

### INFORMATION ON DOCTORAL THESIS

1. Full name : Bui Thi Hong Nhung..... 2. Sex: Female  
3. Date of birth: 25.04.1979 ..... 4. Place of birth: Hanoi  
5. Admission decision number: 642/QĐ-CTSV Dated: September 15, 2014  
6. Changes in academic process: .....  
7. Official thesis title: *Studing of trace representation techniques and trace clustering in process discovery*  
8. Major: Information System 9. Code: 9480104.01

10. Supervisors: .....

Assoc. Prof. Dr. Nguyen Tri Thanh

Assoc. Prof. Dr. Nguyen Cam Tu

11. Summary of the **new findings** of the thesis: .....

This dissertation has obtained the major contributions as follows:

- Firstly, the thesis proposes a trace model based on distance graph in which the activities corresponding to the vertices of the graph, the arc is the connection between the two activities. This representation can describe information about the order and k-distance between activities in a trace, increasing flexibility and efficiency and providing a richer trace representation solution for the differential problem.

- Secondly, the thesis proposes a standardized NTW traces measurement to determine the distance or similarity between traces in an event log using the idea of Google measurement. The NTW measure provides a method of calculating the global distance between traces in the context of comparing them with all other traces in the event log. NTW has been proved to be more effective, contributing to a new trace distance measurement method to improve the quality of trace clustering problems in the field of process mining.

- Thirdly, the thesis proposes a new trace clustering algorithm for ContextTracClus for process mining, including proposing two new concepts of trace context and context tree; propose 3 algorithms for constructing context trees; trace context identification from the context tree and the ContextTracClus trace

clustering algorithm allow automatic detection of the appropriate number of clusters and cluster the traces into a set of sub-traces that have similar execution context. Experimental results show that the trace clusters generated by the ContextTracClus algorithm are highly effective when compared to the K-means clustering algorithm with significantly reduced complexity and computation time.

- Fourthly, the thesis proposes four solutions to apply advanced research results in the world on deep learning to the problem of performing log of event log, including: CompactTrace condensed trace representation solution using network model DNN deep learning neurons; TraceEmbedding solution using embedded model from CBOW; TraceLSTM trace representation solution uses LSTM long-short-term memory model and TraceBERT trace solution uses BERT two-dimensional context learning model. The results have achieved remarkable results compared to traditional trace methods.

12. Practical applicability, if any: .....

13. Further research directions, if any: .....

In the next time, the PhD student will continue to research solutions for the remaining limitations of the thesis and continue to implement proposals to improve the solutions of representation and trace clustering.

- Firstly, it is necessary to study and supplement the application of advanced models in the world to the problem of improving the quality of the process model. Especially, Graph neural networks (GNNs) are one of the current prominent trends and are similar to the structure of the process model in process mining. At the same time, further analyze the results of the proposed solutions of the thesis with a number of new studies.

- Second, modularize the thesis's research so that the research results can be applied in practice.

14. Thesis-related publications: .....

1. [NhunbBTH01] Quang-Thuy Ha, Hong-Nhung Bui, and Tri-Thanh Nguyen (2016), "A trace clustering solution based on using the distance graph model", *Proceeding of the 8<sup>th</sup> International Conference on Computational Collective Intelligence (ICCCI)*, Lecture Note of Artificial Intelligence (LNAI), Springer, pp. 313-320, (Conference Proceedings Citation Index (CPCI), Scopus, DBLP, Clarivate Analytics' Web of Science, EI) (9 references from foreign authors).
2. [NhunbBTH02] Hong-Nhung Bui, Quang-Thuy Ha, and Tri-Thanh Nguyen (2018), "A Novel Similarity Measure for Trace Clustering Based on Normalized Google Distance", *JP Journal of Heat and Mass Transfer*, Special Volume, Issue III, Advances

in Mechanical System and ICT-convergence, Pushpa publishing house, pp. 341-346 (Scopus indexed journal).

3. [NhungBTH03] Hong-Nhung Bui, Tri-Thanh Nguyen, Thi-Cham Nguyen, and Quang-Thuy Ha (2018), "A new trace clustering algorithm based on context in process mining", *Proceeding of International Joint Conference on Rough Sets (IJCRS)*, Lecture Note of Artificial Intelligence (LNAI), Springer, pp. 644-658, (Conference Proceedings Citation Index (CPCI), Scopus, DBLP, Clarivate Analytics' Web of Science, EI).
4. [NhungBTH04] Hong-Nhung Bui, Trong-Sinh Vu, Tri-Thanh Nguyen, Thi-Cham Nguyen, and Quang-Thuy Ha (2019), "A Compact Trace Representation Using Deep Neural Networks for Process Mining", *Proceeding of the 11th IEEE International Conference on Knowledge and Systems Engineering (KSE)*, pp. 312-316, (DBLP, IEEEXplore) (2 references from foreign authors).
5. [NhungBTH05] Hong-Nhung Bui, Trong-Sinh Vu, Hien-Hanh Nguyen, Tri-Thanh Nguyen, and Quang-Thuy Ha (2020), "Exploiting CBOW and LSTM Models to Generate Trace Representation for Process Mining", *Proceeding of the 12th Asian Conference on Intelligent Information and Database Systems (ACIIDS)*, pp. 35-46 (Scopus, DBLP) (1 reference from foreign authors).