

CLB Xử lý Ngôn ngữ và Tiếng nói tiếng Việt
Chi hội của Hội Tin học Việt Nam

CAM KẾT TUÂN THỦ CÁC ĐIỀU KHOẢN CUỘC THI

Cuộc thi Nhận dạng tiếng nói tiếng Việt trong khuôn khổ hội thảo VLSP 2020 được tổ chức theo các điều khoản sau đây.

I. Điều khoản tham gia cuộc thi

1. Đối tượng dự thi bao gồm các đội đã đăng kí tham dự phần thi Nhận dạng tiếng nói (*Automatic Speech Recognition*) trên mẫu đăng kí online trên website <https://vlsp.org.vn/vlsp2020/eval>
2. Mỗi đội tham gia Task-01 (ASR-T1) thì **CHỈ ĐƯỢC PHÉP SỬ DỤNG DỮ LIỆU** do Ban tổ chức cung cấp để phát triển mô hình nhận dạng bao gồm các mô hình âm học, mô hình ngôn ngữ, mô hình tiền và hậu xử lí (*acoustic model, language model, pre-processing model, post-processing model*). Tuyệt đối không được sử dụng thêm bất kì nguồn dữ liệu khác để đảm bảo tính công bằng giữa các đội thi. Mọi vi phạm sử dụng thêm dữ liệu huấn luyện sẽ hủy toàn bộ kết quả của nhóm khi Ban tổ chức công bố kết quả.
3. Các đội được chấm qua vòng sơ khảo sau khi nộp kết quả đánh giá trên dữ liệu kiểm tra do Ban tổ chức công bố sẽ phải nộp báo cáo chi tiết về kiến trúc mô hình, quy trình phát triển mô hình dưới dạng báo cáo kĩ thuật (*technical report*) và chuẩn bị bài trình chiếu trình bày tại hội thảo. Nội dung báo cáo cần bao gồm các thông tin sau:
 - o Kiến trúc của các mô hình đã sử dụng;
 - o Chi tiết các tham số đã sử dụng khi phát triển mô hình (hyperparameter, ví dụ như số DNN layer, số nodes trên mỗi layer, learning rate, normalization methods, số epoch training...);
 - o Chi tiết các tham số đánh giá model: perplexity với language model, final training loss với DNN model, final log-likelihood với Kaldi-based model.
4. Đóng gói và chuẩn bị sẵn sàng API đã tích hợp mô hình được phát triển của nhóm. Ban tổ chức sẽ yêu cầu demo nhận dạng lại một số câu phát âm ngẫu nhiên tại hội thảo.

II. Điều khoản sử dụng dữ liệu

Dữ liệu phát triển cho mô hình nhận dạng tiếng nói VLSP-2020-ASR-Training chỉ được sử dụng cho mục đích nghiên cứu và công bố nghiên cứu khoa học. Nếu sử dụng bộ dữ liệu này các nhóm cần tuân thủ các quy định dưới đây:

1. Khi công bố nghiên cứu, kết quả thử nghiệm mà sử dụng bộ dữ liệu này thì nhóm nghiên cứu phải tham chiếu đến website và mô tả bộ dữ liệu được sở hữu và phát triển bởi VLSP. Bất kì thay đổi nào trên bộ dữ liệu này cũng vẫn phải tham chiếu đến thông tin của VLSP.
2. Không được sử dụng cho mục đích thương mại dưới mọi hình thức.
3. Các bộ dữ liệu có được từ việc chỉnh sửa, thay đổi thông số trên bộ dữ liệu này cũng phải tuân theo các điều khoản như trên.

III. Cách thức nhận dữ liệu

Đại diện nhóm kí bản in của mẫu này và gửi đến địa chỉ huynguyen@tnut.edu.vn. Sau đó ban tổ chức sẽ gửi liên kết tải dữ liệu xuống hợp lệ cho đại diện nhóm.

Tên nhóm:

Đại diện nhóm:

Email:

Điện thoại:

Tôi thay mặt nhóm cam đoan thực hiện đúng theo các điều khoản ở trên.

....., ngày tháng năm 2020
(kí và ghi rõ họ tên)