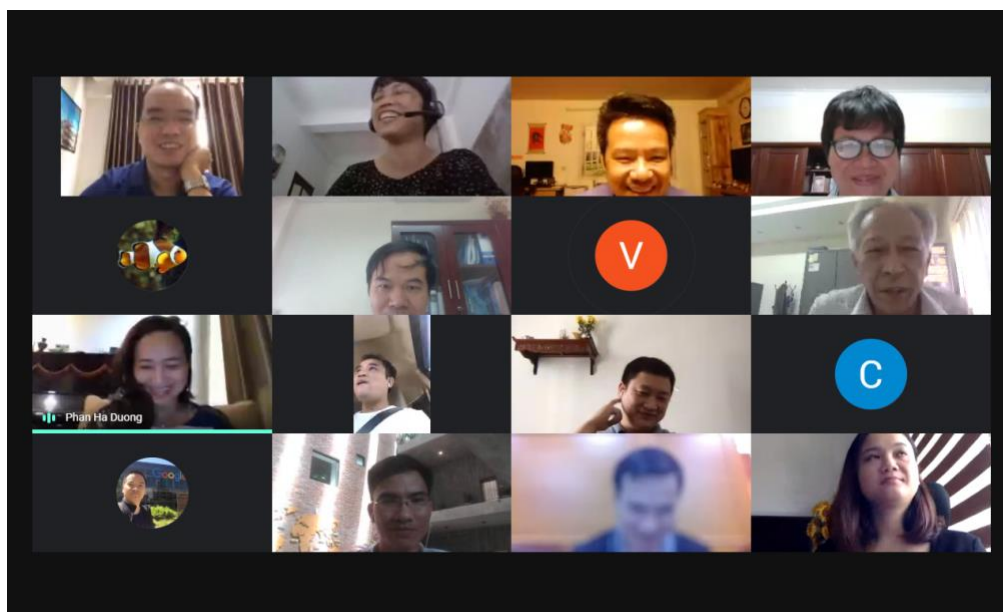


## BÁO CÁO HOẠT ĐỘNG

Câu lạc bộ Xử lý Ngôn ngữ và Tiếng nói Tiếng Việt (gọi tắt theo viết tắt tiếng Anh của Vietnamese Language and Speech Processing là VLSP), chi hội của Hội Tin học Việt Nam được chính thức thành lập thông qua Đại hội thành lập được tổ chức hoàn toàn trực tuyến vào ngày 27/6/2020.



**Họp trừ bị đại hội thành lập VLSP**



**Khách mời tham dự đại hội thành lập VLSP**

Đại hội thành lập CLB VLSP là kết quả của hơn 15 năm hình thành và phát triển của cộng đồng Xử lý ngôn ngữ tự nhiên ở Việt Nam, nơi tập hợp các nhóm nghiên cứu thuộc cả hai khối hàn lâm và công nghiệp trong lĩnh vực xử lý ngôn ngữ và tiếng nói tiếng Việt. Các nhóm nghiên cứu trong cộng đồng Xử lý ngôn ngữ tự nhiên tiếng Việt đã cùng nhau tổ chức hội thảo thường niên kể từ năm 2012, với sự hỗ trợ ban đầu của chuỗi hội nghị IEEE-RIVF.

CLB VLSP được các thành viên tự nguyện thành lập trên cơ sở tiếp thu và chia sẻ các tư tưởng của cộng đồng ngôn ngữ học tính toán thế giới, với mong muốn là nơi kết nối các nhóm nghiên cứu, phát triển về xử lý tiếng Việt, để có tiếng nói chung tham gia vào các hoạt động nhằm đẩy mạnh sự phát triển của nghiên cứu và ứng dụng trong lĩnh vực Xử lý ngôn ngữ tự nhiên nói riêng và Trí tuệ nhân tạo nói chung ở Việt Nam cũng như trên thế giới. Hoạt động chính của CLB VLSP cho đến nay là tổ chức xây dựng xây dựng các bộ dữ liệu chuẩn phục vụ nghiên cứu, đánh giá các hệ thống xử lý ngôn ngữ tự nhiên thông qua các cuộc thi tại hội thảo thường niên VLSP.

Ngay trong năm thành lập 2020, CLB VLSP đã tổ chức 2 sự kiện. Ngày 24-26/10/2020, VLSP tổ chức hội nghị PACLIC (Pacific Asia Conference on Language, Information and Computation) lần thứ 34. Ngày 18/12/2020, hội thảo VLSP lần thứ 7 được tổ chức tại Trường Đại học Bách khoa Hà Nội.



Năm 2021, VLSP lần thứ 8 được Trường ĐH Công nghệ thông tin, ĐHQG TPHCM tổ chức trực tuyến vào ngày 18/12/2021.



Ngày 26/11/2022, VLSP lần thứ 9 được tổ chức tại Viện Nghiên cứu cao cấp về Toán, kết hợp hình thức trực tuyến. Cùng với hội thảo VLSP, CLB cũng đứng ra tổ chức Hội nghị quốc tế

Oriental COCOSDA (International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques) lần thứ 25 (từ 24-26/11).



Các cuộc thi tại VLSP được tổ chức bởi các nhà khoa học Việt Nam ở trong và ngoài nước (Việt Nam, Nhật, Úc, Đức, Singapore, Thụy Điển, Mỹ).

Nhiều nội dung thi được tổ chức trên nền tảng trực tuyến aihub.ml trên máy chủ do công ty INT2 cung cấp.

The image shows a screenshot of the AI HUB website. At the top, there is a navigation bar with links for "Search Competitions", "My Competitions", "Help", "Sign Up", and "Sign In". Below the navigation bar is a banner with the AI HUB logo, which consists of a lightbulb with puzzle pieces inside, and the text "AI HUB" in large green letters, followed by "An Incubator for AI Lovers". Below the banner, there is a white box containing a list of challenges for VLSP 2022. The list includes: "Welcome all participants to register challenges of VLSP 2022:" followed by "VLSP 2022 Evaluation Campaign [link]"; "Join legacy versions of previous challenges:" followed by "Mobile-Captured Image Document Recognition for Vietnamese Receipts (MC-OCR) - Legacy [Link]", "Legacy version of HSD2019-SHARED Task - Hate Speech Detection on Social Networks [Link]", and "VLSP2021 vieCap4H Challenge: Automatic image caption generation for healthcare domains in Vietnamese [Link]".

Các nội dung thi do CLB VLSP tổ chức bao gồm các vấn đề nền tảng của VLSP và các bài toán ứng dụng thực tế có tính hấp dẫn cao. Mỗi năm, có hàng trăm đội thi đến từ các trường đại học trong nước và cả ngoài nước cũng như các doanh nghiệp.

Chẳng hạn, VLSP 2022 có các đội thi đạt giải cao đến từ Trường ĐH Công nghệ, Trường ĐHKHTN thuộc ĐHQGHN, Trường ĐH CNTT và Trường ĐHKHTN thuộc ĐHQG TPHCM, Trường ĐH Bách khoa HN, Học viện Bru chính Viễn thông, ĐH Thái Nguyên, ĐH Bách khoa Tây Bắc (NPU) Trung Quốc, Viện nghiên cứu Xử lý tiếng nói Ấn Độ, JAIST, Trường ĐH Stanford. Khối doanh nghiệp cũng tham gia tích cực với các đội thi đạt giải đến từ Tập đoàn Viettel, Trung tâm không gian mạng Viettel, VCCorp, VinBigData, VNPT-IT, VNG, InterITS (Smartcall), Viện nghiên cứu ứng dụng công nghệ CMC, Samsung SDS R&D Center.

Các cuộc thi VLSP đã tổ chức:

Về xử lý văn bản:

1. Phân tách từ và gán nhãn từ loại tiếng Việt: VSLP 2013
2. Phân tích cú pháp phụ thuộc và cú pháp thành phần: VLSP 2019, 2020
3. Dịch máy Anh - Việt, Trung - Việt: VLSP 2013, 2020, 2022
4. Nhận dạng thực thể có tên: VLSP 2016, 2018, 2021
5. Trích rút quan hệ: VLSP 2021
6. Phân tích tình cảm: VLSP 2016, 2018
7. Phát hiện phát ngôn thù hận (*Hate Speech Detection*): VLSP 2020
8. Đọc hiểu tiếng Việt tự động (*Vietnamese Machine Reading Comprehension*): VLSP 2021
9. Xác định mối quan hệ ngữ nghĩa kéo theo giữa các cặp câu tiếng Việt và tiếng Anh – tiếng Việt (*Vietnamese and English-Vietnamese Textual Entailment*): VLSP 2021
10. Sinh chú thích cho ảnh (*Image Captioning*): VLSP 2021
11. Tóm tắt tóm lược đa văn bản tiếng Việt: VLSP 2022
12. Hỏi đáp đa ngữ về nội dung ảnh: VLSP 2022

Về xử lý tiếng nói:

13. Nhận dạng giọng nói tự động cho tiếng Việt trong hội thoại và bài giảng trực tuyến: VLSP 2018, 2019, 2020, 2021, 2022
14. Tổng hợp tiếng nói tiếng Việt dữ liệu thường và dữ liệu có cảm xúc: VLSP 2018, 2019, 2020, 2021, 2022
15. Xác minh người nói đơn ngữ (tiếng Việt) và đa ngữ (các ngôn ngữ châu Á): VLSP 2021, 2022

Hoạt động của CLB VLSP nhận được sự ủng hộ từ rất nhiều đơn vị. Có thể kể tới các trường đại học như VNU-HUS, VNU-UET, VNUHCM-UIT, HUST, PTIT, TLU, USTH, ... cùng với các doanh nghiệp lớn nhỏ như Aimesoft, Dageoras, INT2, ReML.AI, SMCC, VAIS, VBee, VCCorp, Vietlex, VTTC (Viettel Cyberspace Center), Quỹ VINIF, Zalo-VNG.

Các hoạt động xuất bản của VLSP cũng được sự hỗ trợ của các tạp chí khoa học trong nước. Năm 2021, các báo cáo khoa học có chất lượng từ các cuộc thi được chọn đăng trên Tạp chí CNTT&TT của VNU. Năm 2022, các báo cáo hội thảo sẽ được chọn đăng trên Tạp chí Tin học và Điều khiển.

Trong thời gian tới, CLB VLSP tiếp tục tổ chức các hoạt động liên quan tới dữ liệu và đánh giá liên quan tới xử lý ngôn ngữ và tiếng nói tiếng Việt, đồng thời tăng cường hợp tác quốc tế trong lĩnh vực này thông qua các hội nghị, hội thảo quốc tế được tổ chức ở Việt Nam. CLB đã và luôn khẳng định uy tín của mình trong các hoạt động đánh giá khách quan và tuân theo các chuẩn mực quốc tế.