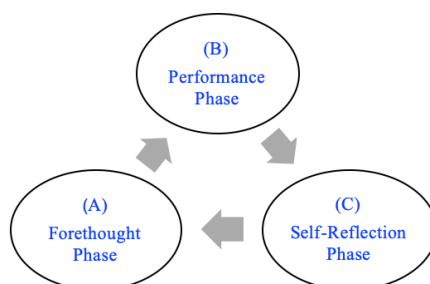


Figure 1. The three phases of self-regulated learning model

There are various works that present the benefit of using the self-regulated learning model in foreign language learning. A range of research papers have presented the benefits of using the self-regulated learning model in foreign language learning. In the English language learning context, incorporating the self-regulated learning model into the curriculum and training programs encourages autonomous, life-long learning. Abadikhah *et al.* [17] investigated EFL university learners' attitudes towards the strategies of self-regulated learning in writing academic papers. The study compared the attitudes of two groups in the application of the self-regulated learning model. It set out to establish whether academic education assists learners in becoming self-regulated writers. Assessing learners' attitudes in applying self-regulated strategies in their writing may benefit the design of academic writing courses. The learners' attitudes assessment can provide detailed and highly relevant information to help instructors enhance their learners' performance. Instructors have an important role in assisting learners to become self-regulated writers. Moreover, Karami *et al.* [4] tried to answer the questions regarding the effect of digital technology on the writing proficiency of learner and the self-regulated strategies usage in the context of English learning as a foreign language. The ability of the self-regulated strategies is correlated to a higher level of writing achievement in an environment of digital technology.

2.2. Natural language processing (NLP) resources and services

NLP aims to use the technique to make the computer system understand the natural language text or speech [18]. There are six levels of NLP tasks [11]: morphological, lexical analysis, syntactic analysis, semantic analysis, pragmatics, and discourse.

In this paper, lexical and syntactic NLP techniques were set as a learning environment to help the learners compose target sentences in English, as shown in Table 1. Moreover, previous works [19]–[21] relate to improving the NLP process with linguistic knowledge for improving word alignment of SMT. Those proposed techniques are also applied to set as learning environments such as the dictionary, Part of Speech (POS) tagging, and tenses detection.

Table 1. List of components in NLP and their grammatical aspects

Level of NLP Processes	Grammatical Aspects	Components
Lexical Level	Vocabulary	Dictionary Plurality
Syntactic Level	Sentence Structure and Tenses	POS Verb Pattern Word Alignment

2.3. Background of learning analytics

Interpreting and evaluating the qualities of activities, strategies, goals and regulation involved in self-regulated learning model is somewhat complicated. Learning behaviors data gathered in a digital learning environment are instrumental to address these challenges [22]. However, the raw data alone are insufficient to guide practice or shape theory. Therefore, learning analytics has a role to play in improving the effectiveness of learning. There are various works for applying learning analytics to the education field. Learning analytics reports data analysis that describes features or factors that influence the self-regulated model [23]. Analysis of

e-learning in factors of culture, technology or infrastructure, and content satisfaction that the analyzed results can be used to develop the proper e-learning in a remote city of Indonesia [24].

Since learning analytics are a supporting tool in the digital environment, this paper uses learning analytics to analyze learning behaviors of writing. All learner behaviors in the log file are analyzed in the learning analytics process. This paper aims to investigate learning behavior that how the provided components in the learning environment reflect the performance of English writing. In addition, learning analytics are used to find learning behavior patterns which categorize a group of the learner.

3. THE PROPOSED SYSTEM

The development of the computer-based learning system for English writing in Thai EFL learners aims for three tasks. First, the system integrated two disciplines between the pedagogical model and components of NLP. The self-regulated learning model is a pedagogical model that encourages self-learning, facilitated by the use of a computer-based learning system. data analysis The components of NLP are helping to learn and compose English sentences. Second, the system aims to collect the learning behavior in case: English writing for Thai EFL learners. The system is designed to automatically record learning behaviors while composing English sentences. Third, writing behaviors are analyzed for finding the English writing behavioral patterns of Thai EFL learners. There are three main tasks that support designing and developing processes of the system.

3.1. System process for learning analytics

This paper proposes three main processes: learning profile acquisition, learning behavior and learning analytics, as shown in Figure 2. All three main processes work coherently as starting with the process of learning profile acquisition. First, the learning profile acquisition process aims to get information on existing English writing skills. Next, the learning behavior collection is the process for recording the learning behavior into the log files store in the data log storage. Finally, the process of learning analytics analyzes data of writing behavior from the log file. The analysis result will conduct to define the behavior pattern of Thai EFL learners in the case of English writing. The patterns of learning behavior use to reflect learners or improve the learning materials or curriculum.

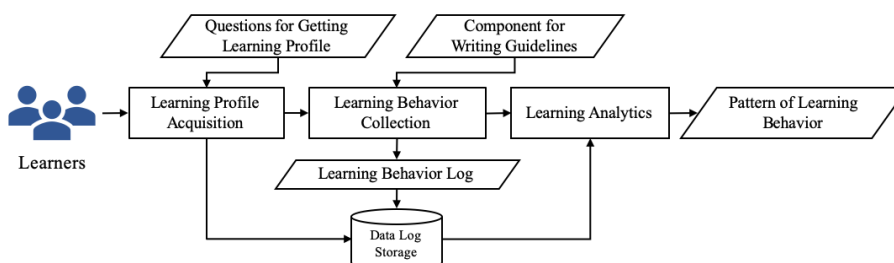


Figure 2. System overview of the computer-based learning system for learning analytics

3.1.1. Learning profile acquisition

The learning profile acquisition is an initial process that uses two steps, registration and acquisition of existing English skills, to get information from the learner. Firstly, learners provide personal and educational information on a registration form. Next, the English grammatical skill acquisition step uses to get the existing writing skills. Learners test to compose the provided sentences without assisting tool for getting a learning profile that reflects the learners existing grammatical skills in three aspects: vocabulary usage, sentence type understanding and tense usage. Then, all answers are scored [25] and categorized into one of three levels (basic, intermediate or advanced) in relation to their English grammatical skill before learners access to the process of collection the learning behavior.

3.1.2. Learning behavior collection

Learning behavior collection is needed for the learning analytics process. This process connects to the data log store for collecting learners' behavior that is important data to analyze by the process of learning analytics. Furthermore, this process is an integrated process for encouraging learning skills by applying the concepts of self-regulated learning model and components of NLP into four subprocesses: source sentence

assignment, component selection, writing behavior monitoring and answering for self-reflection, as shown in Figure 3.

The strategies of the self-regulated learning are applied to the workflow to support self-learning in the computer-based learning environment. The components of NLP are deployed to the component for writing guidelines into the system. When the collected behaviors are analyzed, the data of component selection can reflect the grammatical skills of learners.

The model of self-regulated learning encourages the interaction between person, behavior, and environmental factors for effective learning [8]. According to the definition of three main phases [6], [7], the forethought phase is a goal-setting about the learner's need to learn. The performance phase is collecting the learning behavior. Learners' actions with the provided learning environment and inform their progress. The self-reflection phase is self-assessment and behavior adaption for increasing the effective method of learning. Therefore, this process is designed according to the three main concepts of self-regulated phases for process efficiency [26], as illustrated in Figure 3.

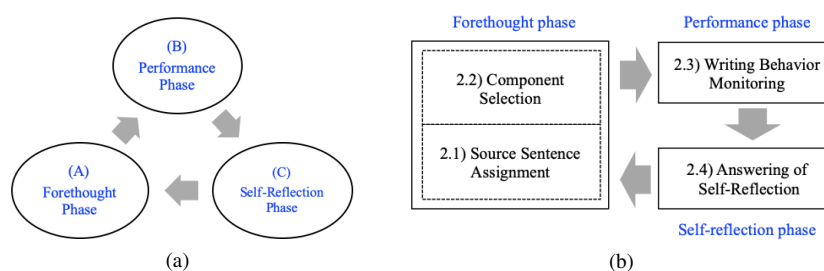


Figure 3. A relation between phases of (a) self-regulated learning model and (b) subprocesses of the system

- a. Source sentence assignment: After learners finish the learning profile acquisition process, they access learning behavior collection for recording writing behavior. The process of learning behavior collection starts with a subprocess of source sentence assignment to practice writing English sentences. The learner selects the source sentence by themselves for trying to compose the target sentence completely. Since learners' decision to select source sentences by themselves. This action relates to set the goal of the forethought phase in the self-regulated learning model as shown in Figure 3. The source sentence selection indicates the learner set the goal for composing the complete target sentence in English.
- b. Component selection: This subprocess is designed to include the components of NLP that is an integration process between the method of NLP and the educational model. The details of the components of NLP are described in section 3.2. These components of NLP are designed to help learners compose English sentences and to motivate them in their writing. The selected component by learners will demonstrate their grammatical needs through the defined components selection. When the source sentence is assigned by the learner, the provided components are used to assist for target sentence composition. All selected components and time usage are recorded in the log file. Moreover, component selection is also related to the forethought phase of the self-regulated learning model, as shown in Figure 3. Component selection by the users themselves indicates they have a plan to write the English sentence properly.
- c. Writing behavior monitoring: The subprocess is designed to allow monitoring learners to monitor their progress in sentence composition. The system records all activities that since the learners select source sentences, selects all NLP components for writing guidelines, until they submit the target sentences. When learners finish composing all target sentences, this subprocess will process the activity parameters in the log file and show results for learners' observation, namely: amount of sentences, the selected component, and time usage. Since the learners monitor or observe their writing performance results by themselves that relates to the concept of self-observation in the performance phase (shown in Figure 3).
- d. Answering for self-reflection: In the last subprocess of learning behavior collection, the learners answer a self-reflection questionnaire containing questions to do with their writing [17]. This subprocess helps learners to reflect on their writing behavior, some of which learners may be able to use in adapting their subsequent writing. The learners' reflection and behavior adaption that related to self-reaction of the self-reflection phase in the model of self-regulated learning are shown in Figure 3.

3.1.3. Learning analytics

Learning analytics are designed to analyze learning behavior from behavioral data in the log files. The analysis process aims to analyze writing behavior using the computer-based learning system. The process also analyzes the provided components in order to reflect the writing performance. This process uses a statistical analysis method [27] to find out the learning behavior pattern. The statistical method determines the writing behavior in the behavior transition form. The result of behavior patterns can support for consideration of writing proficiency. Moreover, the process supports finding the best practice of learning patterns that use the suggestions of other learners.

3.2. Learning environment of the computer-based learning system

The learning environment for writing guidelines consists of two main materials: learning materials and NLP materials as shown in Figure 4. The learning materials are composed of English writing tasks to practice for English sentence composition and instruction for introducing system usage. The NLP materials include components to help the English sentence composition. There are two reasons for setting NLP materials to support English writing tasks. Firstly, the NLP materials involve linguistic understanding through NLP processes such as lexical, syntactic and semantic levels. Second, since many Thai EFL learners think in Thai before translating their ideas into English sentences, a better understanding of the components of linguistics guidelines can help in the writing of appropriate English sentences.

The components of NLP are divided into two levels: lexical and syntactic, as shown in Table 1. The provided assisting components of the lexical level assist learners to write proper vocabulary i.e. dictionary and plurality. The components of the syntactic level guide learners to use appropriate grammar in sentence structure and tenses, including aspects such as part of speech (POS), verb pattern and word alignment. The screen example for assisting components of NLP is shown in Figure 5.

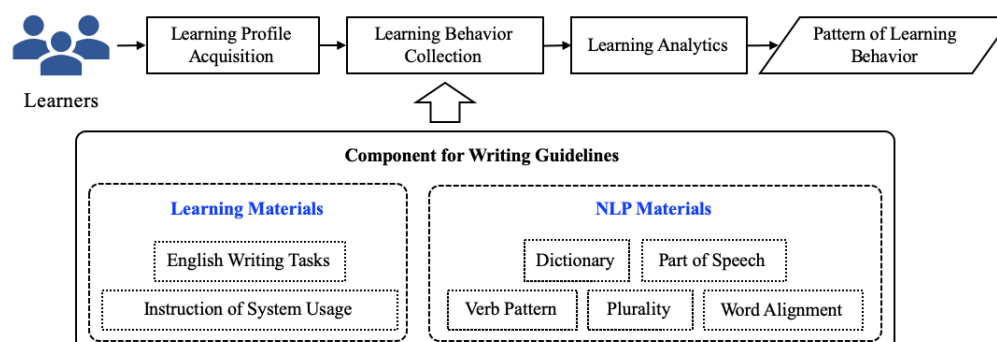


Figure 4. Learning environment of the computer-based learning system for English writing

The background of NLP used for applying to create the provided components that assist learners to compose the complete target sentence as details below:

- POS component: This component uses word segmentation and POS tagging by SWATH [24]. The POS tag set is using based on the ORCHID corpus [28].
- Dictionary component: This component uses word segmentation by LexTo+ [29]. Then, the word- segment of Thai is matched with the English word by using the API of Thai-English LEXiTRON dictionary [30].
- Verb pattern component: This component defines the POS tag by SWATH [31]. Then, the verb or auxiliary verb is identified in the tense of their word by grammatical attributes extraction [20].
- Plurality component: This component uses lemmatization to extract English plural words and transform the word using rules of plurality. Then, machine translation is used to match the plural word in English with its Thai equivalent.
- Word alignment component: This component uses word segmentation and POS tagging by SWATH [31]. Then, the word alignment uses the IBM model of GIZA word alignment [32] to align words of both languages.



Figure 5. The sample of the display for assisting components of NLP

4. EXPERIMENTAL DESIGN

The experiment was designed using behavioral data to analyze the learning behavioral patterns of Thai EFL learners. The behavioral data were collected by automatic data collection (ADC) method [33] that automatically recorded into log files while learners write the English sentence via the computer-based system. The behavioral sequential analysis method was used to explore the learning behavior pattern of Thai EFL learners in the case of English writing.

4.1. Participants

The system collects learning behavior into a log file when learners were writing the English sentence via the computer-based learning system. The learners write English tasks for a duration of about 1 hour. A total of 31 undergraduate students participated in this study. Their personal information was removed during the research processing. All writing activities were recorded in the log file for analysis by the behavioral sequential analysis method.

4.2. Coding scheme

The coding schema is required for sequential analysis method [33], [34]. However, this study uses the computer-based learning system for English writing that automatically records the learning behavior log. The learning system is implemented for getting learning behaviors. The coding process is based on learner behaviors that operating with the system. When learners use the provided learning system to practice English composition, writing behavior such as “composition”, “selection”, “insertion”, “modification” and “deletion” are recorded in the log files. Then, all data of writing behavior are used to generate the patterns of learning behavior.

- a. Composition (CP): When learners type to compose the target sentence in English, they can type in the provided textbox. While learners type each word in the sentence, all typing will be recorded in the log file.
- b. Selection (SL): When learners are interested in the components of NLP for assisting sentence composition, they can select a particular component (or components). Then, all actions of component selection will be collected in the log file. The component of NLP consists of five components: dictionary, POS, verb pattern, plurality and word alignment.
 - Dictionary selection (SL-dict): When learners click the dictionary button, this indicates their interest in the appropriate words for composing each sentence.

- POS selection (SL-POS): Sometimes, learners are confused about which part of speech a word belongs to, such as mistaking noun forms and verb forms in a sentence. When learners click the POS button, this indicates their desire to increase confidence in the part of speech of words.
 - Verb pattern (SL-verb): When learners click the verb pattern button this indicates their interest in the structure of tenses in each sentence.
 - Plurality (SL-plural): Due to differences in relation to singular and plural nouns between Thai and English, there are many different rules regarding pluralization. When learners click the plurality button, this indicates their interest in using the singular or plural nouns in each sentence.
 - Word alignment (SL-align): When learners click the word alignment button, this indicates their interest in the order of words and pairs of words that are aligned in Thai and English sentences.
- c. Insertion (IS): When learners demand to add some words or phrases into the target sentence, they can move the cursor to the desired position and type additional words or phrases into the sentence. Then, these actions will be recorded in the log file.
 - d. Modification (MD): When learners want to delete some words or partial in the target sentence, they can move the cursor to the desired position and click the backspace button to delete some words or parts of the sentence. Then, these actions will be recorded in the log file.
 - e. Deletion (DL): Learners can click the “deletion” button when they want to compose the new target sentence and delete a whole sentence. Then, the text box will be cleared. Next, learners compose the new target sentence into the same text box. These actions are recorded in the log file.

5. BEHAVIORAL LEARNING ANALYTICS

In this paper, learning analytics aims to analyze writing behaviors by using the method of behavioral sequential analysis [27] to determine behavior transitions. The analysis process used to analyze learning behavior with the assisting component in the provided learning environment reflects the behavior of English writing. The analysis of learning behavior is used to investigate all behavior for finding the learning behavior pattern.

The behavioral sequential analysis is a statistical analysis method that uses the sequential analysis matrix to calculate the behavioral transition [34]. The method uses calculation of the frequency of the behaviors sequence and the z-value to determine the behavior transition. Results greater than 1.96 indicated behavior sequences that reached statistical significance [27], [34]. The sample of the matrix of a sequential behavior series is calculated to z- value, as shown in Figure 6. Then, the z-values were greater than 1.96 were selected to generate the learning behavior transition.

Freq.	CP	SL-Dict	SL-POS	SL-Verb	SL-Plural	SL-Align	IS	MD	DL
CP	0	11	3	4	2	7	9	29	0
SL-Dict	31	0	5	0	0	2	0	0	0
SL-POS	2	0	0	5	1	1	0	0	0
SL-Verb	4	5	0	1	2	1	0	0	0
SL-Plural	1	2	1	1	0	2	0	0	0
SL-Align	4	6	0	1	1	0	1	0	0
IS	6	0	0	0	0	0	0	3	0
MD	29	0	0	0	0	1	1	0	0
DL	0	0	0	0	0	0	0	0	0



z-value	CP	SL-Dict	SL-POS	SL-Verb	SL-Plural	SL-Align	IS	MD	DL
CP	-0.39	1.49	0.12	0.29	-0.05	0.81	1.15	4.58	-0.39
SL-Dict	4.92	-0.39	0.47	-0.39	-0.39	-0.05	-0.39	-0.39	-0.39
SL-POS	-0.05	-0.39	-0.39	0.47	-0.22	-0.22	-0.39	-0.39	-0.39
SL-Verb	0.29	0.47	-0.39	-0.22	-0.05	-0.22	-0.39	-0.39	-0.39
SL-Plural	-0.22	-0.05	-0.22	-0.22	-0.39	-0.05	-0.39	-0.39	-0.39
SL-Align	0.29	0.64	-0.39	-0.22	-0.22	-0.39	-0.22	-0.39	-0.39
IS	0.64	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	0.12	-0.39
MD	4.58	-0.39	-0.39	-0.39	-0.39	-0.22	-0.22	-0.39	-0.39
DL	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39	-0.39

Figure 6. The sample of calculation for sequential behavior frequency to z-value

5.1. Analysis of individual learning behavior pattern based on existing english skills

The 31 participants were separated into three groups based on existing English skills (basic, intermediate or advanced). The writing behaviors of individual learners were analyzed using the behavioral sequential

analysis method. This method starts by defining the coding schemes from the writing behaviors which collect behaviors while learners use the provided computer-based system. The coding schemes represent the writing behavior of learners. The frequencies of sequential behavior were calculated into the matrix of a series of sequential behavior. Then, all frequencies are calculated to z-value for conducting to explore the writing behavior patterns. A z-value greater than 1.96 indicates the behavior sequences reach significance. The behavior transition of each learner used to represent the significant behavior sequences as illustrated in Figures 7 to 9.

5.1.1. The individual learning behavior pattern in the basic level

The individual behavior pattern of 14 learners in the basic level as shown in Figure 7. All individual behavior patterns of basic level were separated into five groups:

- Learning Behavior Pattern 1: “modification” has sequential correlations with “composition”
- Learning Behavior Pattern 2: “dictionary selection” has sequential correlations with “composition”
- Learning Behavior Pattern 3: “word alignment selection” has sequential correlations with “composition”
- Learning Behavior Pattern 4: “verb pattern selection” has sequential correlations with “composition”
- Learning Behavior Pattern 5: “insertion” has sequential correlations with “composition”

Analysis of the five groups of learning behavior patterns indicated that the ‘basic’ group learners used the NLP components of dictionary, word alignment and verb pattern to assist them in composing English sentences.

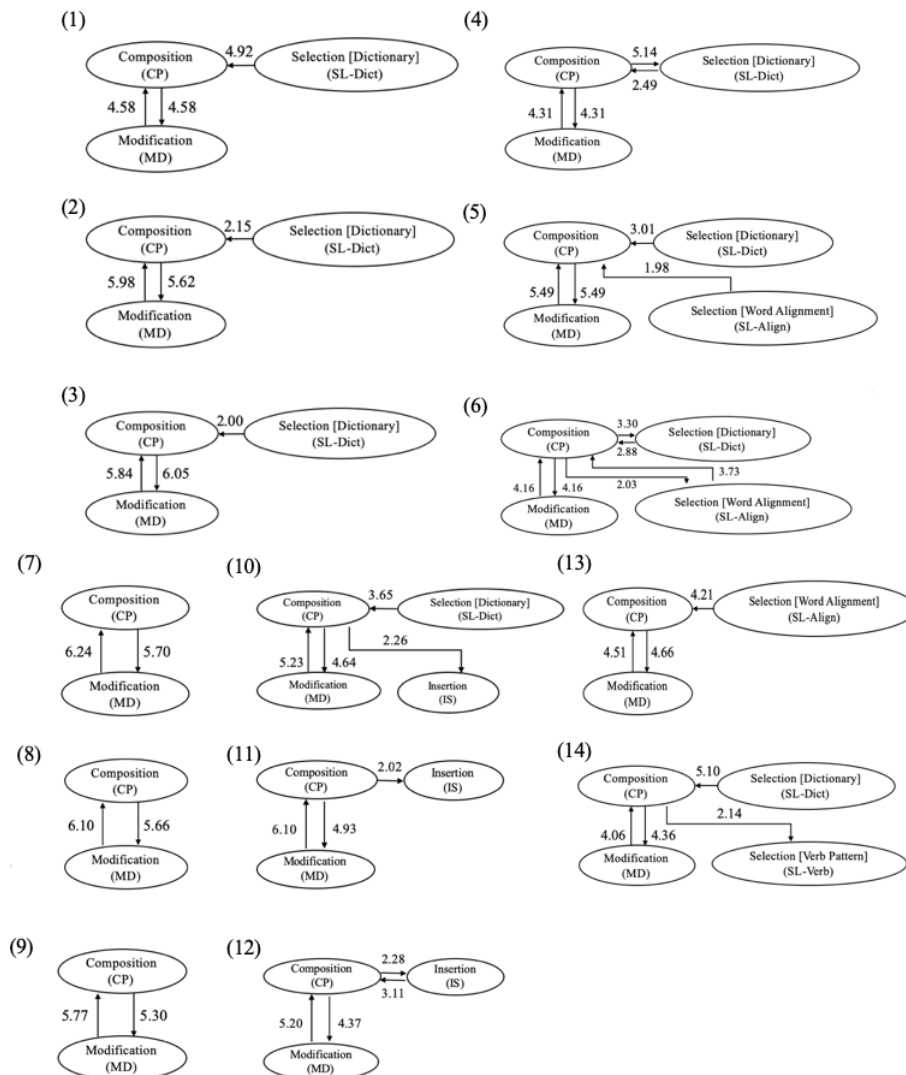


Figure 7. The individual learning behavior transition of learner in the basic level

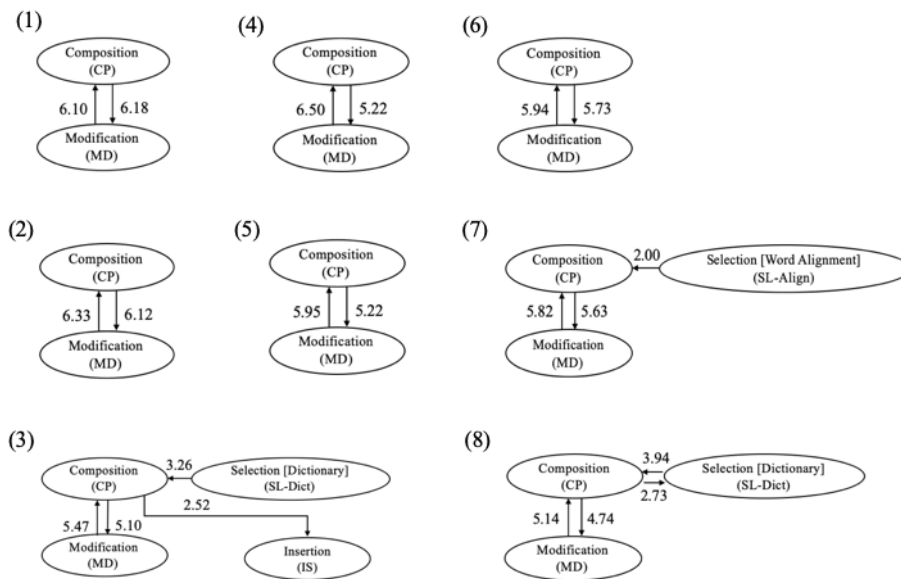


Figure 8. The individual learning behavior transition of learner in the intermediate level

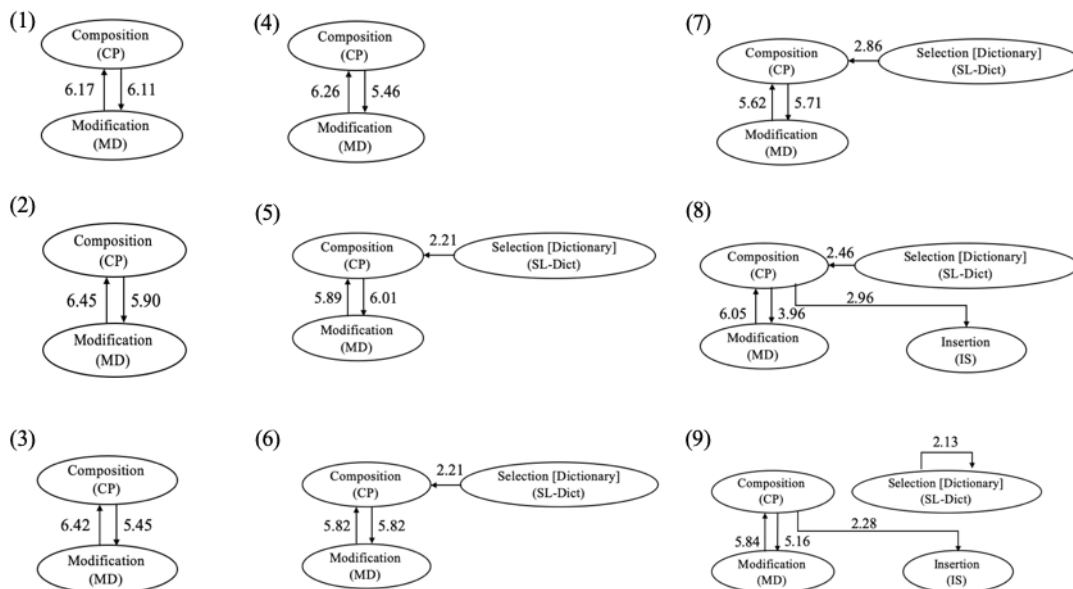


Figure 9. The individual learning behavior transition of learner in the advanced level

5.1.2. The individual learning behavior pattern in the intermediate level

The individual behavior pattern of 8 learners in the intermediate level as shown in Figure 8. All individual behavior patterns of intermediate level are separated into four groups:

- a. Learning Behavior Pattern 1: “modification” has sequential correlations with “composition”
- b. Learning Behavior Pattern 2: “dictionary selection” has sequential correlations with “composition”
- c. Learning Behavior Pattern 3: “word alignment selection” has sequential correlations with “composition”
- d. Learning Behavior Pattern 4: “insertion” has sequential correlations with “composition”

Analysis of the four groups of learning behavior patterns indicated that the ‘intermediate’ group learners used the NLP components of dictionary and word alignment for assisting to compose the English sentences.

5.1.3. The individual learning behavior pattern in the advanced level

The individual behavior pattern of nine learners in the advanced level are shown in Figure 9. All individual advanced level behavior patterns were separated into three groups:

- Learning Behavior Pattern 1: “modification” has sequential correlations with “composition”
- Learning Behavior Pattern 2: “dictionary selection” has sequential correlations with “composition”
- Learning Behavior Pattern 3: “insertion” has sequential correlations with “composition”

Analysis of the three groups of learning behavior patterns indicated that the ‘advanced’ group learners used the NLP components of only dictionary to assist them in composing English sentences .

5.2. Analysis of frequency for the provided components usage

The provided components of NLP support learners with the grammatical aspects of English writing and helped them to compose the target sentences. There are five components: dictionary, part of speech, verb pattern, plurality, and word alignment. As shown in Figure 10, the dictionary component usage of the basic level is used the most of all the provided components. Moreover, the dictionary component usage indicates the component is a satisfactory component to use for all levels. It found that the highest frequency and percentage of each level. The verb pattern and word alignment components are the subordinate components for the provided component usage.

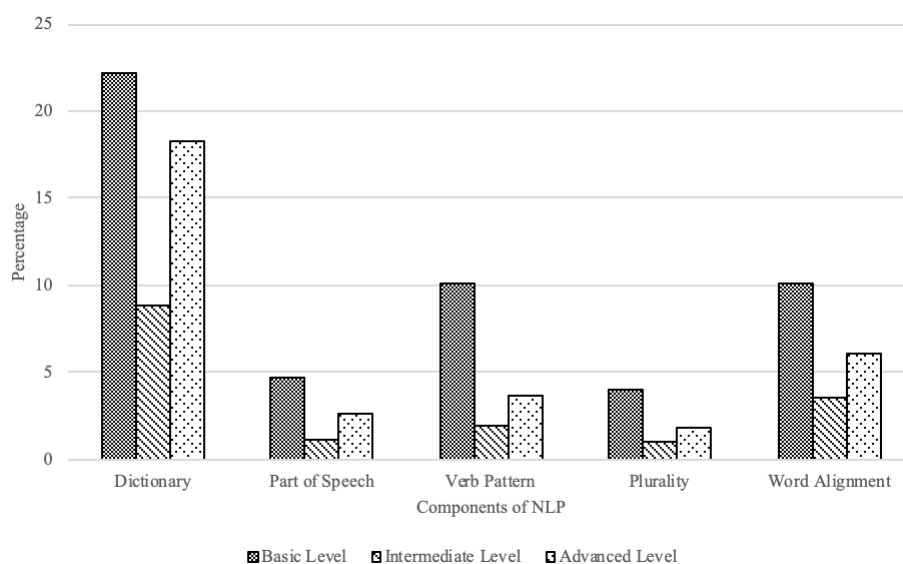


Figure 10. The percentage of component usage for assisting English composition

6. DISCUSSION

This paper aimed to develop a computer-based learning system with which to analyze the learning behavior observed. The system developed to acquire the learning behavior in the case of English writing of Thai EFL learners by using the provided system for composing the English sentences. All learning behavior from the provided system was used for learning behavior analytics. The learning analytics process was used to find out the Thai EFL learners’ pattern of learning behavior.

For the system development, the system was designed by incorporating concepts of the self-regulated learning model with the computer-based learning system. The concepts of self-regulated learning support lifelong autonomous learning for foreign language learning. From the behavior of system usage, we found that learners who assign the source sentences (Thai) by themselves spend less time composing the target sentences (English) than the learners who composed the target sentence from the system randomly. The reason for less time being required is that learners were able to choose the source sentence with which they were confident to compose the target sentence. The learners who set goals by themselves reflect to set goals in an initial phase (forethought) in the self-regulated learning model that indicates the concept of the self-regulated learning model supports the increase the learning performance.

Three groups of learning behavior patterns were observed: basic, intermediate and advanced. The basic level learners used the NLP components more than intermediate and advanced level learners, as shown in Figures 7 to 9. The components of NLP helped basic level learners to compose and modify the target sentences (in English) using components such as dictionary, word alignment and verb pattern. Intermediate level learners chose the dictionary and word alignment components to assist target sentence composition. On the other hand, The dictionary component was the main assisting component to compose and modify the target sentence for learners of advanced level. Moreover, The learners in the basic level used all assisting components more than the intermediate and advanced level learners, as may be observed from the percentage of component usage of basic level in Figure 10. Therefore, the basic level learners' component usage indicates that they have weaknesses in vocabulary and sentence structure, as shown in Table 2.

Table 2. The comparison of natural language processing component usage and grammatical aspects for Thai EFL learners in three levels

	NLP Components	Grammatical Aspects
Basic Level	Dictionary	Vocabulary
	Word Alignment	Sentence Structure and Tenses
	Verb Pattern	
Intermediate Level	Dictionary	Vocabulary
	Word Alignment	Sentence Structure and Tenses
Advanced Level	Dictionary	Vocabulary

7. CONCLUSION





This paper describes the importance of computer-based learning system development and learning analytics. Firstly, the multidisciplinary system integrates elements of the self-regulated learning model and components of NLP. Applying the self-regulated learning model supports personalized foreign language learning. The provided components as a writing guideline tool were used to assist English writing at the lexical and syntactic levels. The system collects the Thai EFL learners' writing behavior. The behavior data were necessary to generate learning behavior patterns. Since a computer-based learning system had not been developed for collecting writing behavior in Thailand, the learning behavior data from this system are useful for the analytics process that may assist in plans to improve Thai EFL learners language learning. Second, learning analytics is a useful process for finding the learning behavior pattern of Thai EFL learners in the case of English writing. Then, the behavioral patterns that use for reflecting learners and improving the learning materials or curriculum. For example, Thai EFL learners of all levels are required to improve vocabulary by observed from the frequency of dictionary component usage more than other components. The behavior sequential analysis was used to analyze the learning logs from the computer-based learning system. The 31 undergraduate students provided samples of their writing behavior via the computer-based learning system. The learning patterns of three groups of participants (basic, intermediate and advanced) were compared and found to be different. Since learners at the basic level have grammatical weaknesses, learners at the basic level use NLP components more than intermediate and advanced level learners. In the future, learning analytics is planned to use the machine learning method for learning behavior analysis. The machine learning method is used to create a pattern prediction model. The model is useful for improving personalized learning, learning material design and in the planning of English writing courses.

REFERENCES

- [1] K. Sermsook, J. Liamnimitr, and R. Pochakorn, "An Analysis of Errors in Written English Sentences: A Case Study of Thai EFL Students," *English Language Teaching*, vol. 10, no. 3, pp. 101–110, 2017, doi: 10.5539/elt.v10n3p101.
- [2] P. Pintrich, "A conceptual framework for assessing motivation and SRL in college students," *Educational Psychology Review*, vol. 16, no. 4, pp. 385–407, 2004, doi: 10.1007/s10648-004-0006-x.
- [3] M. Seker, "The use of self-regulation strategies by foreign language learners and its role in language achievement," *Language Teaching Research*, vol. 20, no. 5, pp. 600–618, 2016, doi: 10.1177/1362168815578550.
- [4] S. Karami, F. Sadighi, M. S. Bagheri, and M. J. Riasati, "The Impact of Application of Electronic Portfolio on Undergraduate English Majors' Writing Proficiency and their Self-Regulated Learning," *International Journal of Instruction*, vol. 12, no. 1, pp. 1319–1334, 2019.
- [5] M. S. Andrade and N. W. Evans, *Principles and Practices for Response in Second Language Writing*, in *Developing Self-Regulated Learners*, 1st ed. New York: Routledge, 2012, doi: 10.4324/9780203804605.





- [6] B. J. Zimmerman, "Academic studying and the development of personal skill: A self-regulatory perspective," *Educational Psychologist*, vol. 33, no. 2-3, pp. 73–86, 1998, doi: 10.1080/00461520.1998.9653292.
- [7] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," *Metacognition in educational theory and practice*, pp. 277–304, 1998.
- [8] L. S. Teng and L. J. Zhang, "Effects of motivational regulation strategies on writing performance: a mediation model of self-regulated learning of writing in English as a second/foreign language," *Metacognition and Learning*, vol. 13, no. 2, pp. 213–240, 2018, doi: 10.1007/s11409-017-9171-4.
- [9] J. B. Jason M. Lodge, Ernesto Panadero and P. G. de Barba, *Supporting self-regulated learning with learning analytics*, in *Learning Analytics in the Classroom*, 1st ed. New York: Routledge, 2019, p. 11, doi: 10.4324/9781351113038-4.
- [10] C. Wang, T. Fang, and Y. Gu, "Learning performance and behavioral patterns of online collaborative learning: Impact of cognitive load and affordances of different multimedia," *Computers and Education*, vol. 143, p. 103683, 2020, doi: 10.1016/j.compedu.2019.103683.
- [11] J. Shruithi and S. Swamy, "A prior case study of natural language processing on different domain," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 4928–4936, 2020, doi: 10.11591/ijece.v10i5.pp4928-4936.
- [12] K. R. Chowdhary, *Natural Language Processing*, in *Fundamentals of Artificial Intelligence*. New Delhi: Springer India, 2020, pp. 603–649, doi: 10.1007/978-81-322-3972-7_19.
- [13] X. Chen, "Research on nature language processing in the application of computer-assisted teaching," in *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, 2017, pp. 1457–1460, doi: 10.1109/ICCSN.2017.8230350.
- [14] A. N. Mathew, V. Rohini, and J. Paulose, "NLP-based personal learning assistant for school education," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 5, pp. 4522–4530, 2021, doi: 10.11591/ijece.v11i5.pp4522-4530.
- [15] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, "Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review," *Online Learning*, vol. 20, no. 2, pp. 13–29, 2016.
- [16] B. J. Zimmerman, "Self-Efficacy: An Essential Motive to Learn," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 82–91, 2000, doi: 10.1006/ceps.1999.1016.
- [17] S. Abadikhah, Z. Aliyan, and S. H. Talebi, "EFL students' attitudes towards self-regulated learning strategies in academic writing," *Issues in Educational Research*, vol. 28, no. 1, pp. 1–17, 2018, doi: 10.3316/ielapa.437737151804355.
- [18] S. Feldman, "NLP meets the jabberwocky natural language processing in information retrieval," *Online (Wilton, Connecticut)*, vol. 23, no. 3, pp. 62–72, 1999.
- [19] P. Luekhong, T. Ruangrajitpakorn, T. Supnithi, and R. Sukhahuta, "Pooja : Similarity-based Bilingual Word Alignment Framework for SMT," in *Proceedings of the 10th International Symposium on Natural Language Processing, Phuket, Thailand*, 2013.
- [20] K. Phodong and R. Kongkachandra, "Improvement of word alignment in Thai-English statistical machine translation by grammatical attributes identification," in *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2016, pp. 1–4, doi: 10.1109/ECAI.2016.7861155.
- [21] K. Phodong and R. Kongkachandra, "Improving Thai-English word alignment for interrogative sentences in SMT by grammatical knowledge," *2017 9th International Conference on Knowledge and Smart Technology (KST)*, 2017, pp. 226–231, doi: 10.1109/KST.2017.7886115.
- [22] I. Roll, R. S. Baker, V. Aleven, and K. R. Koedinger, "On the Benefits of Seeking (and Avoiding) Help in Online Problem-Solving Environments," *Journal of the Learning Sciences*, vol. 23, no. 4, pp. 537–560, 2014, doi: 10.1080/10508406.2014.883977.
- [23] I. Roll and P. H. Winne, "Understanding, evaluating, and supporting self-regulated learning using learning analytics," *Journal of Learning Analytics*, vol. 2, no. 1, pp. 7–12, 2015, doi: 10.18608/jla.2015.21.2.
- [24] S. Sfenrianto, E. Tantrisa, H. Akbar, and W. Mochamad, "E-learning effectiveness analysis in developing countries: East nusa tenggara, Indonesia perspective," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 7, no. 3, pp. 417–424, 2018, doi: 10.11591/eei.v7i3.849.
- [25] A. Pawar and V. Mago, "Calculating the similarity between words and sentences using a lexical database and corpus statistics," *Computation and Language*, pp. 1–14, 2018.
- [26] K. Phodong, T. Supnithi, and R. Kongkachandra, "A Framework of Computer-Based Learning System Based on Self-Regulated Model in English Writing," *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2019, pp. 1–6, doi: 10.1109/iSAI-NLP48611.2019.9045562.
- [27] R. Bakeman and J. M. Gottman, *Observing Interaction: An Introduction to Sequential Analysis*, 2nd ed. Cambridge, UK: Cambridge University Press, 1997, doi: 10.1017/CBO9780511527685.
- [28] V. Sornlertlamvanich, N. Takahashi, and H. Isahara, "Building a Thai part-of-speech tagged corpus (ORCHID)," *Journal of the Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 189–198, 1999, doi: 10.1250/ast.20.189.
- [29] C. Tapsai, P. Meesad, and C. Haruechaiyasak, "TLS-ART: Thai Language Segmentation by Automatic Ranking Trie," *The 9th International Conference Autonomous Systems (AutoSys 2016)*, 2016.
- [30] K. Trakultaweekoon, P. Porkaew, and T. Supnithi, "LEXiTRON Vocabulary Suggestion System with Recommendation and Vote Mechanism," in *The International Symposium on Natural Language Processing (SNLP 2007)*, 2007, pp. 43–48.
- [31] S. Meknavin, P. Charoenpornasawat, and B. Kijisirikul, "Feature-based Thai Word Segmentation," *Proceedings of Natural Language Processing Pacific Rim Symposium, Phuket, Thailand*, 1997, vol. 97, pp. 41–46.
- [32] F. J. Och and H. Ney, "A Systematic Comparison of Various Statistical Alignment Models," *Computational Linguistics*, vol. 29, no. 1, pp. 19–51, 2003, doi: 10.1162/089120103321337421.
- [33] C. Yin *et al.*, "Learning behavioral pattern analysis based on students' logs in reading digital books," *Proceedings of the 25th International Conference on Computers in Education, ICCE 2017 - Main Conference Proceedings*, 2017, pp. 549–557.
- [34] C. Yin, Z. Ren, A. Polyzou, and Y. Wang, "Learning behavioral pattern analysis based on digital textbook reading logs," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11587 LNCS, no. August, pp. 471–480, 2019, doi: 10.1007/978-3-030-21935-2_36.

BIOGRAPHIES OF AUTHORS

Kanyalag Phodong     is a Ph.D. candidate at the Department of Computer Science, Thammasat University, Thailand. She graduated with Bachelor's Degree in Computer Science from Naresuan University, Thailand (2007). She obtained a Master of Computer Science from Thammasat University, Thailand in 2011. Her researches are in fields of Artificial Intelligence, Natural Language Processing, Learning Analytics, and Information System. She can be contacted at email: kanyalagp@gmail.com.





Further info on her homepage: <https://sites.google.com/sci.tu.ac.th/ai-lab/members>



Thepchai Supnithi     is an Artificial Intelligence Research Group Director at National Electronics and Computer Technology Center, National Science and Technology Development Agency, Thailand. He graduated with a Bachelor of Science in Mathematics at Chulalongkorn University (1992). He obtained a Master's degree and a Doctor of Philosophy of Electronics and Computer Engineering at Graduate School of Engineering, Department of Knowledge System, Osaka University, Osaka, Japan (2001). His researches are in the fields of Computer in Education, Knowledge Engineering, Natural Language Processing, Machine Learning, Knowledge Management, and Digital Culture. He can be contacted at email: thepchai@nectec.or.th.

Further info on his homepage: <https://www.nectec.or.th/hccru/lst>



Rachada Kongkachandra     is an Assistant Professor at Data Science and Innovation Program, College of Interdisciplinary Studies, Thammasat University. Her bachelor's degree was B.Sc. (Computer Science), Thammasat University, Thailand, M.Sc. (Computer Technology) from Asian Institute of Technology, Thailand. She obtained Ph.D. in Electrical and Computer Engineering from King Mongkut's University of Technology Thonburi, Thailand. Her researches are in the fields of Artificial Intelligence, Natural Language Understanding, Text Network Analysis, Data Science, and Speech Recognition. She can be contacted at email: krachada@staff.tu.ac.th.

Further info on her homepage: <https://ci.tu.ac.th/uploads/ci/instructor/cv/rachada.pdf>