

XÂY DỰNG KHO NGỮ VỰNG SONG NGỮ VIỆT-Ê ĐÊ DỰA TRÊN MÔ HÌNH TƯƠNG TÁC VIỆT-Ê ĐÊ

BUILDING THE VIETNAMESE-EDE BILINGUAL VOCABULARY DATABASE BASED ON THE VIETNAMESE-EDE INTERACTIVE MODEL

Hoàng Thị Mỹ Lệ¹, Phan Huy Khánh²

¹*Trường Cao đẳng Công nghệ, Đại học Đà Nẵng; hoangthimyle.html@gmail.com*

²*Trường Đại học Bách khoa, Đại học Đà Nẵng; khanhph29@gmail.com*

Tóm tắt - Chính sách song ngữ cho đồng bào dân tộc thiểu số (DTTS) là thật sự cần thiết, nhằm tạo điều kiện để người DTTS được học tiếng nói và chữ viết của dân tộc mình; đồng thời góp phần giữ gìn và phát huy bản sắc văn hoá dân tộc; giúp học sinh người DTTS dễ dàng tiếp thu kiến thức khi học tập trong nhà trường và cơ sở giáo dục khác. Từ nhu cầu về chính sách song ngữ cho đồng bào DTTS, bài báo đã đề xuất giải pháp xây dựng kho ngữ vựng (KNV) song ngữ Việt-Ê Đê dựa trên mô hình tương tác Việt-Ê Đê, để góp phần khắc phục những mặt hạn chế của các KNV song ngữ Việt-Ê Đê hiện có và tạo ra một môi trường tương tác giữa người dùng với KNV. Thông qua môi trường tương tác, triển khai ứng dụng tra cứu từ vựng Việt-Ê Đê, nhằm góp phần nâng cao chất lượng dạy và học tiếng Ê Đê.

Từ khóa - kho ngữ vựng song ngữ; dân tộc thiểu số; môi trường tương tác; Việt-Ê Đê; tiếng Ê Đê

1. Đặt vấn đề

Hiệu quả các hoạt động chính trị, kinh tế, văn hoá của cán bộ Đảng và Nhà nước trong các DTTS được nâng lên khi người nói thông tin trực tiếp cho người nghe nhận biết những điều cần làm. Mỗi công dân DTTS thành thạo hai ngôn ngữ, viết và đọc thành thạo chữ Quốc ngữ và chữ viết dân tộc là một thuận lợi lớn trong quá trình nâng cao trình độ văn hoá, khoa học và mở rộng mối quan hệ giữa các dân tộc.

Song song với chính sách song ngữ cho đồng bào DTTS, trong lĩnh vực công nghệ thông tin, việc xây dựng các KNV song ngữ Việt-DTTS nói chung và KNV Việt-Ê Đê nói riêng luôn là thách thức được đặt ra. Trong những năm gần đây, việc xây dựng các KNV song ngữ Việt-Ê Đê bước đầu đã góp phần giải quyết vấn đề tra cứu từ vựng Việt-Ê Đê trong dịch thuật thủ công các bản tin tiếng DTTS của Đài Tiếng nói Việt Nam khu vực Tây Nguyên. Tuy nhiên, kết quả nghiên cứu về KNV song ngữ Việt-Ê Đê vẫn còn những hạn chế cần được quan tâm nghiên cứu, chẳng hạn như: chưa thống nhất dùng phông chữ Unicode, chưa chia sẻ KNV cho các hoạt động nghiên cứu liên quan đến xử lý tiếng Ê Đê.

Xuất phát từ thực trạng trên, bài báo đề xuất giải pháp phát triển KNV song ngữ Việt-Ê Đê dựa trên mô hình tương tác Việt-Ê Đê, nhằm góp phần khắc phục những hạn chế của KNV song ngữ Việt-Ê Đê hiện có và xây dựng hạ tầng cơ sở để tạo môi trường tương tác giữa người dùng với KNV. Thông qua môi trường tương tác, triển khai ứng dụng tra cứu từ vựng Việt-Ê Đê, để góp phần nâng cao chất lượng dạy và học tiếng Ê Đê.

Cấu trúc của bài báo gồm các nội dung: Trước hết, chúng tôi trình bày tổ chức kiến trúc KNV song ngữ Việt-

Abstract - The bilingualism policy for Vietnam ethnic minorities is really necessary because it has created conditions for ethnic minorities to learn their spoken and written language to preserve and promote the national cultural character, helping ethnic minority pupils to acquire knowledge easily in schools and other educational establishments. From the needs for bilingual policy for ethnic minorities, the paper has proposed the measure of building the Vietnamese-Ede bilingual vocabulary database based on the Vietnamese-Ede interactive model. This solution has contributed to overcoming the restrictions of the current Vietnamese-Ede bilingual vocabulary database and building the interactive environment between users and the vocabulary database. Through the interactive environment, the application to look up Viet-Ede bilingual vocabulary is developed in order to contribute to improving the quality of teaching and learning the Ede language.

Key words - bilingual vocabulary database; ethnic minority; interactive model; Vietnamese-Ede; the Ede language

Ê Đê. Thứ hai, đề xuất mô hình tương tác Việt-Ê Đê trong xây dựng KNV song ngữ Việt-Ê Đê và kết quả thực nghiệm. Thứ ba, đề xuất đánh giá KNV qua độ bao phủ của KNV trên các văn bản. Thứ tư, đánh giá kết quả đạt được, và cuối cùng là Phần kết luận.

2. Tổ chức kiến trúc kho ngữ vựng song ngữ

Kho ngữ vựng là một cơ sở dữ liệu (CSDL) về từ vựng. Trong đó, từ vựng được hiểu là tập hợp các từ, cụm từ thường không đổi, còn được gọi là thành ngữ trong một ngôn ngữ. Các KNV thường được đánh dấu hay gắn nhãn từ loại. Có nhiều quan điểm định nghĩa hay nhìn nhận một KNV, liên quan đến cách tổ chức và mục đích sử dụng [6]. KNV song ngữ trong giải pháp đề xuất là KNV gồm các từ tiếng Việt với từ tiếng Ê Đê.

2.1. Tiêu chí dữ liệu

Với mục tiêu xây dựng KNV song ngữ Việt-Ê Đê làm hạ tầng cơ sở cho xử lý tiếng Ê Đê, các tiêu chí dữ liệu được đặt ra trong KNV song ngữ như sau:

Các từ tiếng Ê Đê chủ yếu được thu thập và ghi theo tiếng Ê Đê thuộc nhóm Kpă. Đây là tiếng địa phương vốn được xem là dễ nghe và dễ hiểu nhất. Các mục từ tiếng Ê Đê phản ánh phần nào vốn văn hóa truyền thống của người Ê Đê và được ghi bằng chữ Ê Đê.

Các ví dụ được đưa vào để làm sáng tỏ cho nghĩa và cách sử dụng của từ ngữ, còn gọi là ngữ cảnh của mục từ.

Các mục từ được gắn nhãn từ loại: gắn nhãn N cho danh từ, gắn nhãn V cho động từ, gắn nhãn A cho tính từ, gắn nhãn O cho các mục từ không phải là danh từ, động từ hay tính từ.

Các từ đa nghĩa được ghi nhận, dịch và đối chiếu với các từ khác nhau tương đương trong ngôn ngữ đích.

Khi giống hệt từ của ngôn ngữ nguồn, tìm từ tương đương trong ngôn ngữ đích, trên cơ sở nghĩa cơ bản, nghĩa thường dùng hiện nay ở cả hai ngôn ngữ.

Dữ liệu được lưu trữ trên máy với phong Unicode chuẩn. Đây là tiêu chí mà các kết quả nghiên cứu KNV song ngữ Việt-Ê Đê từ trước đến nay chưa đề cập đến.

2.2. Nguồn dữ liệu

- Từ điển Việt-Ê Đê gồm 1.000 đơn vị mục từ, phần lớn thuộc vốn từ cơ bản, thông dụng của tiếng Việt [11], [12].
- Từ điển Ê Đê-Việt có khoảng 10.000 mục từ. Phần lớn là các từ ngữ thường dùng hàng ngày của tiếng Ê Đê [13].
- Kho ngữ vựng tiếng Việt với trên 31.000 mục từ, được kế thừa của “Đề tài VLSP” [5].

Nguồn dữ liệu từ điển giấy được nhập theo cách thủ công, theo mẫu chung và thống nhất dùng phông chữ Unicode chuẩn.

Thao tác hiển thị chữ Ê Đê, được thực hiện theo hai cách: dùng bộ gõ WinVNKey qua môi trường IEWVNC [4], hay cũng có thể gõ theo cách gõ qui định và sẽ được chuyển đổi khi đưa vào môi trường tương tác.

2.3. Cấu trúc kho ngữ vựng

Kho ngữ vựng được thiết kế theo mô hình CSDL quan hệ. CSDL quan hệ được sử dụng như là tập hợp các bảng lưu trữ dữ liệu. Các bảng CSDL tương tự như một KNV, được lưu trữ hoàn toàn độc lập về cấu trúc cũng như về dữ liệu. Mô hình CSDL quan hệ có những ưu và nhược điểm sau:

Bảng 1, được thiết kế để lưu các mục từ tiếng Ê Đê. Thuộc tính NOTE, đánh dấu các mục từ được vay mượn từ ngôn ngữ khác.

Bảng 1. Bảng lưu trữ mục từ Ê Đê

Tên thuộc tính	Kiểu dữ liệu	Mô tả
IDE	AutoNumber	Chỉ số Ê Đê
<u>ED</u>	Text	Mục từ Ê Đê
NOTE	Boolean	Ghi chú

Bảng 2, được thiết kế để lưu các mục từ tiếng Việt, có khoảng 31.000 mục từ đã kế thừa được từ KNV trong “Đề tài VLSP”. Thuộc tính ADD để đánh dấu mục từ tiếng Việt tương ứng với tiếng Ê Đê được bổ sung vào KNV.

Bảng 2. Bảng lưu trữ mục từ Việt

Tên thuộc tính	Kiểu dữ liệu	Mô tả
IDV	AutoNumber	Chỉ số Việt
<u>VI</u>	Text	Mục từ Việt
ADD	Boolean	Bổ sung

Bảng 3, được thiết kế để lưu trữ kết quả chỉ số Việt và chỉ số Ê Đê khi thực hiện tương tác theo mô hình tương tác Việt-Ê Đê với các dữ liệu về chỉ số từ Việt, chỉ số từ Ê Đê, từ loại, các ví dụ Việt-Ê Đê theo mục từ Việt.

Bảng 3. Kho ngữ vựng song ngữ Việt-Ê Đê

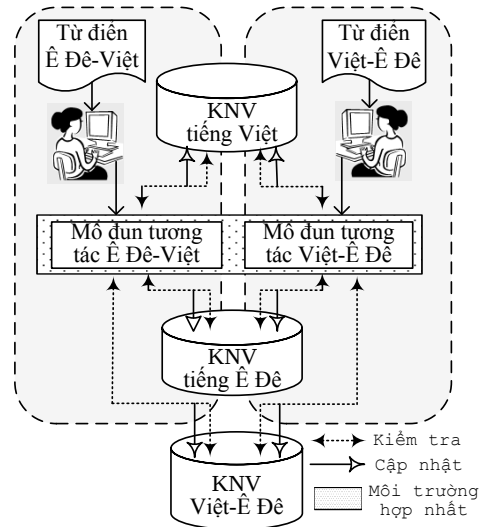
Tên thuộc tính	Kiểu dữ liệu	Mô tả
IDV	Number	Chỉ số Việt
IDE	Number	Mục từ Việt
SP	Text	Từ loại
EXPLE	Text	Ví dụ Việt-Ê Đê

3. Mô hình tương tác Việt-Ê Đê

Mô hình tương tác là môi trường để người sử dụng tương tác chuyển đổi trực tuyến dữ liệu từ từ điển giấy sang từ điển máy tính.

3.1. Hoạt động mô hình tương tác Việt-Ê Đê

Dựa trên các nguồn dữ liệu tiếng Ê Đê có trên thực tế với mục tiêu xây dựng hạ tầng cơ sở cho xử lý tiếng Ê Đê, mô hình tương tác Việt-Ê Đê trong việc phát triển KNV song ngữ Việt-Ê Đê được đề xuất, thể hiện trong Hình 1.



Hình 1. Mô hình tương tác Việt-Ê Đê

Hoạt động trong mô hình:

- Nguồn dữ liệu từ điển Việt-Ê Đê và Ê Đê-Việt được định dạng dữ liệu trước khi chuyển vào môi trường tương tác.
- Môi trường tương tác thực hiện đọc dữ liệu từ nguồn dữ liệu đưa vào và thực hiện tương tác vào hai KNV tiếng Việt và tiếng Ê Đê, kiểm tra có các mục từ trong các KNV hay không để thực hiện cập nhật và đọc chỉ số tương ứng. Với chỉ số Việt và chỉ số Ê Đê, môi trường tiếp tục tương tác vào KNV song ngữ Việt-Ê Đê để thực hiện cập nhật vào KNV song ngữ Việt-Ê Đê.
- Môi trường tương tác trong mô hình được xây dựng qua hai mô-đun tương tác:
 - + Mô-đun tương tác Ê Đê-Việt có nguồn dữ liệu đầu vào là tập tin văn bản từ điển Ê Đê-Việt.
 - + Mô-đun tương tác Việt-Ê Đê có nguồn dữ liệu đầu vào là tập tin văn bản từ điển Việt-Ê Đê.

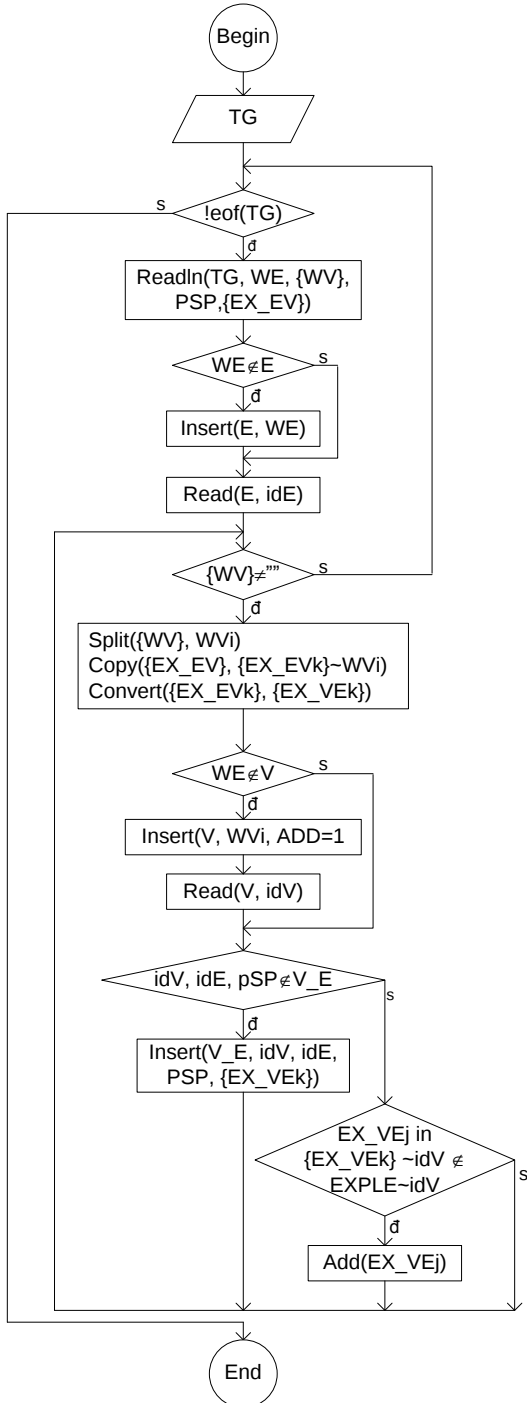
3.2. Xây dựng thuật toán mô-đun tương tác

3.2.1. Mô-đun tương tác Ê Đê-Việt

Đầu vào: TG: file từ điển Ê Đê-Việt đã định dạng,
V: KNV tiếng Việt, E: KNV Ê Đê, V_E: KNV Việt-Ê Đê

Đầu ra: V_E: KNV Việt-Ê Đê, E: KNV Ê Đê,
V: KNV tiếng Việt

Thuật toán mô-đun tương tác Ê Đê-Việt được biểu diễn trong Hình 2.



Hình 2. Thuật toán mô-đun tương tác Ê Đê-Việt

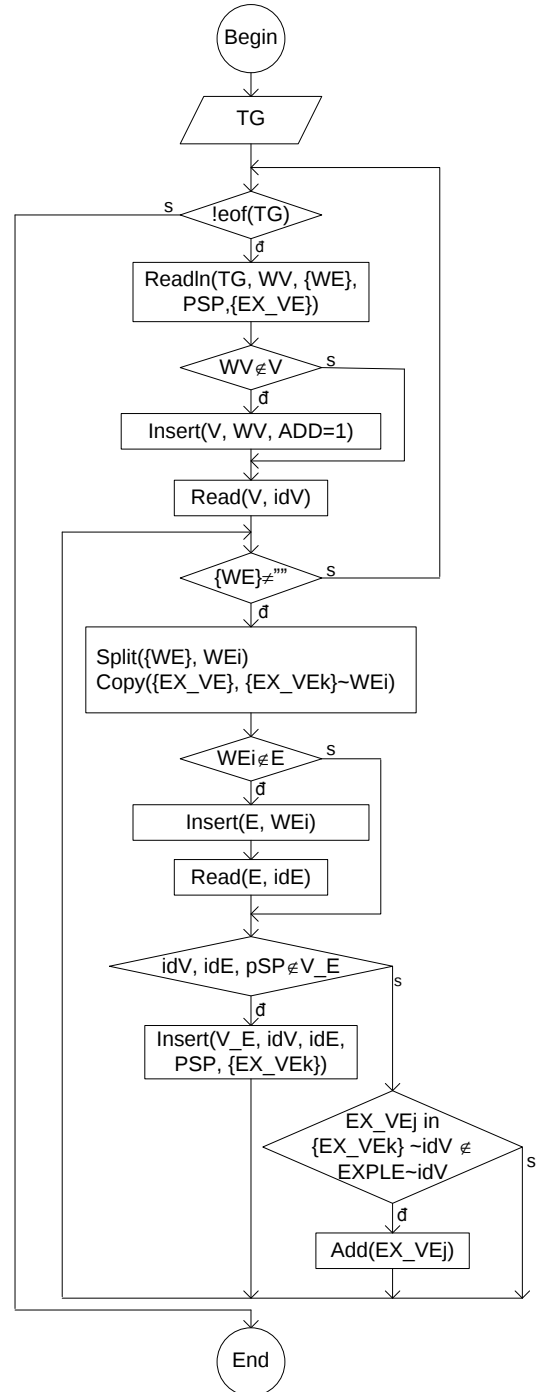
3.2.2. Mô-đun tương tác Việt-Ê Đê

Đầu vào: TG: file từ điển Ê Đê-Việt đã định dạng,
V: KNV tiếng Việt, E: KNV Ê Đê, V_E: KNV Ê Đê

Đầu ra: V_E: KNV Việt-Ê Đê, E: KNV Ê Đê,

V: KNV tiếng Việt

Thuật toán mô-đun tương tác Ê Đê-Việt được biểu diễn trong Hình 3.



Hình 3. Thuật toán mô-đun tương tác Việt-Ê Đê

3.3. Xây dựng môi trường tương tác

Mô hình thiết kế MVC (Model - View - Controller) được chọn để triển khai xây dựng và quản lý KNV song ngữ.

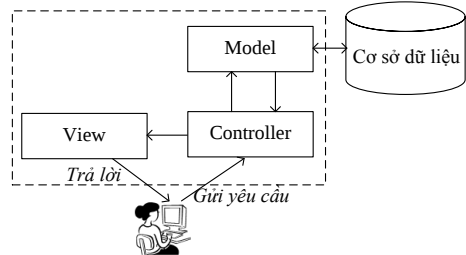
Mô hình MVC là mô hình kiến trúc phần mềm được dùng trong kỹ thuật phần mềm. Mô hình MVC chia ứng dụng thành ba phần thiết kế, với mục tiêu tách phần giao diện và phần code để dễ quản lý, phát triển và bảo trì. Các phần thiết kế của mô hình MVC gồm:

Model có nhiệm vụ thao tác với CSDL, chứa tất cả các hàm, các phương thức truy vấn dữ liệu như *select*, *insert*, *update*, *delete* trong CSDL. Controller sẽ thông qua các hàm, phương thức đó để lấy dữ liệu rồi gửi qua View.

View chịu trách nhiệm hiển thị các thông tin cho người dùng thông qua giao diện. Dữ liệu hiển thị được nhận từ Controller.

Controller nhận yêu cầu từ người dùng và lấy dữ liệu tương ứng từ Model và gửi dữ liệu qua View để xử lý và trả kết quả cho người dùng.

Hoạt động mô hình MVC được minh họa trong Hình 4.



Hình 4. Mô hình thiết kế MVC

Môi trường tương tác được triển khai với hai mô-đun tương tác để thực hiện import dữ liệu từ điển Việt-Ê Đê và từ điển Ê Đê-Việt cập nhật vào các KNV.

Thông qua môi trường tương tác, kết quả cập nhật KNV song ngữ Việt-Ê Đê được thống kê trong Bảng 4.

Bảng 4. Bảng thống kê số mục từ trong các kho ngữ vựng

Kho ngữ vựng	Mục từ có được		Tổng cộng
	Tương tác Ê Đê-Việt	Tương tác Việt-Ê Đê	
Việt	11.357	2.575	13.932
Ê Đê	9.287	2.149	11.436
Việt-Ê Đê	17.980	3.167	21.147

3.4. Đề xuất đánh giá kho ngữ vựng qua độ bao phủ

Để đánh giá các đơn vị mục từ Ê Đê trong KNV được xây dựng có mặt đầy đủ trong các văn bản Ê Đê hiện có, chúng tôi đề xuất đánh giá qua độ bao phủ của các đơn vị mục từ Ê Đê trong KNV. Độ bao phủ được đặt tên DC (Degree of Coverage) và được tính tổng các từ trong văn bản không có trong KNV trên tổng các từ trong văn bản như là một tỷ lệ phần trăm tổng số các từ bao phủ được.

Công thức tính độ bao phủ của KNV:

$$DC = 1 - \frac{X}{Y}$$

Trong đó:

X: tổng số từ trong văn bản chưa có trong KNV

Y: tổng số từ trong văn bản

$$X \leq Y$$

- **Độ bao phủ trong KNV tiếng Ê Đê**, để xác định các từ tiếng Ê Đê trong văn bản chưa có trong KNV, công cụ tách được xây dựng từ để xác định tổng số từ Ê Đê trong văn bản không có trong KNV. Hướng tiếp cận cho bài toán tách từ áp dụng trong bộ công cụ là hướng tiếp cận dựa trên

các mục từ trong KNV Việt-Ê Đê. Phương pháp tách từ được áp dụng trong bộ công cụ là phương pháp so khớp cục đại, để xác định mục từ có trong KNV và các mục từ không có trong KNV.

Để đảm bảo tính chính xác cho nghiên cứu và thử nghiệm, tài liệu được thử nghiệm để tính độ bao phủ của KNV là những tài liệu được nhập thủ công từ sách học tiếng Ê Đê [14], [15], [16], truyện đọc song ngữ Ê Đê-Việt [8], [9], [10] và các bản tin Ê Đê thu thập từ trang web Đài Tiếng nói Việt Nam VOV4 [1]. Các bản tin Ê Đê hiển thị với phông chữ TayNguyenKey, chúng tôi cũng đã xây dựng công cụ chuyển đổi sang phông chữ Unicode để đưa vào làm dữ liệu thử nghiệm.

Các từ tiếng Ê Đê được phát hiện không có trong KNV Ê Đê cũng đã được kiểm tra và kết quả cho thấy các từ này là các trường hợp: các danh từ riêng, các từ do nhập sai và các từ chưa có trong KNV. Bảng 5 thống kê các từ chưa có trong KNV Ê Đê.

Bảng 5. Thống kê số từ không có trong KNV tiếng Ê Đê

Loại văn bản	Số từ tách được	Số từ chưa có trong KNV Ê Đê
Sách học tiếng Ê Đê	1.815	96
Truyện đọc Ê Đê-Việt	2.052	276
Bản tin tiếng Ê Đê	3.735	328

Theo số liệu thống kê có được trong Bảng 5, độ bao phủ KNV trên các văn bản sau khi thực hiện tách từ, kết quả tính được thể hiện trong Bảng 6.

Bảng 6. Độ bao phủ KNV trên các văn bản tiếng Ê Đê

Loại văn bản	Số từ kiểm tra	Số từ chưa có	DC
Sách học tiếng Ê Đê	1.815	96	94,71%
Truyện đọc Ê Đê-Việt	2.052	276	86,55%
Bản tin tiếng Ê Đê	3.725	328	91,19%
Độ bao phủ trung bình			90,82%

Qua kết quả kiểm chứng tính độ bao phủ trên các tài liệu, cho thấy độ bao phủ của KNV Ê Đê trên các tài liệu sách học tiếng Ê Đê chiếm tỷ lệ cao hơn so với các tài liệu khác. Điều này cũng dễ nhận thấy, vì rằng, các ngôn từ trong các sách học tiếng Ê Đê cụ thể, ngắn gọn và các từ vựng chủ yếu là tiếng địa phương Kpă.

- **Độ bao phủ trong KNV tiếng Việt**: để xác định các từ tiếng Việt trong văn bản chưa có trong KNV, phương pháp tách từ tiếng Việt được kế thừa từ bộ công cụ tách từ vnTokenizer [5] sau khi đã được bổ sung vào KNV mở rộng của vnTokenizer các từ tiếng Việt đã phát hiện chưa có trong KVN tiếng Việt. Sau đó, kiểm tra các từ tách được trong KNV song ngữ Việt-Ê Đê.

Các từ được kiểm tra không có trong KNV Việt-Ê Đê đã được kiểm chứng và kết quả nhận thấy các từ này là các trường hợp: danh từ riêng, các từ do nhập sai và từ chưa có trong KNV. Kết quả các từ phát hiện chưa có trong KNV Việt-Ê Đê được thống kê trong Bảng 7.

Bảng 7. Thống kê số không có trong KNV tiếng Việt

Loại văn bản	Số từ tách được	Số từ chưa có trong KNV Ê Đê
Sách học tiếng Việt lớp 3, lớp 4, lớp 5	3.324	225
Truyện đọc tiếng Việt lớp 3, lớp 4, lớp 5	3.538	274
Bản tin Sở Giáo dục và Đào tạo Đắk Lắk	4.777	509

Theo số liệu thống kê có được trong Bảng 7, độ bao phủ KNV trên các văn bản tiếng Việt tính được trong Bảng 8.

Bảng 8. Độ bao phủ KNV trên các văn bản tiếng Việt

Loại văn bản	Số từ kiểm tra	Số từ chưa có	ĐC
Sách học tiếng Việt lớp 3, lớp 4, lớp 5	3.324	225	93,23%
Truyện đọc tiếng Việt lớp 3, lớp 4, lớp 5	3.538	274	92,26%
Bản tin Sở Giáo dục và Đào tạo Đắk Lắk	4.777	509	89,34%
Độ bao phủ trung bình			91,61%

Kết quả tính độ bao phủ của KNV Việt-Ê Đê trên các văn bản tiếng Việt là 91,61% và trên các văn bản tiếng Ê Đê là 90,82%. Theo độ bao phủ đã tính được trong Bảng 6 và Bảng 8, đảm bảo được tính khả thi của KNV cho việc triển khai ứng dụng liên quan đến xử lý tiếng Ê Đê.

4. Đánh giá kết quả đạt được

Để đánh giá KNV song ngữ Việt-Ê Đê được xây dựng dựa trên mô hình tương tác đã đề xuất, chúng tôi lập ra bảng đánh giá, so sánh với các kết quả nghiên cứu liên quan đến KNV song ngữ Việt-Ê Đê đã được nghiên cứu từ trước đến nay. Bảng 9 trình bày những so sánh và đánh giá các kết quả nghiên cứu về KNV song ngữ Việt-Ê Đê.

5. Kết luận

Với bối cảnh xử lý ngôn ngữ DTTS nói chung và ngôn ngữ Ê Đê nói riêng, bài báo đề xuất giải pháp xây dựng KNV Việt-Ê Đê dựa trên mô hình tương tác Việt-Ê Đê, đạt được những kết quả sau:

Thống nhất sử dụng Unicode trong KNV song ngữ.

Góp phần phát triển hạ tầng cơ sở cho bài toán xử lý tiếng Ê Đê nói riêng và tiếng DTTS Việt Nam nói chung.

Chia sẻ KNV cho các hoạt động nghiên cứu liên quan đến xử lý tiếng Ê Đê.

Mô hình tương tác Việt-Ê Đê có thể mở rộng cho việc phát triển các KNV song ngữ Việt-DTTS.

Giải pháp được đề xuất mang tính thực tiễn, vì đã góp phần khắc phục được những hạn chế về KNV song ngữ Việt-Ê Đê mà các nghiên cứu trước đây chưa thực hiện được.

Bảng 9. Đánh giá kết quả nghiên cứu KNV song ngữ Việt-Ê Đê

Kết quả nghiên cứu	Sử dụng Unicode	Ứng dụng thực tiễn	Chia sẻ
Từ điển điện tử Việt-Ê Đê [3]	Chưa	Phục vụ Đài Tiếng nói Việt Nam khu vực Tây Nguyên	Không
KNV Anh-Việt-Ê Đê [7]	Chưa	Thử nghiệm với dung lượng ngữ liệu nhỏ	Không
KNV Việt-Ê Đê cho hệ dịch Việt-Ê Đê [2]	Chưa	Thử nghiệm với dung lượng ngữ liệu nhỏ	Không
KNV song ngữ Việt-Ê Đê bài báo đã xây dựng	Đã sử dụng	Làm hạ tầng cơ sở để phát triển các ứng dụng xử lý tiếng Ê Đê	Có chia sẻ

TÀI LIỆU THAM KHẢO

[1] Đài Tiếng nói Việt Nam. Địa chỉ: <http://vov4.vov.vn/Ede.aspx>, [Truy cập: 20/04/2016].

[2] Đinh Thị Thu Nhân (2013), “Xây dựng kho ngữ vựng song ngữ Việt-Ê Đê trong xử lý tiếng Ê Đê”, *Luận văn Thạc sĩ ngành Khoa học Máy tính*, ĐH Đà Nẵng.

[3] Đoàn Công Quế, Báo cáo đề tài khoa học “Xây dựng từ điển điện tử Việt –Ê Đê”, <http://www.baomoi.com/bao-cao-de-tai-khoa-hoc-xay-dung-tu-dien-dien-tu-viet-ede/c/1137624.epi>, [Truy cập: 17/02/2016].

[4] Hoàng Thị Mỹ Lệ, Vilavong Souksan, Phan Huy Khánh, (2013), “Using Unicode in Encoding the Vietnamese Ethnic Minority Languages, Applying for the EDe Language”, *Proceeding of the International Conference on Knowledge and System Engineering*, Springer, KSE 2013, HaNoi, No: Vol 244, pp. 137-148.

[5] Hồ Tú Bảo, “Đề tài VLSP-Nhánh đề tài Xử lý văn bản” [Trực tuyến], <http://vlsp.hpda.vn:8080/demo/>, [Truy cập: 17/08/2016].

[6] Lưu Tuấn Anh và Yamamoto Kazuhide, “Ứng dụng phương pháp Pointwise vào bài toán tách từ cho tiếng Việt”, <http://viet.jnlp.org/dongdu>, [Truy cập: 17/02/2016].

[7] Nguyễn Thị Tuyết (2003), “Xây dựng cơ sở dữ liệu từ vựng đa ngữ Việt-Anh-Ê Đê”, *Luận văn Thạc sĩ ngành Khoa học Máy tính*, ĐH Đà Nẵng.

[8] Phan Hồng, Nguyễn Văn Thù, *Truyện đọc song ngữ Ê Đê-Việt Tập 1*, NXBGD.

[9] Phan Hồng, Nguyễn Văn Thù, *Truyện đọc song ngữ Ê Đê-Việt Tập 2*, NXBGD.

[10] Phan Hồng, Nguyễn Văn Thù, *Truyện đọc song ngữ Ê Đê-Việt Tập 3*, NXBGD.

[11] Sở Giáo dục - Đào tạo Đắk Lắk (1993), *Từ điển Việt-Ê Đê tập 1*, NXBGD.

[12] Sở Giáo dục - Đào tạo Đắk Lắk (1993), *Từ điển Việt-Ê Đê tập 2*, NXBGD.

[13] Tạ Văn Thông, *Đề tài Từ điển Ê Đê-Việt*, 2014.

[14] Y-Hạ Niê Kdăm (2013), *Klei ÊĐê Hduôn 1*, NXBGD.

[15] Y-Hạ Niê Kdăm (2013), *Klei ÊĐê Hduôn 2*, NXBGD.

[16] Y-Hạ Niê Kdăm (2013), *Klei ÊĐê Hduôn 3*, NXBGD.