

**CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  
**Độc lập - Tự do - Hạnh phúc**

Hà nội, ngày 29 tháng 7 năm 2021

**BÁO CÁO KẾT QUẢ TỰ ĐÁNH GIÁ**  
**NHIỆM VỤ KHOA HỌC VÀ CÔNG NGHỆ CẤP QUỐC GIA**

**I. Thông tin chung về nhiệm vụ:**

1. Tên nhiệm vụ, mã số:

*Nghiên cứu xây dựng hệ dịch tự động văn bản tiếng Việt ra tiếng nói tiếng Mường, hướng đến áp dụng cho các ngôn ngữ dân tộc thiểu số chưa có chữ viết ở Việt Nam - Mã số: ĐTĐLCN.20/17*

Thuộc chương trình: Đề Tài KHCN Độc Lập Cấp Nhà Nước

2. Mục tiêu nhiệm vụ:

- Nghiên cứu, làm chủ công nghệ dịch máy từ văn bản tiếng Việt ra tiếng nói tiếng dân tộc thiểu số không có chữ viết ở Việt Nam.
- Xây dựng được cơ sở dữ liệu song ngữ văn bản tiếng Việt - tiếng nói tiếng Mường.
- Xây dựng hệ thống dịch máy từ văn bản tiếng Việt ra tiếng nói tiếng Mường, với đặc điểm chưa có chữ viết và sự khác biệt về phương/thổ ngữ.
- Phát triển một số sản phẩm sử dụng hệ thống dịch tự động Việt - Mường

3. Chủ nhiệm nhiệm vụ: TS. Mạc Đăng Khoa

4. Tổ chức chủ trì nhiệm vụ: Viện nghiên cứu Quốc tế về Thông tin đa phương tiện, Truyền thông và Ứng dụng (MICA), Trường Đại học Bách Khoa Hà Nội.

5. Tổng kinh phí thực hiện: 4.400 triệu đồng.

Trong đó, kinh phí từ ngân sách SNKH: 4.400 triệu đồng.

Kinh phí từ nguồn khác: 0 triệu đồng.

6. Thời gian thực hiện theo Hợp đồng:

Bắt đầu: 10/2017; Kết thúc: 9/2020; Gia hạn: 6/2021.

7. Danh sách thành viên chính thực hiện nhiệm vụ nêu trên gồm:

Số TT	Họ và tên	Chức danh khoa học, học vị	Cơ quan công tác
1	Mạc Đăng Khoa	Tiến sỹ	Viện MICA ĐHBKHN (2017 - 2019, Viện NC Dữ liệu lớn BigData (2019-đến nay)
2	Đào Trung Kiên	PGS. Tiến sỹ	Viện MICA ĐHBKHN
3	Nguyễn Việt Sơn	Tiến sỹ	Viện Điện - ĐHBKHN
4	Đỗ Thị Ngọc Diệp	Tiến sỹ	Viện MICA ĐHBKHN
5	Nguyễn Việt Tùng	Tiến sỹ	Viện MICA ĐHBKHN

6	Nguyễn Công Phương	Tiến sỹ	Viện MICA ĐHBKHN
7	Vũ Thị Hải Hà	Tiến sỹ	Viện Ngôn ngữ học - Viện Hàn lâm KHXHVN
8	Đinh Thị Hằng	Thạc sỹ	Viện Ngôn ngữ học - Viện Hàn lâm KHXHVN
9	Bùi Đăng Bình	Thạc sỹ	Viện Ngôn ngữ học - Viện Hàn lâm KHXHVN

## II. Nội dung tự đánh giá về kết quả thực hiện nhiệm vụ:

### 1. Về sản phẩm khoa học:

#### 1.1. Danh mục sản phẩm đã hoàn thành:

Số TT	Tên sản phẩm	Số lượng			Khối lượng			Chất lượng		
		XS	Đ	KĐ	XS	Đ	KĐ	XS	Đ	KĐ
1	Hệ thống dịch máy từ văn bản tiếng Việt sang tiếng nói tiếng Mường có định hướng lĩnh vực		X			X			X	
2	Bộ công cụ phục vụ thu thập và xử lý dữ liệu tiếng nói tiếng dân tộc thiểu số chưa có chữ viết ở Việt Nam		X			X			X	
3	Bộ tài liệu công nghệ dịch từ văn bản tiếng Việt ra tiếng nói ngôn ngữ không có chữ viết		X			X			X	
4	Bộ tài liệu thiết kế phần mềm và hướng dẫn sử dụng bộ công cụ phục vụ thu thập và xử lý dữ liệu tiếng nói tiếng dân tộc thiểu số chưa có chữ viết ở Việt Nam		X			X			X	
5	Cơ sở dữ liệu song ngữ văn bản tiếng Việt - tiếng nói tiếng Mường		X			X			X	
6	Bộ tài liệu quy trình xây dựng và mô tả cơ sở dữ liệu song ngữ văn bản tiếng Việt - tiếng nói tiếng Mường		X			X			X	
7	Bộ tài liệu thiết kế, hướng dẫn sử dụng, đánh giá chất lượng và đánh giá kết quả hệ thống dịch máy từ văn bản tiếng Việt sang tiếng nói tiếng Mường có định hướng lĩnh vực		X			X			X	
8	Phần mềm trên thiết bị cầm tay cho phép tra cứu các từ và câu văn bản tiếng Việt ra tiếng nói tiếng Mường		X			X			X	

9	Bộ tài liệu thiết kế phần mềm, hướng dẫn cài đặt và sử dụng phần mềm trên thiết bị cầm tay		X			X			X	
---	--	--	---	--	--	---	--	--	---	--

1.2. Danh mục sản phẩm khoa học dự kiến ứng dụng, chuyển giao (nếu có):

1.3. Danh mục sản phẩm khoa học đã được ứng dụng (nếu có):

## 2. Về những đóng góp mới của nhiệm vụ:

Tính mới và sáng tạo của đề tài được thể hiện ở việc nghiên cứu ra phương pháp, kỹ thuật mới trong dịch tự động từ văn bản một ngôn ngữ ra tiếng nói một ngôn ngữ không có chữ viết. Ở thời điểm hiện tại, cách giải quyết cho bài toán này chưa được hoàn thiện và đi đến đích. Chưa có hệ thống nào thực sự được đưa vào sử dụng. Ngoài việc nghiên cứu phát triển đề xuất thích nghi, kết hợp giữa công nghệ dịch văn bản và công nghệ xử lý tiếng nói không có chữ viết không thông qua biểu diễn trung gian, hay sử dụng công nghệ học sâu hướng trực tiếp để giải quyết bài toán dịch trực tiếp từ văn bản ra tiếng nói cho bài toán này cũng là một đề xuất sáng tạo của đề tài.

Hơn nữa, việc áp dụng cho các ngôn ngữ thiểu số không có chữ viết ở Việt Nam, cụ thể là tiếng Mường ở hai tỉnh Hòa Bình và Phú Thọ cũng là một điểm mới của đề tài. Điều này yêu cầu phải có những nghiên cứu mới về ngôn ngữ, phù hợp các phương pháp đề xuất về công nghệ. Yêu cầu xây dựng nguồn dữ liệu mới, với các hậu xử lý phù hợp với triển khai phương pháp mới này. Có thể nói, các sản phẩm nghiên cứu và ứng dụng của đề tài là những sản phẩm đầu tiên, chưa từng có trong lĩnh vực

## 3. Về hiệu quả của nhiệm vụ:

### 3.1. Hiệu quả kinh tế

Với các lĩnh vực có thể ứng dụng sản phẩm của đề tài kể trên, các kết quả nghiên cứu của đề tài hoàn toàn có thể đưa vào quá trình sản xuất và kinh doanh của các đơn vị, doanh nghiệp như:

- Các doanh nghiệp đơn vị du lịch, lữ hành: Ứng dụng trong các sản phẩm du lịch văn hóa, tại các địa bàn dân tộc thiểu số. Các sản phẩm công nghệ của đề tài có thể góp phần nâng cao trải nghiệm của du khách, hỗ trợ giao lưu và tạo liên kết giữa du khách và đồng bào dân tộc tại địa phương. Việc ứng dụng công nghệ này sẽ tạo ra các sản phẩm du lịch hấp dẫn, tạo sức cạnh tranh và thu hút khách hàng cho doanh nghiệp
- Kết hợp với các doanh nghiệp viễn thông lớn (như VNPT, Viettel) để đưa các dịch vụ viễn thông đến với đồng bào dân tộc thiểu số như:
  - o Tổng đài cung cấp thông tin tự động bằng tiếng nói tiếng dân tộc: Cho phép đồng bào dân tộc thiểu số có thể nghe trực tiếp các thông tin cần thiết từ tổng đài thông tin tự động, ngay cả với các ngôn ngữ không có chữ viết
  - o Dịch vụ chuyển tin nhắn văn bản tiếng việc thành tin nhắn thoại tiếng dân tộc: Sử dụng trong liên lạc giữa người Việt và người dùng dân tộc thiểu số, hoặc cũng có thể sử dụng để tuyên truyền các thông tin của

Đảng và Nhà nước đến đồng bào dân tộc thiểu số thông qua hệ thống tin nhắn thoại.

### 3.2. Hiệu quả xã hội

Như đã trình bày ở trên, các kết quả nghiên cứu và sản phẩm của đề tài sẽ được ứng dụng với các phạm vi khác nhau.

- Các sản phẩm Hệ thống dịch Việt Mường, Phần mềm trên thiết bị cầm tay cho phép tra cứu các từ và câu văn bản tiếng Việt ra tiếng nói tiếng Mường sẽ được ứng dụng trước hết trong phạm vi 2 tỉnh Hòa Bình và Phú Thọ:
  - o Tỉnh Hòa Bình là đơn vị đã kết hợp với Viện MICA trong đề xuất đặt hàng nhiệm vụ khoa học công nghệ cấp quốc gia này (Xem các văn bản về đề xuất đặt hàng nhiệm của UBND Tỉnh Hòa Bình kèm theo hồ sơ). Vì thế, đây là đơn vị đầu tiên sẽ ứng dụng các kết quả nghiên cứu của đề tài này. Ngoài ra Sở Văn hóa, Thể thao và Du lịch của tỉnh Hòa Bình cũng tham gia vào đề tài với vai trò một trong những đơn vị phối hợp nghiên cứu của đề tài. Đây cũng sẽ là đơn vị sẽ tiếp quản kết quả của đề tài và triển khai trọng địa bàn tỉnh. Điều này được thể hiện bằng giấy xác nhận phối hợp của Sở trong hồ sơ đề tài
  - o Tỉnh Phú Thọ cũng là một trong ba tỉnh có số lượng người dân tộc Mường đông nhất cả nước. Các sản phẩm nghiên cứu của đề tài sẽ được tiếp nhận và triển khai trên địa bàn tỉnh Phú Thọ dưới sự hỗ trợ của Sở Văn hóa, Thể thao và Du lịch của tỉnh Phú Thọ. Điều này được khẳng định trong giấy xác nhận của Sở trong hồ sơ đề tài. Sở Văn hóa Thể thao và Du lịch tỉnh Phú Thọ cũng đang có chủ trương xây dựng mô hình các làng du lịch văn hóa Mường. Đây chính là cơ hội để các sản phẩm của đề tài có thể ứng dụng thực tiễn và phát huy hiệu quả
- Các sản phẩm về công nghệ dịch từ văn bản tiếng Việt ra tiếng nói ngôn ngữ không có chữ viết, do được thiết kế không phụ thuộc vào ngôn ngữ nên có thể thích nghi phương pháp và áp dụng trên phạm vi cả nước, tại địa bàn các tỉnh với các dân tộc thiểu số khác nhau. Việc phát triển sản phẩm có thể thực hiện theo các đơn đặt hàng cụ thể, hoặc phối hợp với và đơn vị khác để chuyển giao công nghệ.
- Sản phẩm Cơ sở dữ liệu song ngữ văn bản tiếng Việt - tiếng nói tiếng Mường sẽ được phối hợp với đơn vị hợp tác quốc tế là phòng thí nghiệm LACITO để đưa lên kho ngữ liệu chung OLAC, nhằm chia sẻ với các đơn vị quan tâm trên toàn thế giới.

### III. Tự đánh giá, xếp loại kết quả thực hiện nhiệm vụ

1. Về tiến độ thực hiện: (đánh dấu ✓ vào ô tương ứng):

- Nộp hồ sơ đúng hạn
- Nộp chậm từ trên 30 ngày đến 06 tháng
- Nộp hồ sơ chậm trên 06 tháng

2. Về kết quả thực hiện nhiệm vụ:

- Xuất sắc
- Đạt
- Không đạt

Cam đoan nội dung của Báo cáo là trung thực; Chủ nhiệm và các thành viên tham gia thực hiện nhiệm vụ không sử dụng kết quả nghiên cứu của người khác trái với quy định của pháp luật.

**CHỦ NHIỆM NHIỆM VỤ**  
(Học hàm, học vị, Họ, tên và chữ ký)



TS. Mạc Đăng Khoa

**THỦ TRƯỞNG**  
**TỔ CHỨC CHỦ TRÌ NHIỆM VỤ**  
(Họ, tên, chữ ký và đóng dấu)



**GIÁM ĐỐC**  
VIỆN NGHIÊN CỨU QUỐC TẾ VỀ THÔNG TIN ĐA PHƯƠNG TIỆN  
TRUYỀN THÔNG VÀ ỨNG DỤNG

TS. Nguyễn Việt Sơn