VIETNAM NATIONAL UNIVERSITY, HANOI
VNU UNIVERSITY OF ENGINEERING AND TECHNOLOGY

SOCIALIST REPUBLIC OF VIETNAM
Independence – Freedom – Happiness

INFORMATION ON DOCTORAL THESIS

1. Full name : Lê Hoàng Quỳnh .....2. Sex: Female

3. Date of birth: May 02, 1987 ..... 4. Place of birth: Hà Nội

5. Admission decision number: 985/QD-CTSV Dated December 04, 2014 of the Rector of University of Engineering and Technology (UET), VNU.

6. Changes in academic process: Extend the education period for 2 years

(List the forms of change and corresponding times)

7. Official thesis title: Machine learning-based extraction of semantic relations from biomedical literature

8. Major: Information Systems .....9. Code: 9480104.01 .....

10. Supervisors: Prof. Nigel Collier and Dr. Đặng Thanh Hải

(Full name, academic title and degree)

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

- The Dissertation participates in the research trend of bioNLP in general and biomedical relation extraction in particular. Our focuses are on improving the methods and build an effective architecture for biomedical named entity recognition and relation classification, rather than on developing new machine learning algorithms. In this dissertation, several advanced machine learning architecture for biomedical named entity recognition and relation classification are proposed. These architectures applied and improved many machine learning methods such as unsupervised machine learning with Skip-gram, supervised learning with Conditional Random Fields, Support Vector Machine, Supervised Semantic Indexing, averaged perceptron, Convolutional Neural Networks, Recurrent Neural Network and distant-supervision learning. Most of them bring potential and comparative results.
- We propose a novel representation for document with inter-sentence relations that based on dependency, co-references and knowledge-based information.
- We provide a sentence-level silver standard corpus for the research community.

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

The dissertation research can be applied to build biomedical named entity recognition and relation extraction systems, laying an important foundation to extract the information from scientific literature in the biomedical field.

13. Further research directions: (if any) Our team are continuing the research in this area, towards more complex problems such as multi-document summarization.

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

(List them in chronological order)

1. Hoang-Quynh Le, Mai-Vu Tran, Thanh Hai Dang, Quang-Thuy Ha and Nigel Collier (2016). "Sieve-based coreference resolution enhances semi-supervised learning model for chemical-induced disease relation extraction". *Database (2016),* Vol. 2016: article ID baw102.
2. Hoang-Quynh Le, Duy-Cat Can, Thanh Hai Dang, Mai-Vu Tran, Quang-Thuy Ha and Nigel Collier (2017). "Improving chemical-induced disease relation extraction with learned features based on convolutional neural network". *In proceedings of the 9th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 292-297. IEEE.
3. Thanh Hai Dang, Hoang-Quynh Le, Trang M. Nguyen, Sinh T. Vu (2018). "D3NER: Biomedical named entity recognition using CRF-biLSTM improved with fine-tuned embeddings of various linguistic information". *Bioinformatics*, 34(20), pp 3539-3546.
4. Hoang-Quynh Le, Duy-Cat Can, Sinh T. Vu, Thanh Hai Dang, Mohammad Taher Pilehvar and Nigel Collier (2018). "Large-scale Exploration of Neural Relation Classification Architectures". *In proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2266-2277.
5. Duy-Cat Can, Hoang-Quynh Le and Quang-Thuy Ha (2019). "Improving Semantic Relation Extraction System with Compositional Dependency Unit on Diverse Shortest Dependency Path". *In proceedings of the 11th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2019)*, pp. 140-152. Springer, Cham.
6. Duy-Cat Can, Hoang-Quynh Le*, Quang-Thuy Ha and Nigel Collier. "A Richer-but-Smarter Shortest Dependency Path with Attentive Augmentation for Relation Extraction". *In proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 2902-2912. (*Corresponding author).
7. Mai-Vu Tran, Hoang-Quynh Le, Duy-Cat Can, Huyen Nguyen, Linh Nguyen Tran Ngoc and Tam Doan Thanh. "Overview of VLSP RelEx shared task: A Data

Challenge for Semantic Relation Extraction from Vietnamese News". *In Proceedings of the 7th international workshop on Vietnamese Language and Speech Processing (VLSP 2020)*. Association for Computational Linguistics, 2020.

8. Hoang-Quynh Le, Quoc-An Nguyen, Quoc-Hung Duong, Minh-Quang Nguyen, Huy-Son Nguyen, Tam Doan Thanh, Hai-Yen Thi Vuong, and Trang M. Nguyen. "UETfishes at MEDIQA 2021: Standing-on-the-Shoulders-of-Giants Model for Abstractive Multi-answer Summarization". *In Proceedings of the 20th SIGBioMed Workshop on Biomedical Language Processing, NAACL-BioNLP 2021.* Association for Computational Linguistics, 2021.