Project Title: ASEAN Language Speech Translation thru' U-STAR

Project Leader: Prof Li Haizhou/Ms Aw Ai Ti, Institute for Infocomm Research
Project members: Dr Rapid Sun, National Institute of Posts, Telecommunications, and Information Communication Technology, Cambodia (NIPTICT); Dr. Hammam Riza, Agency for the Assessment and Application of Technology, Indonesia (BPPT); Prof. Sevia M. Idrus, Universiti Teknologi Malaysia, Malaysia (UTM); Prof. Khin Mar Soe, University of Computer Studies, Yangon, Myanmar (UCSY); Dr. Chai Wutiwiwatchai, National Electronics and Computer Technology Center, National Science and Technology Development Agency, Thailand (NECTEC); Prof. NGUYEN Thi Thu Trang, Hanoi University of Science and Technology, Vietnam (HUST); Prof. Luong Chi Mai, Institute of Information Technology, Vietnam Academy of Science and Technology, Vietnam (IOIT)
Budget: US$87000 **Duration**: 1 July 2016 to 30 June 2019

Target of this project: This project aimed to promote innovative application development using U-STAR services within ASEAN to benefit the community in human communication and language learning.

Findings and Outcomes:
A Myanmar-to-English translation prototype was developed. The prototype was trained on 228,767 parallel sentences using a LSTM encoder-decoder syllable-based segmentation. The prototype achieved the best performance when compared to word-based NMT and character-based NMT.

Collaborations:
Leveraging on U-STAR infrastructure, the project team collectively developed a mobile application – uniTRANS, to provide speech-to-speech translation to facilitate communication among ASEAN community. To encourage the use of this application among different language speaking users, the team localized the user interface to support all native languages of the project team members.

On top of the mobile app, the team developed the speech data collection guidelines and language resources to promote the research and development on low-resources ASEAN languages.

Broader Impact and Future Developments:
Most of the ASEAN languages are low resource languages. Data collected through this project can motivate researchers working on low-resources NLP or MT to work on ASEAN languages and promote the research and development on ASEAN languages.
The project team can also continue to leverage on U-STAR infrastructure to collaborate on speech translation technologies and develop innovative applications using U-STAR services to benefit the community in human communication.

Social Contribution:
The project supported Dr. Nguyen Thi Thu Trang, Hanoi University of Science and Technology, to present their paper on "A Hybrid Method for Vietnamese Text Normalization" in the International Conference on Natural Language Processing and Information Retrieval (http://www.nlpir.net), June 28-30, 2019.
Two workshops were conducted to facilitate the sharing of knowledge among the team members. The first workshop was conducted on 4th October 2016 while the second workshop was held on 6th December 2017.