UNIVERSITI PUTRA MALAYSIA

REQUIREMENTS IDENTIFICATION USING HIGH LEVEL CAROTENE TEXT CLASSIFICATION CONCEPT FOR DISTRIBUTED AGILE TEAM COMMUNICATION

NOR HIDAYAH ZAINAL ABDIN

FSKTM 2019 21

REQUIREMENTS IDENTIFICATION USING HIGH LEVEL CAROTENE TEXT CLASSIFICATION CONCEPT FOR DISTRIBUTED AGILE TEAM COMMUNICATION

By

NOR HIDAYAH BINTI ZAINAL ABIDIN

Thesis submitted to the School of Graduate Studies
Universiti Putra Malaysia
in fulfilment of the requirements for the
Master of Software Engineering

JUNE 2019

DEDICATIONS

In the name of Allah S.W.T, the most merciful, the most compassionate all praise to Allah, the lord of this world and peace be upon Muhammad S.A.W his servant and messenger.

This thesis is wholeheartedly dedicated to my husband who has been my source of inspiration, and gave me strength when I felt to giving up and also he who continually provides the warmth of love, a lot of encouragement and financial support.

To my parents, family, family in laws, friends and mentors who shared strength, spirit, encouragement and motivation to complete this work.

Abstract of thesis presented to the University Putra Malaysia in fulfilment of the requirement for the Master of Software Engineering

By

NOR HIDAYAH BINTI ZAINAL ABIDIN

JUNE 2019

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

Communication plays an important role to deliver the correct information. It is one of the biggest issues frequently happened in distributed agile team environment especially in text messaging. Among the issues are exchanging information using unstructured communication platform, misunderstanding on the information communicated and lack of proper documentation. If they require references on the requirements, they need to scroll back to their communicated text messages to identify it. To solve these issues, we studied the High Level Carotene concept that does automatic cascade and cluster text classification. 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 is called as High Level Carotene (HLC) technique that embedded into the tool that can identify the requirements information by classifying the text messages. The cascade classification split the sentences into words and cluster classification grouped it into the same cluster. For the objectives of this study, we identify the text classification technique for requirements identification in text messages, design the solution and evaluate the technique. In this study, there are related issues and solution for agile team communication and brief description of the HLC technique design and implementation for messages text classification function in the Agile Communication Management (AgileCom) tool. Lastly, this study concluded by the HLC technique effectiveness and usability evaluation results.

Abstrak tesis yang dikemukakan kepada Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Sarjana Kejuruteraan Perisian

Oleh

NOR HIDAYAH BINTI ZAINAL ABIDIN

FAKULTI SAINS KOMPUTER DAN TEKNOLOGI MAKLUMAT

Komunikasi memainkan peranan penting dalam menyampaikan maklumat yang betul. Tetapi di dalam kumpulan Agile yang bekerjasama dari kawasan berlainan, ianya merupakan salah satu masalah besar terutama jika ia melibatkan pengunaan mesej teks. Antara masalah-masalah yang timbul adalah menggunakan pelbagai alat komunikasi untuk pertukaran mesej, salah faham dengan mesej yang disampaikan dan ketiadaan dokumen yang diperlukan. Jika mereka perlukan rujukan tentang keperluan produk, mereka perlu merujuk semula kepada mesej yang telah dikomunikasi untuk mengenal pasti keperluan itu. Bagi menyelesaikan masalah ini, kami membuat rujukan kepada konsep umum Carotene yang menawarkan teknik mengklasifikasi teks secara automatik melalui cascade (pecahan perkataan dalam bentuk menegak) dan cluster (mengumpulkan perkataan) kasifikasi. Konsep ini memerlukan set data latihan untuk teks klasifikasi tetapi untuk projek ini, kami menggunakan konsep cascade dan cluster klasifikasi berdasarkan fungsi label hash (tanda pagar). Teknik ini dikenali sebagi konsep umum Carotene (HLC) yang dicipta di dalam alat komunikasi yang membolehkan pengenalpastian keperluan melalui klasifikasi mesej teks. Klasifikasi cascade membuat pecahan ayat kepada perkataan dan klasifikasi cluster mengumpulkan perkataan menjadi ayat di dalam kumpulan yang sama. Objektif tesis ini adalah untuk mengenalpasti teknik untuk klasifikasi teks untuk mengesan keperluan di dalam teks mesej, mereka bentuk penyelesaian dan menilai teknik ini. Di dalam tesis ini juga terdapat penerangan tentang masalah dan penyelesaian berkaitan dengan komunikasi dalam kumpulan Agile dan deskripsi tentang reka bentuk dan pelaksanaan HLC teknik untuk fungsi klasifikasi teks mesej dalam alat pengurusan komunikasi Agile (AgileCom). Akhir sekali, tesis ini dirumuskan dengan penilaian tentang keberkesanan dan kepuasan kegunaan HLC teknik.

ACKNOWLEDGEMENT

First of all, I am thankful to my supervisor Dr. Pathiah Binti Abdul Samat, lecturer of software engineering department that always give guidance, references and support during this project. Without her supervision from time to time, it will be difficult to complete this work.

I also would like to thank my advisor, Dr. Mar Yah @ Mek Yah Binti Said who keep monitoring our progress of work for these research and project work.

Furthermore, I would like to thank the University Putra Malaysia (UPM) for providing the facilities such as online library for me to browse for digital journal databases to complete my literature review and study.

Last but not least, I would like to take this opportunity to express deepest appreciation and special thanks to my families and friends who continuously support and lending help whenever I feel needed during this master journey. Lastly, I also would like to express my sincere gratitude to the persons who have directly or indirectly involved in this project. Without this strong support, this project will not be able to complete. Thank you very much.

DECLARATION

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this project report has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the project report and copyright of project report are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature:  Date: 27/06/2019

Name and Matric No.: NOR HIDAYAH BINTI ZAINAL ABIDIN (GS48465)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
FR	Functional Requirement
NFR	Non-Functional Requirement
AgileCom	Agile Communication Management Tool
HLC	High Level Carotene concept
PHP	Personal Home Page
UME	User Magnitude Estimation
SUS	System Usability Scale

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Organizations worldwide have adopted the technology to replace their manual working culture. This trend is highly demanded and software industries received tremendous job to develop applications that can help to transform these organizations. These digital transformation phenomena that happening across the globe has pushed for better techniques and methodologies to be implemented in their own organization. This trend has created more opportunities for software developer to develop system that demanded by the customers. In order to develop the system, proper software development methodology has to be implemented throughout every single phases of the development.

In today's high demanded for latest system features and quick deliverables requires the businesses to keep changing its direction according to the market demand to compete and stay relevant to the industry. This evolution has led to the emergence of agile development methods. Instead of trying to identify a complete set of requirements before the project begins and to eliminate as much change as early as possible, agile methods embrace change by accepting and incorporating it instead of eliminating it. Agile software development refers to a group of software development

methodologies based on iterative development, where requirements and solutions evolve through collaboration between self-organizing cross-functional teams. It has become the main choices for development methodology for most software development team in the world. Several examples of Agile Software Development Methodologies are Extreme Programming, Crystal Methodologies, Scrum, Adaptive Software Development, Feature-Driven Development and Dynamic Systems Development Methodology etc.

The organizations that adopt agile methodology can have team that are co-located and distributed across the oceans. The co-located agile team is the team that works in the same location within the same room, often called a 'war room' working besides one another that let them to have face to face interaction while communication. Meanwhile, distributed agile team dispersed in some manner even when some of the team members in cubicles or in separate offices or in different floors, it is slightly distributed, and also geographically distributed across the world. Distributed teams can work on projects around the clock, and strong talent can be found in less competitive markets.

In terms of communication and information or data exchange, many distributed agile team uses messaging tool to chat and communicate regarding the product requirements or any project related information. This method produce large number of data in the messaging features in certain messaging tool. So the information is everywhere and the team usually will

scroll back the tool to view previous messages for references. Since project is tied to the timeline and customer requires fast deliverables as per agile methodology promoted, the development team needs to have a tool that allows them to work faster. By classifying the text messages sentences and stored it into structured board will help them to have easy to use platform as reference, better focus and save time. The classification of text messages sentences promote direct and specific interpretation of the information discussed. In industry, there are numerous number of text classification techniques such as Carotene, ConvNet (Convolutional Network) etc.

1.2 Problem Statements

Agile methodology emphasizes on face to face communication and small team. It suits to the “get-together” team and “don’t-do-code” people. Agile fits nicely into the manager’s schedule, but not the maker’s. The Agile methodology is recommended for co-located teams. However, most of the software team is geographically distributed. It will be difficult in terms of communication. Although various technologies offered for communication across the globe, none had fully identified and prepare the one-centralized platform that can identify and categorize the specific requirement that has been discussed by the team in the messaging tool’s discussion board. This shows that market lack of the tool that focusing on the development to fulfill what agile lack in terms of communication. It is an ineffective communication method used for information exchange (Chadli et al. 2016; Verner et al. 2014) when the agile team had to use different platform for communication. These multiple platforms can be confusing and headache because the information

will be unstructured and either party might not familiar with the proposed communication platform.

In addition, agile team had less reference in terms of documentation to work on the product except if customer request. This is because agile manifesto which is working software over comprehensive documentation doesn't include documentation as part of its main deliverables (Voigt, S., 2016). The situation will get worse when they inherited someone else's work for example code or product requirements because the person had to be responsible to create program based on the incorporated legacy references and make it an up-to-date version (Fu et.al., 2017). They can avoid the issue if proper knowledge transition has been communicated beforehand through face to face communication for co-located team or through messaging for the team that is distributed geographically. However, whenever the distributed team communicates about the software requirements through messaging, the data will be unstructured and messy since there will be a lot of waste in the communicated sentences. The team that need to refer back to the data need to scroll over again to find the main point and accurate information of the requirements types. This is applicable to most of the messaging tools.

Other than that, the cross team background with different mother tongue and first language might misunderstand the discussed items during communication. Although English language has been recognized internationally as the global communication medium, some engineers might not well verse with the language's slang (different country had different

English slang), bombastic wording etc. The team background with different mother tongue and first language might misunderstand the discussed items during communication due to invisibility of body language and the terminologies used as well as un-conveyed information (Chadli et al. 2016; Lous et al. 2017; Nguyen-Duc et al. 2015; Niazi, Mahmood, Alshayeb, Qureshi, et al. 2016; Verner et al. 2014; Yaseen et al. 2015; Zahedi et al. 2016).

1.3 Objectives of the Thesis

There are 3 objectives to be achieved by the implementation of this tool:

1) To identify the technique for text classification function for requirements identification in text messages for distributed agile team communication platform.
2) To design the text messages classification function for requirements identification for distributed agile team communication platform.
3) To evaluate the effectiveness and usability of the text messages classification function for requirements identification for the global software development team.

1.4 Scope of the Thesis

The scope of this study is to identify the requirements by classification of text messages into software requirements types of functional or non-functional requirements specifically for software engineers with experience working in distributed agile team environment in order for them to easy understand the software requirements and have centralize structured platform as reference

of product requirements. The message is a text sentences that typed in the messaging room.

1.5 Structure of the Thesis

This thesis contains seven chapters were organized as below. Chapter 1 is Introduction. Chapter 2 is Literature Review, which we do the content analysis from articles, journals and other publications published by previous researchers regarding the challenges and related solutions on the classification of messages in communication for distributed agile team and we the summarize it. Chapter 3 is Methodology, we study the methods and processes involved in the communication management and requirement management especially for distributed agile team. Here, we highlight the research methodology activity and steps taken throughout the every phases. Chapter 4 we elaborate all the design processes involves in HLC Technique and AgileCom tool development. Chapter 5, we illustrated and explained the development and implementation processes. Here, we evaluated the HLC technique in the AgileCom tool for its effectiveness and usability satisfaction to classify messages into FR and NFR during communication. In Chapter 6 we discuss the results and findings from the study and also from the HLC and AgileCom tool evaluation survey. The developed tool will be evaluated by Distributed Agile practitioners whose have experience in Agile Global Software Development project. Lastly is Chapter 7, we explain about the conclusion, study limitation and the recommendation for future work.

REFERENCES

1) Javed, F., Luo, Q., McNair, M., Jacob, F., Zhao, M., & Kang, T. S. (2015, March). Carotene: A job title classification system for the online recruitment domain. In *Big Data Computing Service and Applications (BigDataService), 2015 IEEE First International Conference on* (pp. 286-293). IEEE.
2) Alzoubi, Y. I., Gill, A. Q., & Al-Ani, A. (2015). Distributed Agile Development Communication: An Agile Architecture Driven Framework. *JSW*, *10*(6), 681-694.
3) Chikhale, M. M., & Mansouri, M. (2015). An agile and collaborative framework for effective governance to enhance management in large-scale enterprise business systems: The case of Apple Inc. *Global Journal of Flexible Systems Management*, *16*(3), 283-293.
4) Dikert, K., Paasivaara, M., & Lassenius, C. (2016). Challenges and success factors for large-scale agile transformations: A systematic literature review. *Journal of Systems and Software*, *119*, 87-108.
5) Hoda, R., & Murugesan, L. K. (2016). Multi-level agile project management challenges: A self-organizing team perspective. *Journal of Systems and Software*, *117*, 245-257.
6) Esquivel, S. (2016). Communication Issues in Agile Software Development. In *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*.
7) Papatheocharous, E., & Andreou, A. S. (2014). Empirical evidence and state of practice of software agile teams. *Journal of Software: Evolution and Process*, *26*(9), 855-866
8) Khalil, C., & Khalil, S. (2016). A governance framework for adopting Agile methodologies. *International Journal of e-Education, e-Business, e-Management and e-Learning*, *6*(2), 111.
9) Kaur, H., & Haddad, H. M. (2015, January). Distributed Agile Development: A Survey of Challenges and Solutions. In *Proceedings of the International Conference on Software Engineering Research and Practice (SERP)* (p. 42). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
10) Embretsen, D., & Hyder, L. (2017). Scrum in Global Software Development: An Ethnographic Case Study of Scrum's Mitigation Effects on Global Software Development Challenges.
11) Chadli, S. Y., Idri, A., Fernández-Alemán, J. L., Ros, J. N., & Toval, A. (2016, November). Identifying risks of software project management in global software development: an integrative framework. In *Computer Systems and Applications (AICCSA), 2016 IEEE/ACS 13th International Conference of*(pp. 1-7). IEEE.
12) Verner, J. M., Brereton, O. P., Kitchenham, B. A., Turner, M., and Niazi, M. 2014. “Risks and Risk Mitigation in Global Software Development: A Tertiary Study,” *Information and Software Technology*

(56:1), Elsevier B.V., pp. 54–78. (https://doi.org/10.1016/j.infsof.2013.06.005).
13) Bondoc, E. R., & Ilao, J. (2018). Theoretical Understanding Of Impacts Of Team Separation In Global Software Development.
14) Nguyen-Duc, A., Cruzes, D. S., and Conradi, R. 2015. "The Impact of Global Dispersion on Coordination, Team Performance and Software Quality-A Systematic Literature Review," *Information and Software Technology* (57:1), Elsevier B.V., pp. 277–294. (https://doi.org/10.1016/j.infsof.2014.06.002).
15) Niazi, M., Mahmood, S., Alshayeb, M., Qureshi, A. M., Faisal, K., & Cerpa, N. (2016). Toward successful project management in global software development. *International Journal of Project Management*, *34*(8), 1553-1567.
16) Yaseen, M., Baseer, S., & Sherin, S. (2015, December). Critical challenges for requirement implementation in context of global software development: A systematic literature review. In *Open Source Systems & Technologies (ICOSST), 2015 International Conference on* (pp. 120-125). IEEE.
17) Zahedi, M., Shahin, M., & Babar, M. A. (2016). A systematic review of knowledge sharing challenges and practices in global software development. *International Journal of Information Management*, *36*(6), 995-1019.
18) Qureshi, R., Basheri, M., & Alzahrani, A. A. (2018). Novel Framework to Improve Communication and Coordination among Distributed Agile Teams. *International Journal of Information Engineering and Electronic Business*, *10*(4), 16.
19) Korkala, M., & Maurer, F. (2014). Waste identification as the means for improving communication in globally distributed agile software development. *Journal of Systems and Software*, *95*, 122-140.
20) Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in neural information processing systems* (pp. 649-657).
21) Alzoubi, Y. I., Gill, A. Q., & Al-Ani, A. (2015). Distributed Agile Development Communication: An Agile Architecture Driven Framework. *JSW*, *10*(6), 681-694.
22) Coletta, L. F. S., da Silva, N. F. F., Hruschka, E. R., & Hruschka, E. R. (2014, October). Combining classification and clustering for tweet sentiment analysis. In *2014 Brazilian Conference on Intelligent Systems* (pp. 210-215). IEEE.
23) Lu, M., & Liang, P. (2017, June). Automatic classification of non-functional requirements from augmented app user reviews. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering* (pp. 344-353). ACM.
24) Friesen, E. L. (2017, September). Measuring AT Usability with the Modified System Usability Scale (SUS). In *AAATE Conf.*(pp. 137-143).
25) Kaya, A., Ozturk, R., & Gumussoy, C. A. (2019). Usability Measurement of Mobile Applications with System Usability Scale (SUS). In *Industrial Engineering in the Big Data Era* (pp. 389-400). Springer, Cham.

26) Fu, Q., Grady, F., Broberg, B. F., Roberts, A., Martens, G. G., Johansen, K. V., & Loher, P. L. (2017). Code review and cooperative pair programming best practice. *arXiv preprint arXiv:1706.02062*.
27) Voigt, S., von Garrel, J., Müller, J., & Wirth, D. (2016, September). A study of documentation in agile software projects. In *Proceedings of the 10th ACM/IEEE international symposium on empirical software engineering and measurement* (p. 4). ACM.