

Mining Salient Concepts of IT-Support-Tasks from Daily IT-Support Ticket Reports

การทำเหมืองแนวความคิดงานที่สำคัญของงานสนับสนุนด้านไอที จากรายงานтикเกิดสนับสนุนด้านไอทีรายวัน

Siraporn Palaya¹, Chaveevan Pechsiri¹, Supakpong Jinarat^{1*} and Adisorn Na Ubon¹

ศิริพร ปาลายา¹ ฉวีวรรณ เพ็ชรศิริ¹ สุภัคพงศ์ จินารัตน์^{1*} และอดิสร ณ อุบล¹

¹College of Engineering and Technology, Dhurakij Pundit University

¹วิทยาลัยวิศวกรรมศาสตร์และเทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิตย์

*Corresponding author: supakpong.jin@dpu.ac.th

Received: January 29, 2026

Revised: April 16, 2026

Accepted: April 22, 2026

Abstract

IT help-desk staff at a university computer center create tickets and record end-user request calls as request documents containing explanations of IT-Support-Task (IST) concepts. The ticket's request document contains several elementary discourse units (EDUs, which are simple sentences) including a salient IST-concept EDU. The research focuses on mining, including the extraction of EDUs with salient IST-concepts from consecutive daily IT-support ticket reports to classify the salient IST-concepts EDUs (IT, Network, Mix and Miscellaneous). Classification results are used to represent daily task loads graphically, which helps with planning and decision-making for resource reallocation. There are three main research gaps: how to determine the IST-concept EDUs with noun phrase ellipses in the documents, where the noun phrase ellipses of the EDUs may result in loss of contextual information; how to determine/extract the salient IST-concept EDU as the main content of the request, and how to classify the tickets' salient IST-concept EDUs with ambiguous task concepts. We then propose using a word-Verb cooccurrence (wrVCo) set with IST concepts to determine the IST-concept EDUs (where wrVCo can be used as primary keywords to analyze the meaning, main idea, and context of an EDU). We also applied a salient-concept occurrence rule to extract the salient IST-concept EDU from the document. Machine learning techniques, SVM and MLP, are then applied to classify the tickets' salient IST-concept EDUs. The research results show high F1-scores of the salient IST-concept EDU extraction and classification from the tickets.

Keywords: word cooccurrence, help-desk ticket, salient concept

บทคัดย่อ

สำหรับระบบบริการช่วยเหลือด้านเทคโนโลยีสารสนเทศ (IT help-desk system) ของศูนย์คอมพิวเตอร์ในมหาวิทยาลัย เจ้าหน้าที่ช่วยเหลือด้านไอทีจะสร้างใบรับแจ้งหรือทิกเก็ต (ticket) ที่ถูกบันทึกเป็นเอกสารคำร้องขอของผู้ใช้งาน เอกสารคำร้องอธิบายตามแนวความคิดงานสนับสนุนไอที (IT-Support-Task: IST concept) หรือแนวความคิดไอเอสที (IST concept) เอกสารคำร้องของทิกเก็ตประกอบด้วยหน่วยข้อความระดับวาทกรรมย่อย (Elementary Discourse Unit: EDU/อิตี๋ย หรือประโยคความเดียว) หลายหน่วยที่มีหนึ่งอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญ งานวิจัยนี้มุ่งเน้นการทำเหมืองข้อมูลโดยเฉพาะการสกัดอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญจากรายงานทิกเก็ตสนับสนุนไอทีรายวันแบบต่อเนื่องเพื่อใช้ในการจำแนกอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญเหล่านี้ออกมาเป็นคลาสต่าง ๆ (ไอที เครือข่าย ผสม และอื่นๆ) ผลลัพธ์จากการจำแนกอิตี๋ยดังกล่าวนี้สามารถนำไปใช้ในการแสดงปริมาณงานรายวันในรูปแบบกราฟ ซึ่งช่วยสนับสนุนการวางแผนและตัดสินใจสำหรับการบริหารจัดการทรัพยากร งานวิจัยนี้มีปัญหาหลัก 3 ประการ ได้แก่ (1) วิธีระบุอิตี๋ยที่มีแนวความคิดไอเอสทีที่มีการละนามวลีบางส่วนในเอกสารซึ่งการละนามวลีของอิตี๋ยต่าง ๆ นำไปสู่การสูญเสียบริบทบางส่วนของอิตี๋ยเหล่านั้น (2) วิธีระบุ/สกัด อิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญซึ่งเป็นเนื้อหาหลักของคำร้อง และ (3) วิธีจำแนกอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญของทิกเก็ตที่มีแนวความคิดงานกลุ่มเครือข่าย ดังนั้นเราเสนอการใช้เซตการเกิดร่วมของคำกับกริยา (word-Verb cooccurrence: wrdVCo) ที่มีแนวความคิดไอเอสที มาทำการระบุอิตี๋ยที่มีแนวความคิดไอเอสที (โดย wrdVCo สามารถใช้เป็นคีย์เวิร์ดหลักในการวิเคราะห์ความหมาย สำคัญ และบริบทของอิตี๋ยนั้นได้) และประยุกต์ใช้กฎการปรากฏแนวความคิดที่สำคัญภายในเอกสารมาทำการสกัดอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญออกจากเอกสารนั้น จากนั้นได้นำเทคนิคการเรียนรู้ของเครื่อง ได้แก่ SVM และ MLP มาประยุกต์ใช้ในการจำแนกอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญจากทิกเก็ตต่าง ๆ ผลการวิจัยแสดงให้เห็นว่า การสกัดและการจำแนกอิตี๋ยที่มีแนวความคิดไอเอสทีที่สำคัญ มีค่าคะแนน F1 (F1-Score) สูง

คำสำคัญ: การเกิดร่วมของคำ ใบรับแจ้งปัญหา แนวความคิดที่สำคัญ



Introduction

Recent IT research shows a clear trend toward developing data-driven information systems that improve operational efficiency and decision quality across domains such as retail operations, industrial process measurement, higher-education analytics, and online information credibility assessment (Aun-a-nan & Meesad, 2020; Pakamwang et al., 2020; Sutta et al., 2025; Thimsanthia & Soemphol, 2025).

Most previous studies on data-driven information systems have focused primarily on structured data or tabular entities, without considering attributes containing textual descriptions as unstructured data. In general, textual-description data provide

useful information, including concepts that can be mined or extracted for the workforce planning and service management of the enterprise. Therefore, this research aims to mine, extract, and classify IT-Support-Task: IST concepts, particularly salient IST concepts, from consecutive daily IT-support ticket reports generated by the help desk of the computer center at Dhurakij Pundit University: DPU, Thailand.

According to Paramesh et al. (2018), an IT-support ticket represents a single user request or technical issue submitted through a request message or document that requires attention. In this study, the ticket is logged by help-desk staff without pre-selecting a request category. The daily IT-support ticket report records all tickets

logged each day, where each ticket contains an IT-Support-Task description or detail represented by an IST concept.

The classification results of the extracted IST concepts, particularly salient IST concepts, are aggregated each day and used to generate graphical representations of daily task loads. Each task load represents the total number of classified salient IST concepts that require a person, team, or system to handle within a specific period. A request document typically contains several explanatory sentences or Elementary Discourse Units: EDUs. Among these, one key sentence or EDU represents the salient concept of the request message. An EDU is defined as a simple sentence or clause (Carlson et al., 2003).

In general, the IST concepts include tasks such as setting user accounts, software installation, configuration, troubleshooting, maintenance, the repair of computer systems, network support, and documentation. Several tickets of users are collected

into Ticket which is a ticket entity (or a ticket table) consisting of the following attributes or features: TID: Ticket Identification, date&time_Request, Request-Detail (the request document recorded by help-desk staff based on calls from end users), Responsible-Worker, Start-Time, End-Time, Task-Time (the amount of time between Start-Time and End-Time), Status (finished or pending), and Comment. Thus, Ticket contains a set of collected ticket instances, {ticket_j} where $j=1,2,...,numOfTicketInstances$; i.e., $numOfTicketInstances$ is 1700 tickets.

The daily IT-support ticket report therefore contains multiple ticket instances, each including a request notification document in the Request-Detail attribute/feature that describes IT-service concepts including IT-problem concepts. In addition to Ticket, a ticket_j.Request-Detail feature value (which is a content value from the Request-Detail feature of a ticket_j instance) contains one or more EDU expressions describing event concepts as shown in Example 1.

Example 1: The ticket_j.Request-Detail feature value has the IST concept EDUs in EDU2-EDU4 with the salient concept expression on EDU2 as “be_slow(computer)” where the [...] symbol means an ellipsis.

EDU1: “คอมพิวเตอร์ที่ใช้อยู่ของธุรการคณะ/A Computer is used by the faculty secretary”

EDU2: “คอมพิวเตอร์ช้า/The computer slows down.”

EDU3: “มีวงล้อหมุนหมุน/There is a spinning circle”

EDU4: “[วงล้อ]หมุนไม่หยุด/The spinning circle] keeps spinning.”.

The EDU occurrences in Example1 are based on the following general Thai linguistic pattern after stemming words and eliminating stop words:

EDU → NP1 VP | VP | NP1

VP → Verb NP2 | Verb | Verb Adv

Verb → Verbweak Noun1 | Verbstrong | Neg Verb | Verb Neg | Verb...Neg

NP1, NP2 → Noun2 | Noun2 Phrase; Phrase → Adj| adjectivePhrase | prepositionPhrase

Verbweak → {‘เป็น/be’, ‘มี/have’}

Verbstrong → {‘ติดตั้ง/install’, ‘ดับ/close’, ‘ช้า/beSlow’, ‘ติด/connect’, ‘เข้า/access,login’, ‘ขึ้น,ออก/present’ ...}

Noun1 → {‘สี/color’, ‘เสียง/sound’, ‘เสียงดัง/noise’, ‘ภาพ/image’ ...}

Noun2 → {‘คอมพิวเตอร์/computer’, ‘โปรเจคเตอร์/projector’, ‘สายแลน/LANcable’, ‘จอ/screen’ ...}

Neg → {‘ไม่/not’, ‘ไม่ได้, ไม่ติด, ไม่ขึ้น/not’, ‘ไม่ได้/cannot’ ...}

where NP1 and NP2 are noun phrases; Noun1 and Noun2 are the noun concept sets of properties and computer-hardware respectively; Verb is a verb term consisting of (1) a strong verb concept set (Verbstrong) of the action concepts and (2) a weak verb concept set (Verbweak) plus Noun1 to become the action concepts; Adv is an adverb concept set; Adj is an adjective concept set.

Thus, the ticketj.Request-Detail feature value is represented by its salient concept. The salient concepts of the Request-Detail feature values from the Ticket instances are then categorized into four task-concept classes/groups on expertism; (1) an IT-task concept class (called ‘IT-Task’) handled by an IT expert group in software installs and supports hardware, troubleshoots applications, and carries out user account management, (2) a network-task concept class (called ‘NW-Task’) handled by a network expert group in network security, i.e., firewalls, and VPNs, Wi-Fi management, server maintenance including emails, and file storage, (3) a mixed-task concept class (called ‘Mix-Task’) handled by both an IT expert group and a network expert group, e.g., a finger scan error which involves IT-Task and NW-Task in updating a finger scan data through Wi-Fi accounts, and (4) a miscellaneous-task concept group (called ‘Mis-Task’), e.g., fixing classroom instruments (projectors, speakers, and microphones) and parameters, handled by a computer technician group.

Thus, our research consists of three main research gaps: (1) how to locate and determine IST-concepts from EDU occurrences with some Noun Phrase: NP ellipses in the ticketj.Request-Detail feature values, (2) how to extract the

salient IST-concept EDU from EDUs narrating the IST concepts in each instance of ticketj.Request-Detail feature value, and (3) how to classify the extracted salient IST-concept expression of each instance into the following four task-concept classes: IT-Task, NW-Task, Mix-Task, and Mis-Task; with some ambiguous task concepts. For example, an install software task may be classified either as IT-Task (for operating systems such as Windows) or as Mis-Task (for application software such as Microsoft Office or a web browser).

The classification results of the extracted salient IST-concept expressions are useful to represent the daily task load of each class graphically using the test data from the ticket dataset or the ticket table (Ticket). The daily task-load presentation enables a manager to detect patterns, such as recurring peaks on certain days and supports human resource management in planning or reallocating resources.

Previous studies on task/task-concept determination or classification have relied on machine learning techniques based on noun phrases, verb phrases, or sentence expressions (Al-Hawari & Barham, 2021; Diwanji et al., 2020; Huang et al., 2013; Lin et al., 2018; Paramesh et al., 2018; Pawar et al., 2021) without concerning the salient IST-concept EDU expressions with some noun phrase ellipses. To approach these problems, we develop a framework which combines a machine learning technique with the linguistic phenomena to determine/extract and classify several EDU expressions with the salient IST concepts. We then apply the following request-verb concept set (Vr), collected from the corpus behavior study, to locate each EDU occurrence as an IST event expression.

$V_r = \{ \text{'ติดตั้ง/install'}, \text{'ดับ/fail'}, \text{'ไม่+ออก/not+show, not+appear'}, \text{'เปิด+ไม่ได้, เปิด...+ไม่ได้ /not+open, not+switch on'}, \text{'เข้า+ไม่ได้, เข้า...+ไม่ได้ /not+access'}, \text{'หัก/break'}, \text{'ค้าง/freeze'}, \text{'มี+สี/have+color'}, \text{'เป็น+สี/be+color'}, \text{'ไม่มี+เสียง/not+have+sound'}, \dots \}$,

We also use a relatedness value (r) Guthrie et al. (1991); Pechsiri and Piriyaikul (2016) between words in a word-Verb cooccurrence (called ‘wrVCo’) derived from two-word sets, v_r and Wrd_{co} (a co-occurred word set; $Wrd_{co} = \text{Noun1} \cup \text{Noun2} \cup \text{Adv} \cup \text{Adj}$), within an EDU occurrence that contains an IST event expression with an IST concept based on WordNet (Miller, 1995). After stemming words and eliminating the stop words of an EDU expression, the wrVCo expression consists of wrd_{co} and v_r forming a $v_r \ wrd_{co}$ cooccurrence where $v_r \in V_r$; $wrd_{co} \in Wrd_{co}$; and wrd_{co} is also a head noun of NP1. If NP1 ellipsis occurs, wrd_{co} is the word right after Verb or V_r . In addition, each wrVCo expression with the higher relatedness value of r (v_r, wrd_{co}) between v_r and wrd_{co} , is collected into a wrVCo set (called WVC; see subsection (2) in the Methodology section). We then propose using V_r and WVC to identify and extract each $v_r \ wrd_{co}$ pair as wrVCo representing an IST concept from each EDU in the $ticket_j$.Request-Detail feature value.

According to the corpus behavior study, together with the general rule that salient expressions are often articulated in the opening sentence of the document (Hassan & Mihalcea, 2011; Whiteley, 2024), the $ticket_j$.Request-Detail feature value typically contains the salient IST-concept EDU occurrence within the first EDU of the sequence of EDUs explaining the IST-concepts (see Example1). Finally, we apply supervised learning techniques, e.g., Support Vector Machine (SVM) (Cristianini & Shawe-Taylor, 2000) and Multi-Layer Perceptrons : MLPs (Haykin, 1999), to classify the extracted salient IST-concept EDUs in the $ticket_j$.Request-Detail feature into the IT-Task, NW-Task, Mix-Task, and Mis-Task classes. The classification results of the extracted salient IST-concept EDU classes from the Request-Detail feature values—including the

date&time_Request feature values and the Task-Time feature values—are then used to represent the daily task loads of the IST-concept classes graphically.

The rest of this paper is organized as follows: related works are discussed in the Related Work section, our methodology is described in the Methodology section, the results and discussion are presented in the “Results and Discussion” section. Finally, the conclusions are presented in the Conclusions section.

Related Works

Most of the previous research studies attempted to mine, extract or classify the task concepts from email/chat requests or ticket descriptions for information services. Paramesh et al. (2018), constructed a feature vector of a ticket instance from the ticket description in English using a Term Frequency-Inverse Document Frequency: TF-IDF weighting scheme without considering the NP ellipsis and the salient concept. The feature vectors from 10742 ticket instances with 70% of tickets used as a training set and the remainder were used to test various classification models based on three ensemble methods, i.e., Bagging, Boosting and Voting ensemble methods along with some base classifier models to build the classifiers. The F1-score results of all the ensemble classifiers were better than the corresponding base classifiers. Huang et al. (2013), worked on request-document categorization to analyze internal service requests in Chinese by employing Naïve Bayes: NB, SVM and ID3 without concerning the salient concept to solve the document delivery error. The corpus with the part-of-speech annotation consisted of 950 labeled request sentences to learn five classes. The average F1-score results were 0.56,

0.59, 0.52 for NB, SVM, and ID3 respectively. Lin et al. (2018), worked on the email recipient action classification with RNNs from an English corpus of 399 emails and non-email messages. They employed action-based annotation for word embeddings including Bag-of-Words as an encoder. They reparametrized RNNs outperform on the task classification with an average F1-score of 0.32. Al-Hawari and Barham (2021), classified ticket messages to assist users in selecting the right service where each message was a concatenation of title, description, and comment. SVM was applied with the TF-IDF feature vectorization without concerning the textual concepts and the NP ellipsis in classifying the messages with the 81.4 % accuracy on 331 test tickets and 1254 training tickets. Diwanji et al. (2020) worked on identifying tasks or requests from the 1,000 emails based on verb phrases with task labels from an Enron corpus and chat corpus with 114 dialogues with 300 sentences. They proposed dependency parsing and VerbNet to identify verb phrases for task classification by employing Recurrent Neural Network: RNN without concerning the salient concept. Their methodology F1-score results were 0.80 and 0.89 for the email and chat corpus respectively. Pawar et al. (2021)

proposed two techniques: using linguistic patterns and a BERT-based supervised neural model on the head words, for task extraction of either VPs or NPs without considering the salient task concepts. The corpus consisted of 19268 sentences for training and 1869 sentences for the test with average F1 scores of 0.46 and 0.56 for the linguistic patterns and BERT respectively.

Thus, most of the previous works (Al-Hawari & Barham, 2021; Diwanji et al., 2020; Huang et al., 2013; Lin et al., 2018; Paramesh et al., 2018; Pawar et al., 2021) focused on mining/extracting/ classifying the task/request concept from the texts without concerning their salient concepts with some NP ellipses.

Methodology

There are five steps in our framework: Corpus Preparation and Annotation, Relatedness Value Determination and Collection of WVC having IST Concepts, Extraction of Salient IST Concepts from Tickets' Request Documents, Salient IST-Concept Class Learning, and Determination of Tickets' Salient IST-Concept Classes (see Figure 1).

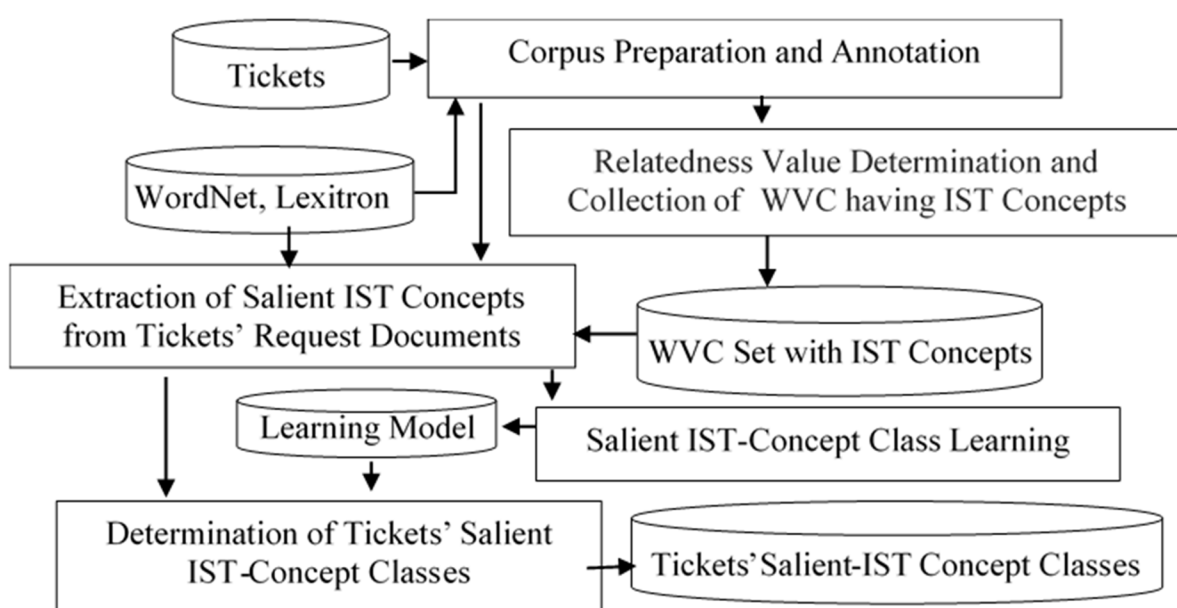


Figure 1 System Overview

1. Corpus Preparation and Annotation

This step involves the preparation of an EDU corpus collected from the ticket instances on the *ticket_j.Request-Detail* feature values ($j=1, 2, \dots, numOfTicketInstances$) as the task request documents in the consecutive daily IT-support ticket reports of the help desk at the DPU computer center. The step involved using Thai word segmentation tools (Sudprasert & Kawtrakul, 2003), Named Entity recognition (Chanlekha & Kawtrakul, 2004; Tongtep & Theeramunkong, 2015), and EDU Segmentation (Chareonsuk et al., 2005; Ketui et al., 2013) to provide the 4500 EDUs' corpus from the Request-Detail feature values of 1700 tickets. The tickets were separated into three parts: 350 tickets (911 EDUs) on the Request-Detail feature for the corpus behavior study; 250 tickets contained the corpus annotation on the Request-Detail feature (see Figure 2) including stemming word and stop word removal to determine the relatedness value $r(v_r, wrd_{co})$ used for the WVC collection from the Request-Detail feature values of this part; and 1100 tickets (2860 EDUs) on the Request-Detail feature values were separated into two parts: one part was used to learn the salient IST-concept classes, and the other part was used as the test data to extract the salient IST concepts through the salient IST-concept classification

2. Relatedness Value Determination and Collection of WVC having IST concepts

After stemming words and eliminating stop words, the relatedness value (Guthrie et al., 1991) was applied in this research to indicate the relatedness value (see equation (1)) between two sequent words without the word-order consideration on the *ticket_j.Request-Detail* feature values

($ja=1,2,\dots,numOfTicketsForCorpusAnotation$ which is 250 tickets of the second part) with the following tags, V_r as a v_r tag, Wco as a wrd_{co} tag and $WV-Co$ as a $wrdVCo$ tag having an IST-concept class as a positive/negative concept class on the annotated corpus (Figure 2).

$$(v_r, wrd_{co}) = \frac{fv_r, wrd_{co}}{fv_r + fwrd_{co} - fv_r, wrd_{co}} \quad (1)$$

where $r(v_r, wrd_{co})$ is a relatedness of $wrdVCo$ with the IST concept. fv_r is the number of v_r occurrences. $fwrd_{co}$ is the number of wrd_{co} occurrences. fv_r, wrd_{co} is the number of v_r, wrd_{co} co-occurrences without the word-order consideration.

If a v_r, wrd_{co} co-occurrence has the IST concept, the relatedness $r(v_r, wrd_{co})$ value is the positive concept. Another relatedness value of the same v_r, wrd_{co} co-occurrence is the non IST concept as the negative concept. Thus, the $wrdVCo$ expression, v_r, wrd_{co} , is collected from the annotated tickets of the *ticket_{ja}.Request-Detail* feature values into WVC if the relatedness $r(v_r, wrd_{co})$ value of the $wrdVCo$ expression with the positive concept is higher than the one of the $wrdVCo$ expression with the negative concept. WVC is used for mining or extracting the salient IST concepts from the *ticket_j.Request-Detail* feature values ($j = 1, 2, \dots, numOfTicketsForSalientISTconceptsExtraction$ which is 1100 tickets).

3. Extraction of Salient IST Concepts

The salient IST-concept Extraction from the *ticket_j.Request-Detail* feature is shown in the following algorithm.

<Ticket id = 1; Request-Detail Feature : “คอมพิวเตอร์เปิดไม่ติด/The computer cannot be turned on.”^{EDU1}
 “กรุณามาดูเร็วหน่อยค่ะ/please come take a look quickly.”^{EDU2} “มีนักเรียนกำลังสมัครเรียนอยู่/There is a student in the process of enrolling.”^{EDU3}>

<EDU1 wrdVCo: yes Salient EDU: yes>
 ((คอมพิวเตอร์/computer)/Noun)/NP1 ((เปิด/be turned on)/Verbstrong (ไม่ติด/not)/Neg)/VP</EDU1>
 <WV-Co: IST concept class: **Positive Concept** = ‘computer/device-unableToTurnOn’>
 <Wco : concept= ‘computer/device’>คอมพิวเตอร์</Wco>
 <Vr concept= ‘not’ + ‘turnOn/switch’> เปิด+ไม่ติด</Vr></WV-co>

<EDU2 wrdVCo: no Salient EDU: no>
 ((กรุณา/please)/Adv (มา/come)/Verb (ดู/look)/Verb (เร็วหน่อยค่ะ/quickly)/Adv)/VP</EDU2>

<EDU3 wrdVCo: no Salient EDU: no>
 ((มี/There be)/Verb (นักเรียน/student)/Noun (กำลังสมัคร/be-applying)/Verb (เรียนอยู่/study)/Verb)/VP
 </EDU3></Ticket id>

<Ticket id = 2; Request-Detail Feature : “[ผู้ใช้]เปิด คอมพิวเตอร์ไม่ได้/[The user]cannot start the computer.”^{EDU1} “เมื่อวาน[ผู้ใช้]ยังเปิดได้/[The user] could still turn on yesterday.”^{EDU2}>

<EDU1 wrdVCo: yes Salient EDU: yes>
 [([ผู้ใช้/user)/Noun)/NP1]
 ((เปิด/turn on)/Verbstrong ((คอมพิวเตอร์/computer)/Noun)/NP2 (ไม่ได้/not)/Neg)/VP</EDU1>
 <WV-Co: IST concept class: **Positive Concept** = ‘computer/device-unableToTurnOn’>
 <Wco : concept= ‘computer/device’>คอมพิวเตอร์</Wco>
 <Vr concept= ‘not’ + ‘turnOn/switch’> เปิด+ไม่ติด</Vr></WV-co>

<EDU2 wrdVCo: no Salient EDU: no>
 (เมื่อวาน/yesterday)/Adv [([ผู้ใช้/user)/Noun)/NP1] (ยังเปิดได้/can turn on)/Verbstrong </EDU2>
 </Ticket id>

<Ticket id = 3; Request-Detail Feature : “จอ [โปรเจคเตอร์]ไม่ขึ้น [ภาพ] / [The projector] screen cannot show [an image].”^{EDU1} “โน้ตบุ๊กเป็นรุ่นเก่า/The laptop is an old model .”^{EDU2} “มันใช้กับ โปรเจคเตอร์ห้องอื่นได้/ It works with projectors in other rooms.”^{EDU3}>

<EDU1 wrdVCo: yes Salient EDU: yes>
 ((จอ [โปรเจคเตอร์]/ [projector] screen)/Noun)/NP1
 ((ไม่/not)/Neg (ขึ้น/show)/Verbstrong ([ภาพ/image)/noun])/VP</EDU1>
 <WV-Co: IST concept class: **Negative Concept** = ‘screen/device-unableToShow’>
 <Wco : concept= ‘projector screen/device’>จอ</Wco>
 <Vr concept= ‘not’ + ‘show’> ไม่+ขึ้น </Vr></WV-co>

<EDU2 wrdVCo: no Salient EDU: no>
 ((โน้ตบุ๊ก/laptop)/Noun)/NP1 ((เป็น/be)/Verbweak (รุ่นเก่า/old model)/Adj)/VP</EDU2>

<EDU3 wrdVCo: no Salient EDU: no>
 ((มัน/laptop)/Noun)/NP1 ((ใช้/use)/Verbweak (กับ/with) (โปรเจคเตอร์/projector)/Noun (ห้อง/room)/Noun (อื่น/other)/Adj (ได้/able)/Aux)/VP</EDU3></Ticket id>

Figure 2 Corpus annotation of the IST concept on the Request-Detail feature values with a wrdVCo tag (named “WV-Co”) including a wrdco tag (named “Wco”) and a vr tag (named “Vr”). If there is a negation (‘not’) in a verb phrase, the negation is added to the Vr concept.

Algorithm: Salient_IST_Concept_Extraction

```

// Process each ticket instance
For  $je = 1$  to  $numOfTicketsForSalientISTconceptsExtraction$ 
Begin_For
  L= the number of EDUs from  $ticket_{je}.Request-Detail$ ;  $ct=0$ ;  $flag=0$ ;
  While ( $ct \leq L$ ) and  $flag=0$  do
    Begin_WhileLoop
      //According to  $EDU_{ct}$ , check Verb as an element of the  $V_r$  set and
      // check an NP1 head noun or a word right after Verb as an element of the  $Wrd_{co}$  set
      If ( $ticket_{je}.Request-Detail.EDU_{ct}.Verb \in V_r$ )  $\wedge$ 
        ( $ticket_{je}.Request-Detail.EDU_{ct}.headNounOfNP1 \in Wrd_{co}$ ) then
        Begin_If  $w_{je,ct} = ticket_{je}.Request-Detail.EDU_{ct}.headNounOfNP1$ ;  $flag=1$ 
        End_If
      Else If ( $ticket_{je}.Request-Detail.EDU_{ct}.Verb \in V_r$ )  $\wedge$ 
        ( $ticket_{je}.Request-Detail.EDU_{ct}.wordRightAfterVerb \in Wrd_{co}$ ) then
        Begin_Elself  $w_{je,ct} = ticket_{je}.Request-Detail.EDU_{ct}.wordRightAfterVerb$ ;  $flag=1$ 
        End_Elself
      If  $flag=1$  then
        Begin_flag_is_true
           $v_{je,ct} = ticket_{je}.Request-Detail.EDU_{ct}.Verb$ ;
          If  $ticket_{je}.Request-Detail.EDU_{ct}.VP$  contain 'is/not' then  $v_{je,ct} = 'not' + v_{je,ct}$ ;
          If  $v_{je,ct} w_{je,ct} \in WVC$  then
             $ticket_{je}.Request-Detail.EDU_{ct}.wrdVCoSalient = v_{je,ct} w_{je,ct}$ 
            // an  $EDU_{ct}$  with wrdVCo concept as an  $EDU_{ct}$  with salient IST Concept
          Else  $flag = 0$ ;
          End_flag_is_true
         $ct++$ 
      End_WhileLoop
    End_For
  Return  $ticket_{je}$ 

```

4. Salient IST-concept Class Learning

According to the Salient_IST_Concept_Extraction algorithm, there were 1060 tickets as the correct extraction of the $ticket_{je}.Request-Detail.EDU_{ct}.wrdVCoSalient$ contents with salient IST concepts from 1100 tickets. The 1060 tickets were then separated into two parts, (1) one part of 670 tickets as $ticket_{jl}$ ($jl=1,2,...,numOfTicketsForLearningClasses$) containing

1740 EDUs of the Request-Detail feature values with 670 salient IST-concept EDUs with wrdVCo expressions for supervised learning the salient IST-concept classes (i.e., IT-Task, NW-Task, Mix-Task, and Mis-Task) and (2) the other part of 390 tickets as $ticket_{jt}$ ($jt=1,2,...,numOfTicketsForClassificationTestSamples$) containing 390 salient IST-concept EDUs with wrdVCo expressions on $ticket_{jt}.Request-Detail$ to become

the test sample. In addition, the learning part of the extracted wrdVCo expressions is based on 10-fold cross validation by using Weka (Bouckaert et al., 2016; Hall et al., 2009) with the following SVM and MLPs.

SVM learning (Cristianini & Shawe-Taylor, 2000) with the linear kernel: The linear function, $y=f(x)$, y is a binary vector, of the input $x = (x_1 \dots x_n)$ assigned to the positive class if $f(x) \geq 0$, and otherwise to the negative class if $f(x) < 0$, can be written as follows:

$$\begin{aligned} f(x) &= \langle w \cdot x \rangle + b \\ &= \sum_{k=1}^n w_k x_k + b \end{aligned} \quad (2)$$

where: x is a dichotomous vector; w is a weight vector; b is a bias; and $(w, b) \in R_n \times R$ are the parameters that control the function. Regarding our learning samples, SVM is applied to determine that the combination between Vr and Wrdco is used to predict the ticket concept classes, particularly the salient-concept classes. The SVM learning then determines w_k and b for v_{r-c} ($c=1,2,\dots,numOfVrElements$) and wrd_{co-d} ($d=1,2,\dots,numOfWrdElements$) features or variables (x_k) in each $v_{r-c} wrd_{co-d}$ pair with either the positive class (ISTConceptClass, $y_u=1$; $u = 1,2,\dots, numberOfTestSamples$) or the negative class (nonISTConceptClass, $y_u=0$) which relies on the One-vs-All techniques and the supervised learning instances.

MLPs: According to Haykin (1999), Artificial Neural Networks (ANNs) are composed of neuron-like units connected together through input and output paths that have adjustable weights. Each node (neuron) produces an output signal, which is a function of the sum of its inputs. The function is formulated in equation (3).

$$y_k = f\left(\sum x_k w_k\right) \quad (3)$$

where w_k represents the weight, x_k is the input feature or variable of the input node. There are two input nodes of v_{r-c} , wrd_{co-d} ($c=1,2,\dots,numOfVrElements$; $d=1,2,\dots,numOfWrdElements$) from EDUs with the salient IST concepts. $f(\cdot)$ is the activation function such as a sigmoid function, and y_k is the output of the k th node. MLP consists of an input layer, hidden layers, and an output layer which produces the output pattern/class. At the output layer, there are four nodes of different classes: IT-Task, NW-Task, Mix-Task, and Mis-Task. Thus, MLP applied in determining the salient IST-concept classes is based on the binary classes with multilevels since the activation function is generally a binary-value function (either 0 or 1). Each layer includes a different number of processing nodes. The net weighted input can then be solved by equation (4) where n is the number of neuron inputs, θ_j is the threshold value of the neuron at the j^{th} node in the hidden layer, and the number of hidden layers $p=2$

$$y_j(p) = \sum_{k=1}^n x_k(p) w_{kj}(p) - \theta_j \quad (4)$$

5. Determination of tickets' Salient IST-concept Classes

SVM: The salient IST-concept class of each ticket from the test samples (*ticketjt*. Request-Detail.EDU_{ct}.wrdVCo) containing the TID feature and the $v_{r-c} wrd_{co-d}$ pair on the salient IST-concept EDU is determined by the weight vector (w) of all v_{r-c} and wrd_{co-d} variables from the learning step. w_k and the bias obtained from the SVM learning are used to determine the $v_{r-c} wrd_{co-d}$ pair with the IT-Task, NW-Task, Mix-Task, and Mis-Task classes by Equation (3) based on One-vs-All. If $f(x) \geq 0$, the $v_{r-c} wrd_{co-d}$ pair with a positive class (ISTConceptClass), otherwise the negative class (nonISTConceptClass) occurs.

MLPs: The weight w from learning the following classes, IT-Task, NW-Task, Mix-Task,

and Mis-Task, from the previous step is used to determine the salient IST-concept classes by equation (4) from the wrdVCo expressions (v_{r-c} wrd_{co-d} pairs) on the salient IST-concept EDUs of the test samples (*ticket_{jt}.Request-Detail.EDU_{ct}*. wrdVCo, including TID).

Results and Discussion

Three evaluations of the proposed research were conducted by three expert judgments with max win voting: the first evaluation is the IST-concept EDU determination: we determined the IST-concept EDUs from the test sample by Vr and WVC, where the test sample of 390 tickets contained 995 EDUs of the ticket_{jt}.Request-Detail feature values having 923 IST-concept EDUs. The second evaluation is mining or extracting the salient IST-concept EDUs: we could extract the salient IST-concept EDUs of the 1100 tickets with the correct extraction of 1060 salient IST-concept EDUs on the ticket_{jt}.

Request-Detail feature values by using the rule (based on the corpus behavior study). The third evaluation is the salient IST-concept classification from the test sample by SVM and MLP. The three evaluations are based on precision and the recall (see Table 1).

Since most of the documents on the *ticke_{jt}.Request-Detail* feature values ($j=1,2,...,1700$), which are short notes recorded by help-desk staff based on calls from end-users, words are omitted until only the verb remains (after stemming word and stop word removal) causing low recall in the IST-concept EDU determination as shown in Example 2 (the [...] symbol means an ellipsis).

Other than the word ellipses, there are some *ticke_{jt}.Request-Detail* feature values containing two salient IST-concept expressions on two EDUs, e.g., EDU1 and EDU2 in Example3, which results in the low recall of the salient IST-concept extraction (or the salient IST-concept EDU extraction)

Table 1

The accuracy of extracting the IST-concept EDUs and the salient IST-concept EDUs, and classifying the salient IST-concept EDUs

Evaluation	Precision	Recall	F1-Score
IST-Concept EDU Determination	0.985	0.941	0.962
Salient IST-Concept EDU Extraction	0.980	0.960	0.970
Salient IST-Concept Classification by SVM	0.896	0.859	0.877
Salient IST-Concept Classification by MLP	0.940	0.856	0.896

Example 2 EDU1: “[((ผู้ใช้/user)/Noun)/NP1] ((เปิด/open)/Verb_{strong} (PDF [ไฟล์/file])/NP2)/VP”
 (“[The user] open a PDF [file].”)

EDU2: “(แล้ว/then)/conj [(PDF ไฟล์/PDF file)/NP1] (ตั้ง/pop up)/Verb_{strong}”
 (“[the file] is then popped up.”)

Example 3 EDU1: “[ผู้ใช้] ขอเปลี่ยนเมาส์/[The user] requests to change the mouse.”

EDU2: “แถบกูเกิลข้างบนหายหมด/The Google toolbar has completely disappeared.”

EDU3: “[ผู้ใช้] ค้นหาอะไรไม่ได้/[The user] can’t search for anything.”

One reason for the low recall in salient IST-concept classification by both SVM and MLP is the non-mutually exclusive characteristic occurrence, e.g., the software installation task without mentioning the particular software—for example EDU: “ติดตั้งซอฟต์แวร์เพื่อใช้ในการประชุม/Install software for use in the meeting.”. Some software installation events, such as Windows, belong to the IT-Task class, while others, such as WordNet, belong to the Mis-Task class. According to Diwanji et al. (2020), their F1-score results of classifying or identifying tasks or requests from the email corpus and the chat-message corpus based

on parsing and VerbNet in RNN classification were 0.80 and 0.89 respectively which rely on several sentences’ verb phrases per instance, which is time consuming whereas our method is based on one salient IST-concept EDU per instance/ticket. Since MLP achieves a higher F1-score in salient IST-concept classification than SVM, the MLP results of the salient IST-concept classification are used for the daily task load representation as shown in Figure 3.

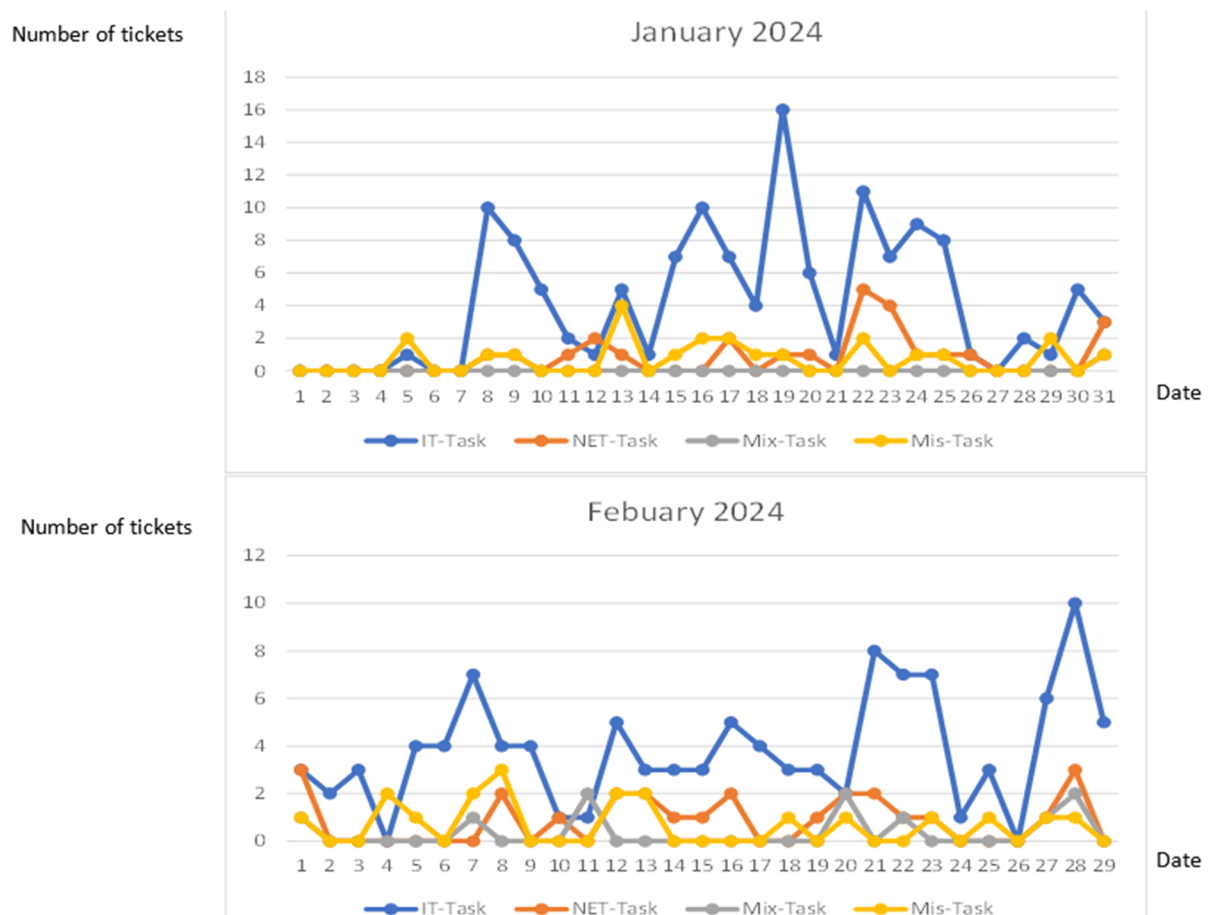


Figure 3 The daily task loads for IT-Task, Net-Task, Mix-Task, and Mis-Task from January through February 2024. The average task times of the Task-Time attribute per ticket during this period were 68.66 minutes, 66.61 minutes, 37.06 minutes, and 37.34 minutes, respectively. IT-Task shows the highest load during this period, while Mix-Task has the lowest load.

Conclusions

This research applied word-co expression ($wrdVCo$) based on v_r and wrd_{co} to mine and extract the salient IST-concept EDUs to represent the salient IST concepts of the corresponding Request-Detail feature values in the Ticket entity. The extracted salient IST-concept EDUs were classified by SVM and MLP under supervised learning for the graphical representation of the daily task loads. The limitations of our salient IST-concept extraction are as follows: (1) one salient IST concept must be expressed within a single EDU occurrence, and (2) $v_r wrd_{co}$ represents $wrdVCo$ after determining the relatedness value between two sequent words within EDU(s) of each Request-Detail feature value where wrd_{co} serves as a head noun of NP1; and if NP1 ellipsis occurs, wrd_{co} is the word immediately following v_r . The salient IST-concept classification of the Request-Detail feature value for each ticket entry relies on 2-3 words of the $wrdVCo$ expression as the element of WVC. In contrast, the previous researches—such as Lin et al. (2018), which used Bag-Of-Words and RNN to classify emails, and Diwanji et al. (2020),

which relied on Lin et al. (2018) and used verb phrases and RNN to classify tasks from emails and chat messages—shows that both BoW and verb phrases affected the time it took to determine or classify each task class, even though their F1-scores are comparable to ours. Moreover, our graphical daily task-load representation can be applied to resource optimization and enable better decision-making in scheduling and workload balancing. With regard to Hee (2018) and Pechsiri et al. (2025), our methodology can be applied to several languages, particularly in Asia, e.g., Lao and Chinese, which have similar characteristics to the Thai language, e.g., noun phrase ellipsis and a lack of explicit sentence or word delimiters which affect computational tasks to handle the textual data effectively.

Acknowledgements

The authors would like to acknowledge Mr. Taweesak Cha-nah for providing the information on solving IT-support requests.



References

- Al-Hawari, F., & Barham, H. (2021). A machine learning based help desk system for IT service management. *Journal of King Saud University-Computer and Information Sciences*, 33(6), 702-718. <https://doi.org/10.1016/j.jksuci.2019.04.001>
- Aun-a-nan, A., & Meesad, P. (2020). The classification of credibility of Thai news source websites using data mining techniques [In Thai]. *EAU Heritage Journal Science and Technology (online)*, 14(2), 101–116. <https://he01.tci-thaijo.org/index.php/EAUHJSci/article/view/241729>
- Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2016). *WEKA manual for version 3-9-1*. University of Waikato.
- Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In J. van Kuppevelt & R. W. Smith (Eds.), *Current directions in discourse and dialogue* (pp. 85–112). Springer.

- Chanlekha, H., & Kawtrakul, A. (2004). Thai named entity extraction by incorporating maximum entropy model with simple heuristic information. In K. Su, J. Tsujii, J. Lee & O. Y. Kwong (Eds.), *Proceedings of the First International Joint Conference on Natural Language Processing* (pp. 1–7). Springer-Verlag.
- Chareonsuk, J., Sukvakree, T., & Kawtrakul, A. (2005). Elementary discourse unit segmentation for Thai using discourse cue and syntactic information [In Thai]. In International Annual Symposium on Computational Science and Engineering (Ed.), *The 9th Annual National Symposium on Computational Science and Engineering (ANSCSE 9)* (pp. 85–90). ANSCSE.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. Cambridge University Press.
- Diwanji, P., Guo, H., Singh, M. P., & Kalia, A. (2020). Lin: Unsupervised extraction of tasks from textual communication. In D. Scott, N. Nuria & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 1815-1819). International Committee on Computational Linguistics
- Guthrie, J. A., Guthrie, L., Aidinejad, H., & Wilks, Y. (1991). Subject-dependent co-occurrence and word sense disambiguation. In D. E. Appelt (Ed.), *29th Annual Meeting of The Association for Computational Linguistics* (pp. 146-152). The Association for Computational Linguistics.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10-18.
<https://doi.org/10.1145/1656274.1656278>
- Hassan, S., & Mihalcea, R. (2011). Semantic relatedness using salient semantic analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 25(1), 884–889.
<https://doi.org/10.1609/aaai.v25i1.7971>
- Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall.
- Hee, Y. E. (2018). A comparative study of subject ellipsis in Chinese and Korean sentence. *International Journal of Humanities and Social Science Invention (IJHSSI)*, 7(4), 11-21.
[https://www.ijhssi.org/papers/vol7\(4\)/Version-1/A0704010110.pdf](https://www.ijhssi.org/papers/vol7(4)/Version-1/A0704010110.pdf)
- Huang, Y., Yu, L., & Lin, I. C. (2013). Text categorization for service request classification. *International Journal of Signal Processing Systems*, 1(1), 54-58. <https://doi.org/10.12720/ijsp.1.1.54-58>
- Ketui, N., Theeramunkong, T., & Onsuwan, C. (2013). Thai elementary discourse unit analysis and syntactic-based segmentation. *International Information Institute (Tokyo) Information*, 16(10), 7423-7436. <https://bit.ly/4wlKBfN>
- Lin, C. C., Kang, D., Gamon, M., & Pantel, P. (2018). Actionable email intent modeling with reparametrized RNNs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 4856-4864.
<https://doi.org/10.1609/aaai.v32i1.11931>
- Miller, G. A. (1995). WordNet: A lexical database. *Communications of the ACM*, 38(11), 39-41
<https://doi.org/10.1145/219717.219748>

- Pakamwang, J., Khoomsab, K., & Timsorn, K. (2020). Investigation of factors for student's decision in studying at Phetchabun Rajabhat University by data mining technique [In Thai]. *EAU Heritage Journal Science and Technology (online)*, 14(1), 24–33. <https://he01.tci-thaijo.org/index.php/EAUHJSci/article/view/240508>
- Paramesh, S. P., Ramya, C., & Shreedhara, K. S. (2018). Classifying the unstructured IT service desk tickets using ensemble of classifiers. In M. N. Vijayalakshmi (Ed.), *The 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS 2018)* (pp. 221-227). IEEE.
- Pawar, S., Palshikar, G., & Banerjee, A. S. (2021). Weakly supervised extraction of tasks from text. In S. Bandyopadhyay, S. L. Devi & P. Bhattacharyya (Eds.), *Proceedings of the 18th International Conference on Natural Language Processing (ICON)* (pp. 583-592). NLP Association of India (NLPAl).
- Pechsiri, C., & Piriyaikul, R. (2016). Extraction of a group-pair relation: Problem-solving relation from web-board documents. *SpringerPlus*, 5, 1265. <https://doi.org/10.1186/s40064-016-2864-3>
- Pechsiri, C., Piriyaikul, I., & Pechsiri, J. S. (2025). Grouped semantic-feature relation extraction from texts to represent medicinal-plant property knowledge on social media. *Frontiers in Artificial Intelligence*, 8, 1579357. <https://doi.org/10.3389/frai.2025.1579357>
- Sudprasert, S., & Kawtrakul, A. (2003). Thai word segmentation based on global and local unsupervised learning [In Thai]. In National Computer Science and Engineering Department (Ed.), *The 7th National Computer Science and Engineering Conference* (pp. 1–8). National Computer Science and Engineering Department.
- Sutta, S., Jivnot, T., & Settawatcharawanit, T. (2025). Development of an inventory management system on a mobile application using machine learning technology [In Thai]. *EAU Heritage Journal Science and Technology (online)*, 19(3), 76–90. <https://he01.tci-thaijo.org/index.php/EAUHJSci/article/view/277291>
- Thimsanthia, R., & Soemphol, C. (2025). Design and development of a video-based work time analysis system for industrial production lines [In Thai]. *EAU Heritage Journal Science and Technology (online)*, 19(3), 212–226. <https://he01.tci-thaijo.org/index.php/EAUHJSci/article/view/279822>
- Tongtep, N., & Theeramunkong, T. (2015). Pattern-based extraction of named entities in Thai news documents. *Science & Technology Asia*, 15(1), 70–81. <https://ph02.tci-thaijo.org/index.php/SciTechAsia/article/view/41311>
- Whiteley, E. K. (2024). Order-based salience patterns in language: What they are and why they matter. *Ergo: An Open Access Journal of Philosophy*, 11(26), 689–715. <https://doi.org/10.3998/ergo.6166>

