

THAM LUẬN

Số hóa và xây dựng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt -Hrê, Việt –Co

CN. Trương Thị Diệu Thủy

Trưởng phòng Kỹ thuật và Hạ tầng số,

Trung tâm Công nghệ thông tin và Truyền thông Quảng Ngãi, Thư ký Đề tài

Tóm tắt: Trình bày các nội dung số hóa và xây dựng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt -Hrê, Việt –Co trong phạm vi nội dung các công việc thực hiện của Đề tài, làm rõ quá trình thực hiện, một số khó khăn vướng mắc và giải pháp khắc phục, kết quả đạt được.

Từ khóa: số hóa, cơ sở dữ liệu điện tử, tiếng đồng bào dân tộc thiểu số, Hrê, Co.

1. GIỚI THIỆU

1.1. Thông tin về Đề tài

Xuất phát từ nhu cầu thực tiễn ứng dụng công nghệ số vào phục vụ các lĩnh vực của đời sống, đặc biệt là góp phần bảo tồn, lưu giữ ngôn ngữ tiếng đồng bào dân tộc thiểu số của tỉnh Quảng Ngãi, Trung tâm Công nghệ thông tin và Truyền thông Quảng Ngãi đã đề xuất nghiên cứu, xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số tỉnh Quảng Ngãi (mà cụ thể sẽ xây dựng cho tiếng Việt – Hrê, Việt – Co) nhằm thu hẹp khoảng cách giao tiếp, ngôn ngữ, giúp cho người đồng bào dân tộc thiểu số học hỏi, nâng cao kiến thức, hòa nhập với sự phát triển chung của tỉnh, của đất nước; đồng thời cũng giúp cho các cá nhân người Kinh đang làm việc với người đồng bào có thể hiểu và giao tiếp thuận lợi hơn, qua đó truyền đạt các kiến thức, các chủ trương, chính sách của chính quyền đến người đồng bào, tiếp thu kinh nghiệm và các vấn đề của người đồng bào được sâu sắc hơn.

Đề tài khoa học công nghệ cấp tỉnh “Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co” đã được phê duyệt tại Quyết định số 2109/QĐ-UBND ngày 23/12/2021 của Chủ tịch UBND tỉnh Quảng Ngãi về việc phê duyệt danh mục nhiệm vụ khoa học và công nghệ cấp tỉnh triển khai

thực hiện năm 2022:

- Tên đề tài: **Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co**
- Cấp quản lý: **Cấp tỉnh**
- Thời gian thực hiện: **18 tháng** (Từ tháng 02/2023 đến tháng 08/2024)
- Tổ chức chủ trì đề tài: **Trung tâm Công nghệ Thông tin và Truyền thông Quảng Ngãi**
- Tổ chức phối hợp chính thực hiện đề tài: **Ban Dân tộc tỉnh Quảng Ngãi**
- Chủ nhiệm đề tài: **ThS. Trần Duy Linh – ThS. Trần Văn Mẫn**

1.2. Mục tiêu thực hiện của Đề tài

Đề tài **Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co** thực hiện mục tiêu nghiên cứu, xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co và ngược lại nhằm hỗ trợ công tác giảng dạy, học tập, nghiên cứu về tiếng nói, chữ viết của người Hrê, người Co được thuận lợi, góp phần thu hẹp khoảng cách phát triển về kinh tế, văn hoá, xã hội giữa các vùng dân tộc thiểu số với các địa bàn trong tỉnh, đặc biệt là khoảng cách về giao tiếp giữa đồng bào các dân tộc; thúc đẩy phát triển kinh tế, xã hội; gìn giữ và giới thiệu bản sắc văn hóa, phong tục tập quán của người đồng bào dân tộc Hrê, Co.

Đề tài **Xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co** thực hiện và hoàn thành **bốn nhiệm vụ lớn** đặt ra, gồm:

Nhiệm vụ thứ nhất: Xây dựng hệ thống dữ liệu - kho ngữ vựng Việt – Hrê, Việt – Co: Trên cơ sở nguồn dữ liệu tiếng Hrê và tiếng Co tại 02 bộ tài liệu đã được UBND phê duyệt: “Tài liệu Đào tạo, bồi dưỡng tiếng Hrê” (*dùng cho cán bộ công chức tại miền núi và công tác dân tộc tỉnh Quảng Ngãi*) của UBND tỉnh Quảng Ngãi (*ban hành tạm thời kèm theo Quyết định số 893/QĐ-UBND ngày 27/4/2007 của UBND tỉnh Quảng Ngãi*) và Tài liệu “BAI HOC APOK KOOL (BÀI HỌC TIẾNG CO)” (*được UBND tỉnh cho phép tại công văn số 4796/UBND-NC ngày 27/8/2019*), Đề tài sẽ tiếp cận và nghiên cứu toàn bộ vốn từ Việt, Hrê và Co cơ bản thường xuyên sử dụng trong cuộc sống trong 02 bộ tài liệu. Từ đó, thực hiện việc số hóa đồng nhất giữa các từ tương ứng để thành lập kho dữ liệu Việt – Hrê, Việt – Co, gồm:

- Kho ngữ vựng song ngữ Việt – Hrê và ngược lại;
- Kho ngữ vựng song ngữ Việt – Co và ngược lại;

Nhiệm vụ thứ hai: Xây dựng công cụ hỗ trợ gõ tiếng dân tộc (Hrê, Co) trên giao diện phần mềm: Trong phạm vi kết quả của đề tài bao gồm: Cơ sở dữ liệu điện tử Việt – Hrê và ngược lại (Hrê – Việt); Cơ sở dữ liệu điện tử Việt – Co và ngược lại (Co – Việt). Đối với tiếng Việt đã có các công cụ hỗ trợ gõ tiếng Việt tích hợp font Unicode nên việc sử dụng để tra từ trên giao diện phần mềm dễ dàng. Đối với tiếng Hrê và Co, hiện chưa có công cụ hỗ trợ gõ có tích hợp font Unicode, vì vậy để tra cứu từ Hrê – Việt, Co – Việt phải nghiên cứu và xây dựng công cụ cho phép gõ tiếng Hrê, Co bằng font Unicode trực tiếp trên giao diện phần mềm.

Nhiệm vụ thứ ba: Xây dựng bộ phần mềm Cơ sở dữ liệu điện tử Việt – Hrê, Việt – Co: Nhờ vào kho dữ liệu đã xây dựng, đề tài sẽ phát triển bộ phần mềm Cơ sở dữ liệu điện tử có thể sử dụng được trên các thiết bị điện tử như: máy tính, điện thoại thông minh,... nhằm mục đích tra ngữ nghĩa giữa các ngôn ngữ Việt – Hrê, Việt – Co. Bên cạnh đó, phần mềm còn trang bị thêm một kênh tương tác để người dùng có thể đóng góp thêm một số từ chưa có trong cơ sở dữ liệu.

Nhiệm vụ thứ tư: Ứng dụng trí tuệ nhân tạo (AI) để xây dựng hệ thống nhận diện giọng nói tiếng dân tộc và tích hợp vào hệ thống CSDL điện tử: Hiện nay, trí tuệ nhân tạo (AI) đã có nhiều ứng dụng rộng rãi trong nhiều lĩnh vực, trong đó có lĩnh vực từ điển điện tử với công nghệ nhận diện giọng nói (SpeechToText). Tuy nhiên, việc nhận diện giọng nói là tiếng đồng bào dân tộc thiểu số Hrê, Co thì chưa có trước đây nên đây là một nghiên cứu mới, vừa mang lại hiệu quả thiết thực với sản phẩm của Đề tài và sẽ nền tảng cho các nghiên cứu khác liên quan đến ứng dụng công nghệ vào lĩnh vực ngôn ngữ. Nhóm thực hiện cần phải nghiên cứu, thử nghiệm các mô hình huấn luyện dữ liệu nhận diện tiếng Hrê, Co, chọn một mô hình khả thi, hiệu quả nhất để tích hợp kết quả huấn luyện vào các sản phẩm phần mềm (phiên bản web và phiên bản app), cho phép tra cứu bằng âm thanh, giọng nói.

Ngoài ra, còn có các báo cáo khoa học và công nghệ theo từng chuyên đề nghiên cứu, các bài báo khoa học công nghệ được đăng tải trên các tạp chí uy tín của ngành, phim tư liệu để quảng bá sản phẩm.

1.3. Các nội dung công việc của Đề tài

Theo Thuyết minh Đề tài được phê duyệt, gồm có 09 nội dung công việc:

Nội dung 1: Nghiên cứu tổng quan: gồm 04 công việc nghiên cứu nhằm tạo cơ sở lý thuyết cho thực hiện Đề tài.

Nội dung nghiên cứu bao gồm: Nghiên cứu về công nghệ xây dựng Cơ sở dữ liệu điện tử; Nghiên cứu tổng quan về ngôn ngữ, ngôn ngữ Hrê và ngôn ngữ Co; Nghiên cứu về bộ gõ tiếng Việt và các bộ gõ tiếng đồng bào dân tộc thiểu số, nghiên cứu phương pháp để tích hợp bộ mã Unicode cho tiếng Hrê, Co; Nghiên cứu về công nghệ trí tuệ nhân tạo AI nhận diện giọng nói.

Nội dung 2: Xây dựng công cụ hỗ trợ gõ tiếng đồng bào dân tộc thiểu số Hrê, Co.

Nội dung này gồm 02 công việc là xây dựng công cụ hỗ trợ gõ tiếng đồng bào dân tộc thiểu số Hrê và tiếng Co.

Nội dung 3: Xây dựng kho ngữ vựng song ngữ Hrê-Việt, Việt-Hrê, Co-Việt, Việt-Co

Nội dung này gồm 04 công việc, bao gồm: Nhóm nghiên cứu đi khảo sát tại các huyện, phối hợp với phòng Dân tộc huyện để thu thập thông tin, dữ liệu từ vựng tiếng đồng bào dân tộc thiểu số Hrê, Co thông qua các hình thức: thu thập các nguồn tài liệu đã có và đồng thời đi khảo sát trực tiếp tại các địa phương có người đồng bào dân tộc thiểu số Hrê và người Co sinh sống, gồm các huyện: Ba Tơ, Minh Long, Sơn Hà, Sơn Tây và Trà Bồng.

Từ kết quả của việc khảo sát thu thập là thông tin, dữ liệu về từ vựng tiếng đồng bào dân tộc thiểu số Hrê, Co, nhóm nghiên cứu đề tài thực hiện số hóa các dữ liệu, xây dựng kho ngữ vựng song ngữ Hrê-Việt, Việt-Hrê và Co-Việt, Việt-Co.

Nội dung 4: Xây dựng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt-Hrê và Việt-Co.

Trên cơ sở các nghiên cứu, thông tin, dữ liệu đã hình thành tại các công việc trước, tiến hành xây dựng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt-Hrê và Việt-Co theo quy trình phát triển phần mềm tạo ra sản phẩm phần mềm có đầy đủ các chức năng và yêu cầu phi chức năng theo Thuyết minh Đề tài đã được phê duyệt. Quy trình gồm các bước: Xác định các yêu cầu của

Cơ sở dữ liệu điện tử; Phân tích, thiết kế Cơ sở dữ liệu điện tử; Xây dựng Cơ sở dữ liệu của Cơ sở dữ liệu điện tử; Thiết kế, hiệu chỉnh giao diện của Cơ sở dữ liệu điện tử; Lập trình, xây dựng các mô-đun chức năng của Cơ sở dữ liệu điện tử; Kiểm thử, hiệu chỉnh Cơ sở dữ liệu điện tử; Cập nhật dữ liệu vào Cơ sở dữ liệu điện tử.

Kết quả của nội dung công việc này là sản phẩm cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt-Hrê và Việt-Co phiên bản web.

Nội dung 5: Xây dựng app hoạt động trên các thiết bị di động

Thực hiện xây dựng app hoạt động trên các thiết bị di động gồm các bước theo quy trình xây dựng app, bao gồm: Thiết kế, hiệu chỉnh giao diện app; Lập trình chức năng app (bao gồm app trên Android và iOS); Kiểm thử và hiệu chỉnh app; Cài đặt app trên Android và iOS.

Kết quả của nội dung công việc này app Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co được cài đặt trên kho ứng dụng CHPlay/Google Play, AppStore.

Nội dung 6: Ứng dụng công nghệ AI - trí tuệ nhân tạo xây dựng mô-đun nhận diện giọng nói

Nội dung thực hiện gồm 06 công việc, bao gồm: Thu thập bộ mẫu phát âm tiếng Hrê, Co; Dựng hệ thống máy chủ (server) và các máy trạm, tiến hành cài đặt, cấu hình hệ thống huấn luyện (train) dữ liệu mẫu; Tiến hành xử lý mẫu và huấn luyện (train) dữ liệu mẫu; Đánh giá mô hình xử lý dữ liệu đã thu thập (xử lý file âm thanh) và tiến hành huấn luyện dữ liệu theo mô hình được chọn; Xây dựng và tích hợp mô-đun nhận diện tiếng Hrê, Co vào Cơ sở dữ liệu điện tử.

Nội dung 7: Ứng dụng cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt-Hrê và Việt-Co vào thực tiễn

Để sản phẩm của Đề tài được ứng dụng một cách hiệu quả, mang lại những lợi ích thiết thực cho cộng đồng, nhóm nghiên cứu Đề tài sẽ xây dựng và đề xuất mô hình ứng dụng Cơ sở dữ liệu điện tử vào thực tiễn. Trên cơ sở nhóm nghiên cứu Đề tài phối hợp với Ban dân tộc tỉnh triển khai mô hình ứng dụng Cơ sở dữ liệu điện tử vào thực tiễn, với mục tiêu đưa kết quả của Đề tài ứng dụng có hiệu quả vào thực tiễn, mang lại những lợi ích thiết thực cho xã hội, cộng đồng.

Nội dung 8: Đào tạo, tập huấn, hội thảo quảng bá kết quả

Vì Đề tài xây dựng sản phẩm hướng đến đông đảo người sử dụng, nên cần tổ chức hội thảo và tập huấn để chuyển giao và hướng dẫn sử dụng cho các đối tượng sẽ tham gia sử dụng CSDL điện tử này, cụ thể: Tổ chức 01 hội thảo cấp tỉnh để giới thiệu về kết quả của Đề tài; Tổ chức 05 lớp tập huấn hướng dẫn quản lý và hướng dẫn khai thác CSDL điện tử cho các đối tượng sử dụng và 01 lớp quản trị, vận hành hệ thống.

Nội dung 9: Viết bài báo khoa học, xây dựng phim tư liệu và quảng bá sản phẩm; viết báo cáo tổng hợp

Nội dung thực hiện gồm: Tổng hợp nội dung, viết 01 bài báo khoa học về việc xây dựng kho ngữ vựng song ngữ Việt – Hrê, Hrê – Việt, Việt – Co, Co – Việt, 01 bài báo khoa học về việc xây dựng công nghệ AI nhận diện giọng nói vào xây dựng Cơ sở dữ liệu điện tử tiếng đồng bào Việt – Hrê, Việt – Co. Kết quả là 01 bài báo khoa học được đăng trên Tạp chí Khoa học và Công nghệ Việt Nam – Bộ KH&CN hoặc Tạp chí Thông tin và Truyền thông – Bộ TT&TT và 01 bài báo khoa học được đăng trên tạp chí Khoa học và Công nghệ Việt Nam – Bộ KH&CN hoặc Tạp chí Thông tin và Truyền thông.

Đồng thời, tổng hợp nội dung, xây dựng 01 phim tư liệu và viết tin bài quảng bá sản phẩm quảng bá trên các nền tảng số như: website, kênh zalo của Trung tâm Công nghệ thông tin và Truyền thông, Sở Thông tin và Truyền thông, Ban Dân tộc tỉnh; Cổng thông tin điện tử tỉnh Quảng Ngãi, Cổng thông tin điện tử Bộ Thông tin và Truyền thông để sản phẩm CSDL điện tử được biết và sử dụng trong đông đảo người dân có nhu cầu.

Cuối cùng là viết các báo cáo chính và báo cáo tóm tắt đề tài theo mẫu quy định.

2. NỘI DUNG THỰC HIỆN

2.1. Số hóa nguồn dữ liệu

Theo Luật Công nghệ thông tin năm 2006, "*số hóa*" là việc biến đổi các loại hình thông tin sang thông tin số.

Trong phạm vi thực hiện Đề tài, loại hình thông tin thu thập được làm cơ sở cho việc số hóa là **nguồn dữ liệu dạng giấy**: 02 cuốn giáo trình đã được UBND tỉnh Quảng Ngãi phê duyệt và đang được sử dụng để giảng dạy tại trường Đại học Phạm Văn Đồng:

Tài liệu Đào tạo, bồi dưỡng tiếng Hrê (dùng cho cán bộ công chức tại miền núi và công tác dân tộc tỉnh Quảng Ngãi) của UBND tỉnh Quảng Ngãi (ban hành tạm thời kèm theo Quyết định số 893/QĐ-UBND ngày 27/4/2007 của UBND tỉnh Quảng Ngãi). Ban biên soạn: Chủ biên: Đỗ Ảnh; cộng tác viên chính: Trương Hồng Hoà, Đinh Cơi, Nga Ri Vê.

Tài liệu “BAI HOC APOK KOOL (BÀI HỌC TIẾNG CO)” do nhóm biên soạn: Nguyễn Minh Trí, Tạ Văn Thông, Hồ Văn Năm, Tạ Quang Tùng thực hiện. Tài liệu này là một công trình thuộc Đề tài KHCN cấp tỉnh: “Nghiên cứu, biên soạn tài liệu dạy – học tiếng Co cho cán bộ, công chức (người Kinh) công tác tại huyện Trà Bồng và huyện Tây Trà, tỉnh Quảng Ngãi” do chủ nhiệm đề tài ThS. Nguyễn Minh Trí thực hiện năm 2016.

Phương pháp thực hiện số hóa là sử dụng các công cụ cần thiết để tiến hành số hóa gồm nhập lại văn bản (gồm từ và câu) vào máy vi tính từ 02 bộ giáo trình nêu trên, sau đó tiến hành xử lý dữ liệu, rồi tiến hành thu âm phát âm của các từ và câu.

Công cụ để tiến hành số hóa là máy vi tính, thiết bị ghi âm, phần mềm xử lý văn bản và phần mềm xử lý âm thanh. Bên cạnh đó, để nhập tiếng Hrê, Co thì cần phải có bộ gõ tiếng đồng bào dân tộc thiểu số Hrê, Co. Ngoài ra, cần có công cụ số để lưu trữ kết quả là dữ liệu số sau khi hoàn thành, là thiết bị lưu trữ (USB, ổ cứng) hoặc ổ đĩa mạng (Drive hoặc Cloud).

Dạng thông tin số được biến đổi thành là dữ liệu số, gồm dạng file văn bản lưu dữ liệu văn bản (text), và dạng file âm thanh lưu dữ liệu âm thanh (phát âm từ và câu).

Kết quả, đã số hóa được trên 3.349 từ tiếng Hrê và tiếng Co, 671 câu tiếng Hrê và tiếng Co:

SỐ LƯỢNG TỪ, CÂU TIẾNG HRÊ, CO ĐÃ THU THẬP VÀ SỐ HÓA DẠNG VĂN BẢN (TEXT)		
	Tiếng Hrê	Tiếng Co
Từ	1.934	1.409
Câu	602	69
Tổng cộng	3.349	671

Đồng thời, đã thu âm các đọc của các từ và câu tiếng Hrê, Co ở hai giọng đọc nam và nữ, với tổng số file âm thanh đã số hóa là trên 7.000 file:

SỐ LƯỢNG TỪ, CÂU TIẾNG HRÊ, CO ĐÃ THU THẬP VÀ SỐ HÓA DẠNG ÂM THANH (AUDIO)		
Tổng số file âm thanh thô tiếng Hrê:		5.072 file
Trong đó:		
Từ tiếng Hrê	Giọng nam:	1.934 file
	Giọng nữ:	1.934 file
Câu tiếng Hrê	Giọng nam:	602 file
	Giọng nữ:	602 file
Tổng số file âm thanh thô tiếng Co:		2.887 file
Trong đó:		
Từ tiếng Co	Giọng nam:	1.409 file
	Giọng nữ:	1.409 file
Câu tiếng Co	Giọng nam:	69 file
	Giọng nữ:	0 file

2.2. Xây dựng Cơ sở dữ liệu điện tử

Theo từ điển tiếng Việt của Trung tâm Từ điển học – Viện Ngôn ngữ học, xuất bản năm 2002, "**cơ sở dữ liệu**" là tập hợp dữ liệu được tổ chức sao cho các chương trình máy tính có thể khai thác, sửa đổi thông tin từ các dữ liệu đó.

Trong lĩnh vực công nghệ thông tin, "**cơ sở dữ liệu**" được hiểu là là một hệ thống các thông tin có cấu trúc, được lưu trữ trên các thiết bị lưu trữ nhằm thỏa mãn yêu cầu khai thác thông tin đồng thời của nhiều người sử dụng hay nhiều chương trình ứng dụng chạy cùng một lúc với những mục đích khác nhau.

Khái niệm này có liên quan chặt chẽ đến khái niệm "**phần mềm**" là chương trình máy tính được mô tả bằng hệ thống ký hiệu, mã hoặc ngôn ngữ để điều khiển thiết bị số thực hiện chức năng nhất định (*theo Luật Công nghệ thông tin năm 2006*).

"**Cơ sở dữ liệu điện tử**" được hiểu là Cơ sở dữ liệu tồn tại ở dạng kỹ thuật số và có thể được truy cập thông qua một số phương tiện khác nhau. Cơ sở dữ liệu điện tử có thể được tìm thấy ở nhiều dạng, bao gồm phần mềm được cài đặt trên máy tính bảng hoặc máy tính để bàn, ứng dụng di động, ứng dụng web.

Trong phạm vi của Đề tài, "**Cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số**", Cơ sở dữ liệu điện tử, hay Cơ sở dữ liệu điện tử (viết tắt là CSDL điện tử) được hiểu là một phần mềm máy tính có chức năng quản lý, cho phép tìm kiếm tra cứu dữ liệu tiếng đồng bào dân tộc thiểu số, được cài đặt và sử dụng trên các thiết bị số.

Xây dựng CSDL điện tử là quá trình thực hiện các công việc theo quy trình phát triển phần mềm để tạo ra sản phẩm phần mềm có đầy đủ các chức năng và yêu cầu phi chức năng theo Thuyết minh Đề tài đã được phê duyệt.

Quy trình phát triển phần mềm gồm các bước:

- Xác định các yêu cầu của Cơ sở dữ liệu điện tử;
- Phân tích, thiết kế Cơ sở dữ liệu điện tử;
- Xây dựng Cơ sở dữ liệu của Cơ sở dữ liệu điện tử;
- Thiết kế, hiệu chỉnh giao diện của Cơ sở dữ liệu điện tử;
- Lập trình, xây dựng các mô-đun chức năng của Cơ sở dữ liệu điện tử;
- Kiểm thử, hiệu chỉnh Cơ sở dữ liệu điện tử;
- Cập nhật dữ liệu vào Cơ sở dữ liệu điện tử;

Kết quả là sản phẩm cơ sở dữ liệu điện tử tiếng đồng bào dân tộc thiểu số Việt-Hrê và Việt-Co phiên bản web và phiên bản app.

Phiên bản web của CSDL điện tử đã hoàn thành, được cài đặt và hoạt động trên môi trường mạng tại tên miền: <https://csdlhresco.nuian.vn>, phục vụ tra cứu tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co và ngược lại. Các từ đều được dịch nghĩa, hướng dẫn phát âm, cung cấp các từ đồng nghĩa, ví dụ minh họa, có hình ảnh về phong tục, tập quán, văn hoá của người đồng bào dân tộc thiểu số Hrê, Co. CSDL điện tử được thiết kế linh hoạt, cho phép khai thác và sử dụng trên nhiều hệ thống: máy vi tính, điện thoại thông minh, máy tính

bảng; cung cấp đầy đủ chức năng để tra cứu từ thông qua bộ gõ ký tự hoặc giọng nói với chức năng thêm mới, cập nhật, chỉnh sửa từ ngữ, cập nhật hình ảnh; đồng thời có chức năng cho phép tiếp nhận phản hồi và tương tác với người dùng qua chức năng “Đóng góp cho CSDL”.

Đồng thời, phiên bản app của CSDL điện tử đã được hoàn thành và cài đặt trên các kho ứng dụng của Google Play và App Store, có thể tìm kiếm với tên app là “Hre – Co – Việt” và tải về cài đặt cho các dòng điện thoại sử dụng hệ điều hành iOS và Android.

CƠ SỞ DỮ LIỆU ĐIỆN TỬ TIẾNG ĐỒNG BÀO DÂN TỘC THIỂU SỐ VIỆT – HRÊ, VIỆT – CO		
PHIÊN BẢN APP CHO ANDROID <i>trên Google Play</i>	PHIÊN BẢN WEB <i>https://csdllhrecovietnam.vn</i>	PHIÊN BẢN APP CHO IOS <i>trên App Store</i>
		

Mã QR để truy cập khai thác các sản phẩm kết quả

3. MỘT SỐ KHÓ KHĂN, VƯỚNG MẮC TRONG QUÁ TRÌNH THỰC HIỆN VÀ GIẢI PHÁP KHẮC PHỤC

Vướng mắc trong thủ tục thuê máy chủ, thiết bị phục vụ thu âm:
Thuyết minh Đề tài được xây dựng từ năm 2022. Tuy nhiên, trong quá trình thực hiện, luật Đấu thầu năm 2023 ra đời, với một số quy định mới trong lựa chọn nhà thầu. Việc phê duyệt Kế hoạch lựa chọn nhà thầu theo quy định bị vướng, phải kéo dài thời gian (theo Luật đấu thầu 2023 và tại điểm c, khoản 1, điều 91 Nghị định số 24/2024/NĐ-CP ngày 27/02/2024 Quy định chi tiết một số điều và biện pháp thi hành Luật đấu thầu về lựa chọn nhà thầu). Vì vậy để kịp thời có thiết bị phục vụ các công việc của Đề tài, Tổ chức chủ trì đã có phương án tạm thời mượn thiết bị, hạ tầng hiện có để tiếp tục thực hiện các nội dung công việc thuộc Đề tài.

Khó khăn trong thu thập dữ liệu, số hóa, ghi âm: qua quá trình khảo sát,

thu thập dữ liệu tại các huyện Ba Tơ, Minh Long, Sơn Hà, Sơn Tây, Trà Bồng, nhóm thực hiện nhận thấy đối với tiếng Hrê được sử dụng bởi người đồng bào dân tộc Hrê và Ca đông, phân bố chủ yếu tại các huyện Ba Tơ, Minh Long, Sơn Hà, Sơn Tây. Tuy nhiên, cách đọc tại các vùng miền có sự khác nhau, cụ thể là ở Quảng Ngãi, có sự phân biệt hai giọng “Co” ở “đường nước” (*truôk đhak*) và “Co” ở “đường rừng” (*truôk gôk*). Để lựa chọn giọng đọc cho việc thu âm, nhóm thực hiện đã tham khảo ý kiến chuyên gia về ngôn ngữ Hrê, Co để chọn giọng ở vùng miền phổ biến nhất.

Khó khăn trong việc nghiên cứu phương pháp nhận diện giọng nói tiếng Hrê, Co để ứng dụng vào đề tài: Nội dung “Ứng dụng công nghệ AI - trí tuệ nhân tạo xây dựng mô-đun nhận diện giọng nói” là nội dung mới, có độ khó, phức tạp cao, có nhiều phương pháp nghiên cứu, tại thời điểm Đề tài thực hiện vẫn đang được các chuyên gia nghiên cứu, triển khai. Đơn vị chủ trì cùng các thành viên thực hiện Đề tài phải trải qua quá trình thu thập mẫu, phân tích, huấn luyện dữ liệu thử nghiệm, đánh giá kết quả, tuy nhiên kết quả những lần đầu không như mong đợi nên thời gian thực hiện bị kéo dài so với tiến độ đặt ra. Nhóm thực hiện đã tìm và tham khảo ý kiến các chuyên gia đầu ngành khác trong lĩnh vực nghiên cứu, ứng dụng công nghệ AI – nhận diện giọng nói ngoài thành viên Đề tài, kết nối và mời tham gia là thực hiện Đề tài. Từ đó, đã giải quyết được khó khăn, đến nay đã thực hiện thành công mô hình huấn luyện dữ liệu tiếng Hrê, Co và đã tích hợp vào sản phẩm, cung cấp chức năng nhận diện giọng nói tiếng Hrê và Co.

Khó khăn trong việc cài đặt app lên các kho ứng dụng của Google Play, App Store: Quá trình mua tài khoản nhà phát triển của các kho ứng dụng để cài đặt, phát hành app lên kho cũng gặp một số khó khăn ban đầu, khi nhóm thực hiện tiến hành đăng ký mua tài khoản nhà phát triển của Apple nhưng không được duyệt. Qua tìm hiểu chính sách của các kho ứng dụng, nhóm đã tuân thủ quy trình, khai báo thông tin đầy đủ theo yêu cầu, thử lại nhiều lần đến khi được duyệt.

4. KẾT LUẬN

Qua thời gian tháng thực hiện, với sự tích cực và nỗ lực, quyết tâm cao, quy tụ nhiều chuyên gia trong lĩnh vực ngôn ngữ lẫn công nghệ số, nhóm thực hiện Đề tài đã hoàn thành các công việc, các nội dung nghiên cứu của Đề tài.

Trong đó, nội dung công việc số hóa và xây dựng CSDL điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co là các nội dung công việc lớn trong các nội dung công việc của Đề tài, quyết định đến sản phẩm cuối cùng cũng đã hoàn thành, góp phần tạo nên thành công chung của Đề tài.

Sản phẩm đã được cài đặt, triển khai trên các nền tảng, cho phép truy cập, khai thác, sử dụng. Đồng thời, sản phẩm cũng đã bước đầu được ứng dụng vào thực tiễn, phục vụ đa dạng các đối tượng người dùng như cán bộ công chức, lực lượng vũ trang, giáo viên đang công tác tại các huyện miền núi có người đồng bào dân tộc thiểu số sinh sống, các nhà nghiên cứu về ngôn ngữ tiếng đồng bào dân tộc thiểu số, người dân và các em học sinh miền núi,... mang lại một số kết quả ban đầu khả quan.

Với xuất phát và mục tiêu phục vụ cộng đồng, nhóm thực hiện Đề tài đã mang lại một phương tiện, công cụ ứng dụng công nghệ số để hỗ trợ thêm cho công tác học tập, tra cứu ngôn ngữ, nhóm thực hiện Đề tài đã hoàn thành các công việc; đồng thời, việc thực hiện cũng mang lại giá trị văn hóa và nhân văn sâu sắc, khi sản phẩm CSDL điện tử tiếng đồng bào dân tộc thiểu số Việt – Hrê, Việt – Co cũng sẽ là nơi lưu giữ, bảo tồn ngôn ngữ như một bản sắc văn hóa của đồng bào dân tộc thiểu số tỉnh Quảng Ngãi cho các đời sau.