

VLSP 2020

**Proceedings of the 7th International Workshop on
Vietnamese Language and Speech Processing**

18 December, 2020

Hanoi University of Science and Technology

Hanoi, Vietnam

©2020 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Introduction

In 2020, the Association for Vietnamese Language and Speech Processing (VLSP) is officially founded, as a chapter of Vietnam Vietnam Association for Information Processing. VLSP consortium is an initiative which came from Institute of Information Technology - Vietnam Academy of Science and Technology, to establish a community working on speech and text processing for Vietnamese language. The first national project KC01.03/06-10 from 2007-2009 which received strong support from Ministry of Science and Technology, gathered eight active research groups from universities and institutes in Vietnam and overseas. A main goal of the first project is to set up long term strategy on Vietnamese language processing in order to provide and to involve community to enrich shared language resources and tools for R&D purpose. Since 2012, the VLSP Consortium has organized a series of workshops, in conjunction with large international conferences organized in Vietnam. Until 2020, six events have taken place with different forms of activities such as technical reports, activity reports, discussion panel, shared tasks on VLSP. From VLSP 2016, organizing shared tasks on Vietnamese processing became the main activities of the workshop series in order to promote the development of essential tools and resources for VLSP.

The seventh VLSP workshop constitutes the first workshop organized by the Association for VLSP, held in Hanoi in December 2020. To mark this event, seven shared tasks have been proposed and in the end six challenges have been organized:

1. Vietnamese Universal Dependency Parsing (UDP): the task of determining syntactic dependencies between words in a sentence.
2. Vietnamese Relation Extraction (RE): the task of identifying and determining the semantic relations between pairs of named entity mentions within a single sentence.
3. English-Vietnamese Machine Translation (MT): text translation from English to Vietnamese in the news domain.
4. Reliable Intelligence Identification on Vietnamese Social Network Sites (ReINTEL): the task of identifying a piece of information shared on social network sites (SNSs) as reliable or unreliable.
5. Automatic Speech Recognition for Vietnamese (ASR): the task includes two evaluation sub-tasks. For the first sub-task, all participants are required to use only the provided data to develop ASR models including acoustic and language models. For the second one, participants can use all available data sources to develop their ASR models without any limitation.
6. Vietnamese Text-To-Speech on Common Datasets (TTS): the task of building a TTS system with a training voice from the same speech database released by the organizers.

The participants to the evaluation campaign were asked to present their system in a dedicated paper.

We very much hope that you have had an enjoyable and inspiring time!

Huyen T M. Nguyen, Xuan-Son Vu, Chi Mai Luong

Hanoi & Umeå

February 2021

Organisers:

General Chair: Nguyễn Thị Minh Huyền

Organizing Co-chairs: Nguyễn Văn Huy, Hà Thành Lê, Vũ Xuân Sơn, Nguyễn Thị Thu Trang, Trần Thế Trung, Trần Mai Vũ

Program Committee Co-chairs: Nguyễn Lê Minh, Lương Chi Mai

Publication Co-chairs: Vũ Xuân Sơn, Nguyễn Thanh Sơn, Hà Thành Lê, Vũ Tiến Thành

Program Committee:

Ngô Xuân Bách (PTIT, Vietnam), Nguyễn Việt Cường (INT²), Lê Anh Cường (TDTU, Vietnam), Phạm Hiến (IOL, VASS), Phan Xuân Hiếu (UET, VNU), Lê Quang Hùng (QNU, Vietnam), Lê Thanh Hương (HUST, Vietnam), Nguyễn Văn Huy (VINBDI), Nguyễn Thị Minh Huyền (HUS, VNU) Hà Thành Lê (KIT, Germany & VINBDI, Vietnam), Lương Chi Mai (IOIT, VAST), Nguyễn Lê Minh (JAIST, Japan), Phạm Quang Nhật Minh (Aimesoft Co., Vietnam), Nguyễn Lưu Thuỳ Ngân (UIT, VNUHCM), Lê Hồng Phương (HUS, VNU), Vũ Xuân Sơn (Umeå University, Sweden), Nguyễn Phương Thái (UET, VNU), Lê Công Thành (InfoRe), Nguyễn Thị Thu Trang (HUST, Vietnam), Trần Thế Trung (FPT, Vietnam), Nguyễn Văn Vinh (UET, VNU), Trần Mai Vũ (UET, VNU), Lê Đức Trọng (UET, VNU), Harry Nguyen (University of Glasgow, Singapore), Alex To, (ReML.AI) Nguyễn Thuỳ Trinh (ReML.AI), Nguyễn Quang (UET, VNU), Linh Le (ReML.AI), Hoàng Minh Đức (ReML.AI), Nguyễn Anh Tuấn (ReML.AI), Lê Nghĩa (ReML.AI)

Table of Contents

ReINTEL Challenge 2020: Vietnamese Fake News Detection using Ensemble Model with PhoBERT embeddings	1
<i>Thuan Nguyen Hieu, Hieu Cao Nguyen Minh, Hung To Van and Bang Vo Quoc</i>	
ReINTEL Challenge 2020: A Comparative Study of Hybrid Deep Neural Network for Reliable Intelligence Identification on Vietnamese SNSs	6
<i>Hoang Viet Trinh, Tung Tien Bui, Tam Minh Nguyen, Huy Quang Dao, Quang Huu Pham and Ngoc N. Tran</i>	
An Empirical Study of Using Pre-trained BERT Models for Vietnamese Relation Extraction Task at VLSP 2020	13
<i>Minh Quang Nhat Pham</i>	
Improving prosodic phrasing of Vietnamese text-to-speech systems	19
<i>Phuong Pham Ngoc, Chung Tran Quang, Quang Minh Nguyen and Quoc Truong Do</i>	
Development of Smartcall Vietnamese Text-to-Speech for VLSP 2020	24
<i>Khuong Duy Trieu, Ba Quyen Dam and Quoc Bao Nguyen</i>	
Vietnamese Relation Extraction with BERT-based Models at VLSP 2020	30
<i>Thuật Nguyễn and Hiếu Mẫn</i>	
Vietnamese Text-To-Speech Shared Task VLSP 2020: Remaining problems with state-of-the-art techniques	35
<i>Thi Thu Trang Nguyen, Hoang Ky Nguyen, Quang Minh Pham and Duy Manh Vu</i>	
ReINTEL Challenge 2020: A Multimodal Ensemble Model for Detecting Unreliable Information on Vietnamese SNS	40
<i>Manh Duc Tuan Nguyen and Quang Nhat Minh Pham</i>	
ReINTEL Challenge 2020: Exploiting Transfer Learning Models for Reliable Intelligence Identification on Vietnamese Social Network Sites	45
<i>Kim Nguyen Thi Thanh and Kiet Nguyen Van</i>	
A Joint Deep Contextualized Word Representation for Deep Biaffine Dependency Parsing	49
<i>Xuan-Dung Doan</i>	
Applying Graph Neural Networks for Vietnamese Dependency Parsing	54
<i>Nguyen Duc Thien, Nguyen Thi Thu Trang and Truong Dang Quang</i>	
Implementing Bi-LSTM-based deep biaffine neural dependency parser for Vietnamese Universal Dependency Parsing	60
<i>Lien Nguyen</i>	
Vietnamese-English Translation with Transformer and Back Translation in VLSP 2020 Machine Translation Shared Task	64
<i>Le Duc Cuong and Trang Nguyen Thi Thu</i>	
The UET-ICTU Submissions to the VLSP 2020 News Translation Task	71
<i>Ngo Thi-Vinh, Nguyen Minh-Thuan, Nguyen Hoang Minh Cong, Nguyen Hoang-Quan, Nguyen Phuong-Thai and Nguyen Van-Vinh</i>	
VLSP 2020 Shared Task: Universal Dependency Parsing for Vietnamese	77

Ha My Linh, Nguyen Thi Minh Huyen, Vu Xuan Luong, Nguyen Thi Luong, Phan Thi Hue and Le Van Cuong

ReINTEL: A Multimodal Data Challenge for Responsible Information Identification on Social Network Sites 84

Duc-Trong Le, Xuan-Son Vu, Nhu-Dung To, Huu-Quang Nguyen, Thuy-Trinh Nguyen, Thi Khanh-Linh Le, Anh-Tuan Nguyen, Minh-Duc Hoang, Nghia Le, Huyen Nguyen and Hoang D. Nguyen

Overview of VLSP ReIEx shared task: A Data Challenge for Semantic Relation Extraction from Vietnamese News 92

Vu Tran Mai, Hoang-Quynh Le, Duy-Cat Can, Thi Minh Huyen Nguyen, Tran Ngoc Linh Nguyen and Thanh Tam Doan

Goals, Challenges and Findings of the VLSP 2020 English-Vietnamese News Translation Shared Task 99

Thanh-Le Ha, Van-Khanh Tran and Kim-Anh Nguyen