

12. Practical applicability, if any:

The results of the thesis are to increase the quality of English – Vietnamese SMT system. Furthermore, the results have been published in journals and international conferences, thus contributing to science in the field of natural language processing, especially for the development of machine translation systems have the ability to use in reality.

13. Further research directions, if any:

- Firstly, for building the bilingual corpus, we will extend our work on extracting smaller parallel components such as paragraphs, sentences or phrases. This work will also be interesting in the case the quality of translation between bilingual web pages is not good.
- Secondly, the problem of word alignment for SMT: we will handle exceptions occur for constraints as proposed.
- Thirdly, to identify English - Vietnamese parallel phrases for SMT: we will use much more syntactic patterns to increase the number of extracted parallel phrases. Also, we plan to use the proposed method for building a dictionary of parallel phrases for the language pair English-Vietnamese and will improve the quality of English-Vietnamese statistical machine translation by using this dictionary.

14. Thesis-related publications:

- [1] Le Quang Hung and Le Anh Cuong (2010), “Extracting parallel texts from the web”, *Proceedings of the Second International Conference on Knowledge and Systems Engineering, IEEE Computer Society*, pages 147-151.
- [2] Le Quang Hung and Le Anh Cuong (2012), “Improving Word Alignment for Statistical Machine Translation Based on Constraints”, *Asian Language Processing (IALP), International Conference on, IEEE Computer Society*, pages 113-116.
- [3] Le Quang Hung and Le Anh Cuong (2012), “Statistical Word Alignment with Part-of-Speech Constraint”, *Kỹ yếu hội thảo Quốc gia lần thứ XV "Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông"*, trang 410-416.
- [4] Quang-Hung LE, Duy-Cuong NGUYEN, Duc-Hong PHAM, Anh-Cuong LE, and Van-Nam HUYNH (2013), “Paragraph Alignment for English-Vietnamese Parallel E-Books”, In *Knowledge and Systems Engineering, Springer International Publishing*, pages 251-259.
- [5] Quang-Hung LE, Anh-Cuong LE, and Van-Nam HUYNH (2013), “Parallel phrase extraction from English-Vietnamese parallel corpora”, In *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2013 IEEE RIVF International Conference on*, pages 175-179.

- [6] Le Quang Hung and Le Anh Cuong (2013), “An effective method to sentence alignment for the English-Vietnamese parallel e-book”, *Kỷ yếu hội thảo Quốc gia lần thứ XVI "Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông"*, trang 12-16.
- [7] Le Quang Hung (2014), “A new approach to extract parallel corpus”, *Tạp chí khoa học Trường Đại học Quy Nhơn*, Số 4 - Tập VIII, trang 12-24.
- [8] Quang-Hung LE and Anh-Cuong LE (2014), “Syntactic pattern based Word Alignment for Statistical Machine Translation”, *The International Journal of Knowledge and Systems Science (IJKSS) IGI Global Publishing*, Volume 5 Issue 3, pages 36-45.

Supervisor

Phd Student

Date: *09-Sep-2015*
Signature:

Date: *09-Sep-2015*
Signature:

Full name: **Le Anh Cuong**

Full name: **Le Quang Hung**