ISSN: 2302-9285, DOI: 10.11591/eei.v10i1.2031

Requirements identification for distributed agile team communication using high level carotene

Nor Hidayah Zainal Abidin, Pathiah Abdul Samat

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia

Article Info

Article history:

Received Aug 4, 2019 Revised Dec 11, 2019 Accepted Jun 18, 2020

Keywords:

Carotene
Distributed agile
Requirement identification
Text classification
Text messaging

ABSTRACT

Communication plays an important role to deliver the correct information. However, the communication became challenging especially for agile software teams, which are in geographical distributed. The problem arise when there are exchanging information using unstructured communication platform, misunderstanding on the information communicated and lack of documentation. The aim of this study is to propose a text classification technique for requirements identification in text messages. In this study, we adopted the cascade and cluster classification concept of Carotene that relies on the hash tag function. It classifies the text messages into requirements types instead of job title. This technique called as high-level carotene (HLC) technique that embedded into the tool to identify the functional requirement and non-functional requirements. The result shows that most of criterias evaluated have achieved more than 85% of effectiveness in identifying both of requirement in text messaging by using this technique.

This is an open access article under the CC BY-SA license.



249

Corresponding Author:

Pathiah Abdul Samat, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia. Email: pathiah@upm.edu.my

1. INTRODUCTION

The replacement of manual work through technologies made the software industries received tremendous job to develop applications that suits to the organizations need. The evolution of demand for quick deliverables requires the businesses to keep changing its direction which led to the emergence of agile development methods. Instead of trying to identify a complete set of requirements before the project begins and eliminate changes, agile methods embrace change by accepting and incorporating it. Agile software development refers to a software development methodologies based on iterative development, where requirements and solutions evolve through collaboration between self-organizing cross-functional teams. The organizations that adopt agile methodology can have team that are co-located and distributed across the oceans. The co-located agile team works in the same location within the same room, besides one another and can have face to face interaction while communication. Meanwhile, the distributed agile team dispersed in cubicles or in separate offices or in different floors, or geographically distributed across the world. In terms of communication, many distributed agile team uses messaging tool to discuss the product requirements or any project information. This method produce large number of data in the messaging features. Since project is tied to the timeline and customer requires fast deliverables, the development team need a suitable tool. By classifying the text messages sentences and stored it into structured board will help them to have an effective platform as reference, better focus and save time. In industry, there are numerous number of text classification techniques such as Carotene and Convolutional Network (ConvNet).

Journal homepage: http://beei.org

250 🗖 ISSN: 2302-9285

Although various technologies offered for communication across the globe, none prepare one-centralized platform that can identify and categorize the specific requirement communication. In agile, the different communication platform will cause ineffective of information exchange [1, 2]. However, when the distributed team communicates through messaging, the data will be unstructured. Keeping track of huge amount of message text nature is challenging and time-consuming to understand and organizing its data. Most chatter failed to extract its value. For example, the team that need to refer back to the data have to scroll over the text messages interface to find the main point of the requirements types information may gets lost its objective. Other than that, the global team background with different first language might misunderstood the discussed items due to invisibility of body language and un-conveyed information [3, 4]. The message's real meaning gets lost in conversation through the medium. The issues mentioned above are applicable to the present text messages tool. Therefore, the agile team require an invention of text messages tool that function as organizer that provide the service to identify and organize the text messages of requirements. By using text classifiers function, the chatter able to structure chat conversations, save time when analyzing text data and help identify the requirements types.

However, the challenges in communication are distance, different works hour, communication methods, knowledge transition, knowledge management, language barrier and role of specialist. Distance transforms most communications into electronic communication platform of variety choices [5, 6]. There are number of reseachers [7, 8] reveal that it is more challenges when the scale of the projects become larger. Team members need to align their working hours to get clear clarification to improve productivity [9, 10]. They also found that team with different time zones introduced additional challenges when global organizations started to use agile [11, 12]. Ineffective communication method used for information exchange make it difficult to meet customer's expectation which can only be solved through face to face meeting [13]. Sometimes due to the quality of the sound and video and lack of the visibility on the body language contributed to the challenges [14]. Effective knowledge sharing, methods, decision and skills must be gathered during the development process to reduce redundant work and cost [15]. Different language cause mismatching language, the misinterpretation of the terminologies used and un-conveyed information [16-18]. The software organizations expected to utilize the specialist. Agile methodology does not have mechanism to request such expertise. Even the specialists may face problems similar to a new team member about gaining the understanding of the project requirements [19].

There are several related solutions such as existing communication tool and identification techniques as well as available text classification. It is recommended to set video conferencing meeting using Skype as it will saves time, expenses and flexible for project members [20]. The examples of effective communication tool are internet messaging (IM). Thus, this issue can avoided by using standard synchronous communication tools for each project members [21]. Others researcher proposed novel Scrum framework methodology that guide the Scrum Master to effectively coordinate and communicate to the distributed agile teams [22]. Another studies proposed corrective actions to the identified communication waste in terms of lack of involvement, shared understanding, outdated and restricted access and scattered information [23]. Text classification technique using ConvNets provided an empirical research and solution on the use of character-level convolutional networks for text classification to avoid misunderstanding between communicating participants [20]. Other than that, Carotene text classification is a semi-supervised classification for job title and taxonomy discovery system [15]. From the job datasets, the taxonomy discovery component discovers job titles by utilizing clustering techniques, it has combination of cascade architecture and clustering component that relies on a training dataset. In addition, the text sentences also can be classified using the combination of the classification and clustering for tweets Sentiment Analysis [24, 25]. The messages of text are classify either as positive or negative using algorithm where the similar instances from the same cluster, which assumed will be share the same class label.

In this study, we propose a technique for text classification function for requirements identification in text messages for agile global software development team communication platform, design the solution and evaluate its' effectiveness. Our technique is focus on identifying the requirements by text messages classification software requirements types of functional or non-functional requirements specifically for software engineers with experience working in distributed agile team environment. This paper is organize as follows: The following section discuss the research method. Section 3 will discuss detail our result and analysis. Finally, the conclusions and directions for future research are discuss in section 4.

2. RESEARCH METHOD

In this paper, we study high level carotene text classification concept that does automatic cascade and cluster text classification as our proposed solution of text classification function. This concept relies on training dataset. For this study, we adopted the cascade and cluster classification concept but rely on the hash

tag function. This technique called as high-level carotene (HLC) technique that embedded into the AgileCom tool. The user can use its service to identify requirements and organize it into functional requirement (FR) and non funtional requirement (NFR). They also can perform the hash tag function to initiate the text sentences classification process. Once the sentence submitted, the cascade classification split the sentences into words to find the hash tag function. During this process, it tries to detect the hash tag function in the text sentences to identify the requirements. After it detected, the cluster classification will grouped back the splitted words into the sentences. The cluster classification organized the identified requirements by storing them into their respective clusters before it is display for user's view.

ISSN: 2302-9285

2.1. The design

In design phase, we perform the process of embedding the HLC classification concept into the AgileCom tool. The cascade and cluster classification implemented at the backend for automatic classification function and the hash tag function labeled at the front end. This technique also known as *topic labeling*, that is understanding what is the given text is talking about. The combination of cascade and cluster classification used for structuring data by its topic or subject labeling. Figure 1 shows the flow of the classification technique for the text messages in the AgileCom tool.

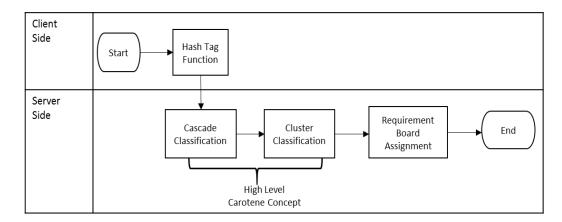


Figure 1. The workflow of the HLC classification technique in the AgileCom tool

As shown in Figure 1, user start activities by using the hash tag function. All the text messages, automatically classify vertically in the array list during backend processing through cascade classification. The identified same hash tag function in the array list in the cascade classification later automatically grouped into the same cluster using cluster classification. Any text sentences that not contain the hash tag function will automatically eliminated. The hashtag function takes the following code as shown below:

'The hashtag symbol" the first letter of the 2 Requirement Types word'

The purpose of code above is to label the text messages sentence into topic function. The classified messages that available in the cluster assigned to the requirements board as functional requirement or non functional requirement. The assigned classified requirements displayed in the requirement board for user references. The codes of classification take the following as shown below:

a) Cascade classification: Split (string \$pattern): array

b) Cluster classification:

Word wrap (string \$str) + Match (string \$pattern)

The code of cascade classification taken from the existing PHP *preg_split*. The purpose of cascade classification is to split the sentence into word of cascase of array to identify the word that contain hashtag function by using regular expression pattern. Whereas the code of cluster classification taken from the existing PHP *wordwrap* and *preg_match*. The purpose of cluster classification is to find the match of the hashtag function and wrap back the splitted word into sentences using regular expression pattern.

Figure 2 shows the activity diagram of the AgileCom tool. It starts with the user to login into the system and type the text messages in the messaging room. The user can either type normal text messages or use hash tag function in messaging room. Based on Figure 2, if the user chose to use hash tag function, the user can view the classified text messages in the requirements board. For the login alternative, user can directly go to requirements board to view requirements as reference. Lastly, they can logout from the system.

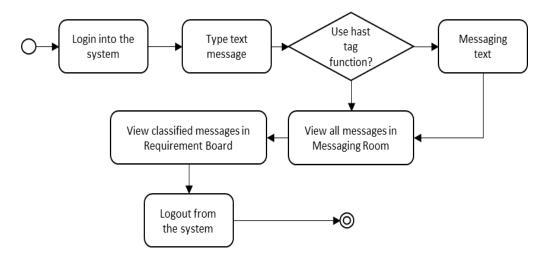


Figure 2. The activity diagram for the AgileCom tool

2.2. The implementation

We develop AgileCom as a web based application, use HTML and JavaScript for front-end scripting, PHP programming language as server side scripting and PHPMyAdmin as database that uses MySQL scripting. Figure 3 shows the interface for chatting in the messaging room. This process can done by any members in the distributed Agile Team project. Based on Figure 3, business analyst (BA) login into the messaging room and typing the messages to communicate. The user can use the '#FR' or '#NFR' hash tag function to input the message. Once user input the messages the automatic cascade perform the classification process at server side to find the match for hash tag function in the sentences. The sentence split into an array list of word to detect the '#FR' words vertically. Figure 4 shows the process of cluster classification at server side.



Figure 3. The textbox message of messaging room

```
#FR The system shall allow the user to login either as student or supervisor. Both had different access once login.
#FR login features need to be in black color
For the #FR, the system shall be able to allow the user to chat in the Messaging Room.
#FR The system shall be able to display the classified messages in the Requirements Board.
#FR The system shall be able to display the classified messages in the Requirements Board.
#FR The requirements board need to be divided int o: 1 for FR, another 1 for NFR
```

Figure 4. Automatic cluster classification process at server side based on the hash tag

As shown in Figure 4, after the hash tag function of '#FR' identified, it wrapped back the split word into the sentences and grouped it in the same cluster. In this case, '#FR' is classify into its cluster of #FR. Figure 5 shows the requirements board interface. Based on Figure 5, the requirement board displays the classified text sentences of functional and non-functional requirement. Figure 6 shows the algorithm of to the HLC technique which is embedded into the AgileCom tool.

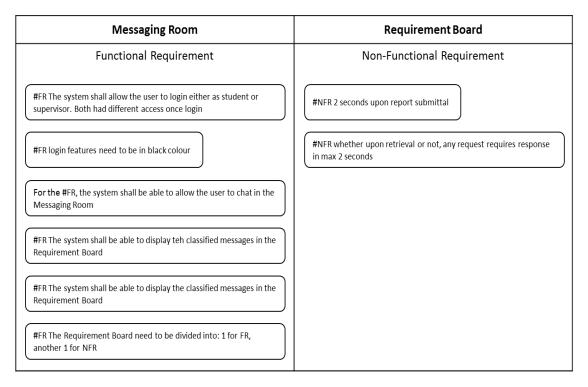


Figure 5. The requirements board interface.

Based on Figure 6, step 1 describes the user "Start" by login into the system and procees to step 2, which is 'Output "Messages Text Box in Messaging Room" by displaying the textbox for chatting in the Messaging Room. In step 3, the "Input String" allow user to input text messages. In the step 4, the "Input hash tag function for software requirements types sentences." accept the user input for hash tag function. Step 5 shows the classification of text messages by HLC technique based on the hash tag function. In step 6, the tool will perform "Output HLC Text Classification" to display the classified software requirements types. Lastly, the users can "Stop" by logout from the system.

```
Step 1: Start.

Step 2: Output "Messages Text Box in Messaging Room".

Step 3: Input string.

Step 4: Input hash tag function for software requirements types sentences.

Step 5: HLC Text Classification ← hash tag function.

Step 6: Output HLC Text Classification.

Step 7: Stop.
```

Figure 6. The algorithm of HLC technique

3. RESULTS AND DISCUSSION

In this study, we distributed survey to seven respondents through email and mostly using instant messaging applications such as WhatsApp to the targeted respondent. Figure 7 shows the background of repondents. As shown in Figure 7, the total of 57.1 % of the respondents work as Software Developer and the remaining of 14.3% work as Project manager, Business Analyst and Other types of position related to Software Engineering field. The user background achieved our target because we focus on the respondents who work in software development team. For example, to adopt the real environment, we require higher number of developers only 1 business analyst and 1 project manager in a project team.

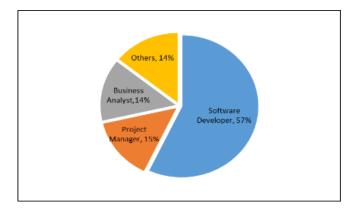


Figure 7. The evaluation result on respondents background

To ensure that our study is meaningfull, we provide a questionnaire "Do you have work experience in Distributed Agile Team". The objective of this questionnaire is to ensure only those who are experience in Agile environment and distributed geographically understand the constraint and challenges in distributed communication team. The result shows that all of seven respondents have experience working in distributed agile methodology team. In term of years the working experience in distributed agile team, five of respondents or 71.4% have 4 to 7 years experience whereas only two of respondents have 1 to 3 years experience. Most of seven respondents using messaging tool in their daily work and have experience facing communication issues with their team members that distributed at different places. However, six of respondents agreed that they have face difficulties to communicate with other team in different location while one disagreed. The result also shows that approximate 85.7% or 6 respondents agreed that the classification of message data into FR and NFR would be able to avoid requirement misunderstanding during messages communication while one respondent disagreed. Table 1 shows the evaluation results summary of the effectiveness of the technique and tool. The purpose of this evaluation is to obtain valuable feedback of our design, method and implementation.

Table 1. The Summary of the effectiveness of HLC technique in the agilecom tool

No.	Statement	Yes	Fair	No
1.	Awareness on the importance of requirements types classification in software development process.	100%	0%	0%
2.	The effectiveness of the AgileCom tool and the HLC technique able to expedite the understanding of the software requirements and reduce requirements freeze duration.	100%	0%	0%
3.	The effectiveness of the AgileCom tool and the HLC technique support the text sentences classification process.	100%	0%	0%
4.	The effectiveness of the Requirements Board able to improve requirements management.	100%	0%	0%
5.	The effectiveness of the AgileCom tool and the HLC technique able to maximize the value delivered during sprint.	85.7%	14.3%	0%
6.	The effectiveness of the AgileCom tool and the HLC technique used in improvements for communication and requirement management for global agile software development team.	85.7%	14.3%	0%

The questions in Table 1 is to measure the effectiveness of the HLC technique embedded in the AgileCom tool. From the result of question 1 until 4, 100% of the user agrees that they are aware on the importance of requirements types' classification during the software development process. This is because they have to communicate to each other to clarify the requirements. They usually used messages application to

chat about it. However, the size of text messages data they sent during chatting with one another is huge. It is difficult for them to find the FR and NFR sentences when they need to refer to the information again. They have to scroll the interface multiple times to allocate the data. With this technique, it is able to eliminate this problem and improve the team's productivity. Apart from that, the technique also helps them to expedite the understanding of their communicated software requirements type's data easily because it identifies and classifies the text messages into FR and NFR. This helps to discard the requirements freeze that allow them proceed to development phase faster. They also agree that it support proper text sentences classification process since it is able to to identify and categorize the raw text messages into FR and NFR in the chat conversation. The categorized text sentences of FR and NFR value stored at the Requirements Board able to help organizing their requirements data. For question 5, approximate 85.7% or six users agrees that this technique allow the agile team to move to development phase faster, maximize the value delivered during sprints because they manage to meet the timeline but able to produce a quality product. Based on question 6, approximate 85.7% or six users also agree it is able to improve their communication management because working in global agile team is difficult and time consuming. By having one centralized communication platform with focus of requirements identification of FR and NFR, they had less worry on the messiness of their unstructured way of communication.

Based on our pre survey, approximate 85.7% or six of respondents face difficulties to communicate with other team in different location. All of respondents also agreed having experience in wrong software requirements delivered during messaging communication. In order to eliminate the above problem, we evaluate the suitability of communication using HLC technique in the Agilecom Tool. We also evaluate the successful of this technique in classifying FR and NFR to avoid wrong software requirement. Table 2 shows the summary of suitability evaluation. Based on Table 2, all of the respondents agreed that Agilecom tool is suitable practice by distributed agile team in different location. The result also shows that six of the respondent or 85.7% are strongly agreed the technique and tool successful in delivering the right software requirements during messages once they obtained it from the Product Owner. This is to ensure that they can plan the sprint correctly and reduce requirements freeze duration. However, they did not focus on the communication management due to the development of the product is more important for them to achieve the timeline for product release. In overall, we conclude that hash tag and HLC technique able to answer to the problem highlighted.

Table 2. The summar	v the suitability and	d successful of HLC	technique and	l agilecom tool

No.	Question	Strongly Agree	Agree	Fair	Disagree	Strongly Disagree
1.	The AgileCom and the technique successful in support the exchanging the classified messages into FR and NFR to avoid wrong software requirement.	85.7%	14.3%	0%	0%	0%
2.	The technique of HLC and Agilecom tool is suitable can be practiced by Distributed Agile Team in different location.	100%				

4. CONCLUSION

In this study, we proposed the HLC technique that does automatic cascade and clustering text classification based on the hash tag function to identify requirements types during text messages communication. Due to the classification of text messages to the FR and NFR, it is able to reduce the misunderstanding in communicating the types of requirements between the distributed agile team members. Although there are many other available communication platform as mentioned by the user like Trello and WhatsApp but only AgileCom have the technique that solely focus on classifying the requirements types into FR and NFR. As we believe, the user also agrees this technique can overcome the problem highlighted earlier. The users now have structured text messages tool that function as organizer that provide service to identify and organize text messages. This process contributed uniquely only for agile team usage, which makes it new in agile, team communication management area. The algorithm and system design stated in section 2 describe the novel technical contribution to this work. By using web-based application, the requirement is accessible at any time with different location. The technique also able to improve their communication and requirements management that crucial to deliver quality output and increase team productivity, where six respondents (85.7%) agreed.

ACKNOWLEDGEMENTS

This research sponsored by Ministry of Education Malaysia (MOE) and Universiti Putra Malaysia (UPM) via the Fundamental Research Grant Scheme (FRGS, Vot No: 5524956)

REFERENCES

- [1] S. Y. Chadli, A. Idri, J. L. Fernández-Alemán, J. N. Ros and A. Toval, "Identifying risks of software project management in Global Software Development: An integrative framework," 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, pp. 1-7, 2016.
- [2] Saad, Y. C., and Ali, F. "Identifying and Mitigating risks of software project management in global software development," *In Proceeding of the 27th International Workshop on Software Measurement and 12th International Conference on Software Process and Product Measurement*, pp. 12-22, 2019.
- [3] Niazi, M., et al., "Toward successful project management in global software development," *International Journal of Project Management*, vol. 34, no. 8, pp. 1553-1567, Nov 2016.
- [4] Krunal, B., et al., "Scrum: An Agile Process Reengineering In Software Engineering," International Journal of Innovative and Exploring Engineering, vol. 9, no. 3, pp. 840-848, 2020.
- [5] Esquivel, S. "Communication Issues in Agile Software Development," In XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016), pp. 275-484, 2016.
- [6] Paul, T. R. "Communication Network in an Agile Distributed Software Development Team," *In Proceeding of the 14th ACM/IEEE International Conferece on Global Software Engineering*, pp. 100-104, 2019.
- [7] Papatheocharous, E., and Andreou, A. S. "Empirical evidence and state of practice of software agile teams," *Journal of Software: Evolution and Process*, vol. 26, no. 9, pp. 855-866, 2014.
- [8] Khalil, C., and Khalil, S. "A governance framework for adopting Agile methodologies," *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 6, no. 2, pp. 111-119, 2016.
- [9] Kaur, H., et al., "Distributed Agile Development: A Survey of Challenges and Solutions," In Proceedings of the International Conference on Software Engineering Research and Practice (SERP), pp. 42-48, 2015.
- [10] Alzoubi, Y. I., et al., "A measurement model to analyse the effect of agile enterprise architecture on geographically distributed agile development," Journal of Software Engineering Research and Development, vol. 6, no. 4, pp. 1-24, 2018.
- [11] Dikert, K., et al., "Challenges and success factors for large-scale agile transformations: A systematic literature review," *Journal of Systems and Software*, vol. 119, pp. 87-108, Sep 2016.
- [12] Hoda, R., and Murugesan, L. K. "Multi-level agile project management challenges: A self-organizing team perspective," *Journal of Systems and Software*, vol. 117, pp. 245-257, July 2016.
- [13] Nguyen-Duc, A., et al., "The Impact of Global Dispersion on Coordination, Team Performance and Software Quality-A Systematic Literature Review," *Journal of Information and Software Technology*, vol. 57, pp. 277-294, Jan 2015.
- [14] Embretsen, D., and Hyder, L. "Scrum in Global Software Development: An Ethnographic Case Study of Scrum's Mitigation Effects on Global Software Development Challenges," *Master Thesis, Management, Communication and IT, Uppsala University*, pp. 1-55, 2017.
- [15] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao and T. S. Kang, "Carotene: A Job Title Classification System for the Online Recruitment Domain," 2015 IEEE First International Conference on Big Data Computing Service and Applications, Redwood City, CA, pp. 286-293, 2015.
- [16] Zahedi, M., et al., "A systematic review of knowledge sharing challenges and practices in global software development," *International Journal of Information Management*, vol.36, no.6, part A, pp. 995-1019, 2016.
- [17] Qureshi, R., et al., "Novel Framework to Improve Communication and Coordination among Distributed Agile Teams," *International Journal of Information Engineering and Electronic Business*, vol. 10, no. 4, pp.16-24, 2018.
- [18] Constantin, H., et al., "Development of an Kanban for distributed part-time teams and an introduction framework," In Proceeding of the 8th Conference on Learning Factories 2018-Advanced Engineering Education and Training for Manufacturing Innovation, pp. 45-50, 2018.
- [19] Coletta, L. F. S., *et al.*, "Combining classification and clustering for tweet sentiment analysis," In 2014 Brazilian Conference on Intelligent Systems, pp. 210-215. 2014.
- [20] Alzoubi, Y. I., et al., "Distributed Agile Development Communication: An Agile Architecture Driven Framework," Journal of Software, vol. 10, no. 6, pp. 681-694, 2015.
- [21] Pathiah, A.S., and Mohamed, Z.M.N., "SMMS: Document Management in agile model for software maintenance," *International Journal of Advanced Science and Technology*, vol. 28, no. 2, pp. 321-328, 2019.
- [22] Areej Al-Zaidi and Rizwan Qureshi. "Global software development geographical distance communication challenges," *The International Arab Journal of Information Technology*. vol. 14, no. 2, pp. 215-222, 2017.
- [23] Zhang, X., et al., "Character-level convolutional networks for text classification," In Advances in neural information processing systems, pp. 1-9, 2015.
- [24] Voigt, S., et al., "A study of documentation in agile software projects," In Proceedings of the 10th ACM/IEEE international symposium on empirical software engineering and measurement, no. 4, pp. 1-6, 2016.
- [25] Shagufta, S., Voigt, S., et al., "A study of documentation in agile software projects," In Proceedings "Towards Scrum Based Agile Framework for Global Software Development Teams," Journal of Engneering & Technology, vol. 38, no. 4, pp. 979-998, 2019.

BIOGRAPHIES OF AUTHORS



Nor Hidayah Zainal Abidin is a master student in Department of Software Engineering and Information System, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia.



Pathiah Abdul Samat is a senior lecturer at Department of Software Engineering and Information System of Faculty Computer Science and Information Technology, Universiti Putra Malaysia. She has received her PhD from Universiti Kebangsaan Malaysia. Her focus research area is in software modeling and software verification.