

CHƯƠNG 1: GIỚI THIỆU LUẬN VĂN

1.1 GIỚI THIỆU ĐỀ TÀI

Trong cuộc sống hiện nay có rất nhiều thiết bị công nghệ được áp dụng rộng rãi và phục vụ cho nhiều mục đích khác nhau trong cuộc sống và hiện nay đang càng được quan tâm để mở rộng. Một trong những vấn đề đã và đang được phát triển đó chính là dịch máy. Trên thế giới có rất nhiều ngôn ngữ nói, viết khác nhau trên thế giới và sự khác biệt về ngôn ngữ là một trở ngại lớn trong hầu hết các mặt của đời sống. Sự khác biệt đó làm cho con người khó tiếp thu được các kiến thức đến từ các ngôn ngữ khác và nó cũng làm cho sự toàn cầu hoá bị tác động chậm lại.

Vì thế, con người mong muốn có thêm những công cụ hỗ trợ họ một phần giúp họ có thể tạm thời bỏ qua sự khác biệt ngôn ngữ khác nhau để hoàn thành một mục đích của họ. Từ đó các dịch vụ về ngôn ngữ được ra đời và phát triển một cách nhanh chóng.

Trong những năm gần đây, dịch máy đóng vai trò như một công cụ hỗ trợ con người có thể bỏ qua rào cản ngôn ngữ để cập nhật thông tin từ nhiều nguồn khác nhau một cách dễ dàng. Nó đã giúp con người rất nhiều trong các công việc học tập và làm việc một cách hiệu quả hơn.

Nếu như ngày xưa để hiểu được một câu tiếng Anh bạn có thể cần phải tra trong từ điển dày cộm tốn thời gian và không hiệu quả. Thế nhưng ngày nay với các phần mềm dịch tiếng Anh sang tiếng Việt online, bạn sẽ dễ dàng tra cứu ý nghĩa của các từ, câu, đoạn văn bản mình cần trong thời gian ngắn gần như là ngay lập tức.

Với sự phát triển kinh tế và toàn cầu hoá, các dịch vụ dịch cũng như các ứng dụng sử dụng dịch vụ này không chỉ hấp dẫn người sử dụng mà còn thu hút giới công nghệ trên thế giới. Một dịch vụ dịch máy thường sở hữu những đặc điểm nổi bật như:

- ❖ Sử dụng hoàn toàn miễn phí, kho từ phong phú.
- ❖ Phần mềm hỗ trợ dịch với độ chính xác cao.
- ❖ Sử dụng phần mềm một cách dễ dàng.
- ❖ Đôi khi không cần kết nối mạng, chúng ta vẫn có thể sử dụng các chức năng phiên dịch từ của những ứng dụng này.

Từ đó, việc dịch máy và các ứng dụng sử dụng dịch máy đã đem lại một lợi ích khổng lồ trong cuộc sống của con người.

1.2 LÝ DO LỰA CHỌN ĐỀ TÀI

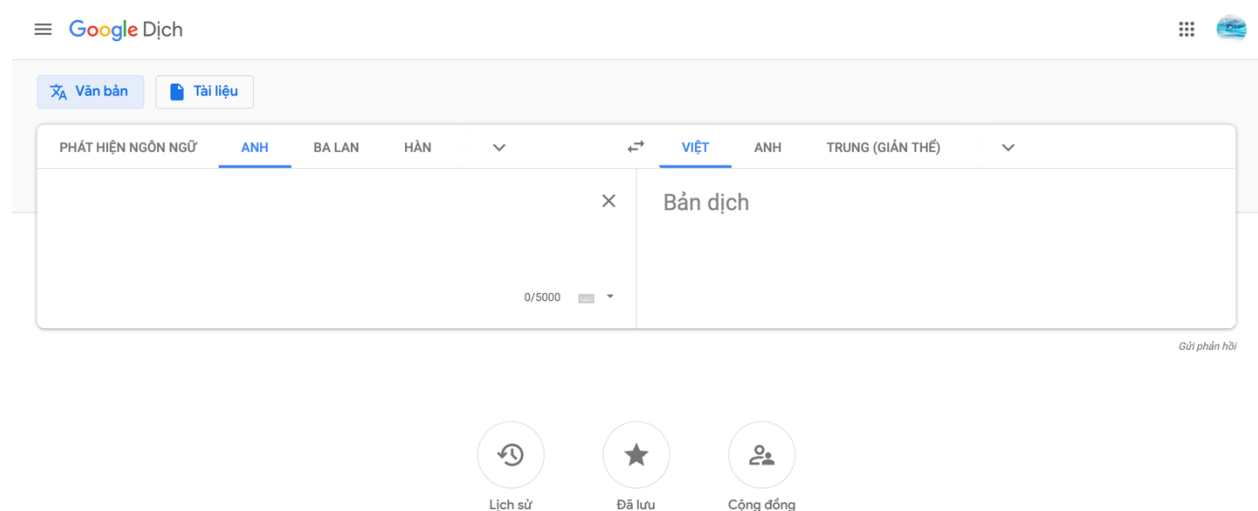
Nhằm mục đích phát triển kiến thức của bản thân, chúng em quyết định lựa chọn đề tài “Xây dựng mô hình dịch máy từ tiếng Anh sang tiếng Việt” để tìm hiểu thêm các kiến thức, có sản phẩm mang ý nghĩa thực tế và có tiềm năng trong tương lai.

Ngoài ra việc lựa chọn đề tài này giúp chúng em có thêm các kiến thức về các thư viện, các mã nguồn mở là những nguồn tài liệu quý giá đối với chúng em. Với các lợi ích kể trên và sau khi thực hiện luận, chúng em sẽ có thêm hiểu biết về các quy trình phát triển dự án thực, không những thế chúng em còn được học hỏi các kiến thức nền tảng và chuyên sâu liên quan đến dịch máy. Chính những điều đó sẽ là nền tảng giúp chúng em phát triển hơn trong con đường học tập cũng như trong công việc của chúng em.

Như đã trình bày ở trên, các ứng dụng sử dụng dịch vụ dịch máy ngày càng thu hút sự đầu tư và các nhà phát triển phần mềm lẫn người sử dụng phần mềm. Việt Nam đang đi trên con đường toàn cầu hoá nên thị trường Việt Nam là một thị trường đầy tiềm năng. Nhu cầu học tiếng Anh của các học sinh, sinh viên cũng như người dân là rất lớn. Trên thị trường hiện nay có khá nhiều các công cụ miễn phí cũng như thu phí nhưng những ứng dụng này cũng có một số ưu điểm cũng như khuyết điểm của nó.

Google Translate

Đây là một sản phẩm rất hữu ích của Google và là phần mềm dịch tiếng Anh sang tiếng Việt đầu tiên được rất nhiều người sử dụng cho các nhu cầu như là học tập.



Ưu điểm:

- Google Translate có giao diện dễ nhìn, dễ sử dụng, đặc biệt có độ chính xác khá cao.

- Có thể sử dụng phần mềm này trên máy tính, điện thoại hay cả máy tính bảng.
- Miễn phí và dịch được nhiều ngôn ngữ khác nhau trên thế giới.

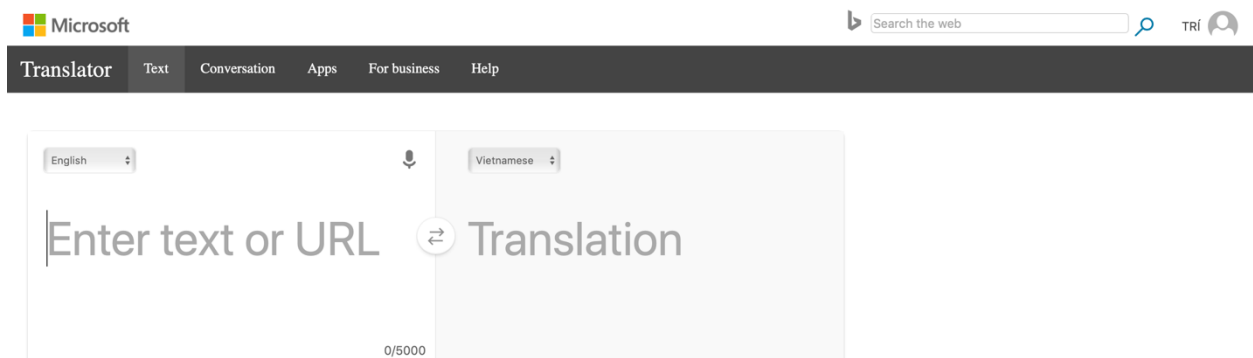
Nhược điểm:

- Để sử dụng phần mềm này cần phải có kết nối mạng cho thiết bị hoặc phải tải bộ dữ liệu để dùng khi không có mạng.
- Chỉ hiển thị giới hạn 5000 từ mỗi lần và cắt đoạn chưa hợp lí.

Link tham khảo: <https://translate.google.com>

Microsoft Translator

Nối tiếp Google, Microsoft cũng như cho ra đời phần mềm dịch trực tuyến của riêng mình.



Ưu điểm:

- Giao diện dễ nhìn, độ chính xác cao.
- Có thể sử dụng trên các thiết bị khác nhau như: máy tính, điện thoại.
- Sử dụng miễn phí và dịch được nhiều ngôn ngữ..

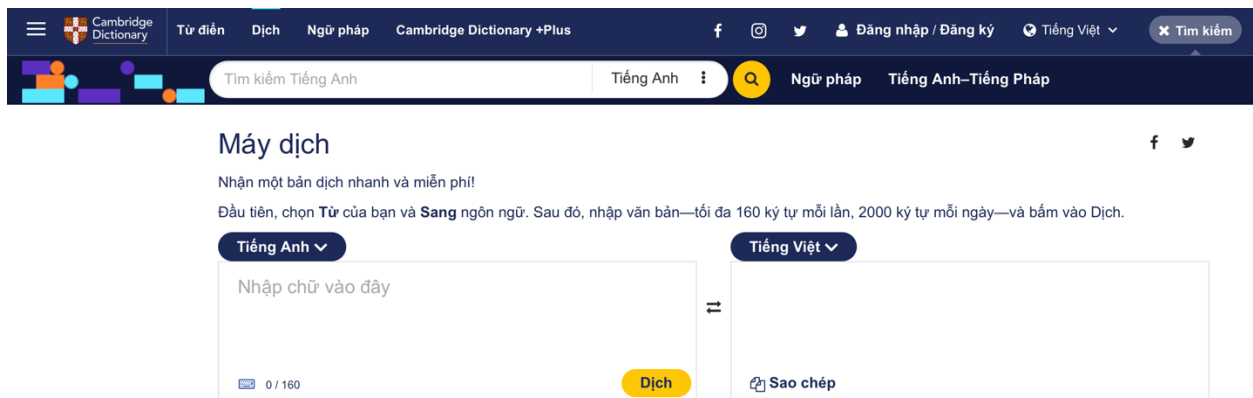
Nhược điểm:

- Cần có kết nối mạng cho thiết bị.
- Chỉ giới hạn 5000 từ và không tự cắt đoạn văn bản để dịch tiếp.
- Giao diện không cân đối.

Link tham khảo: <https://www.bing.com/translator>

Cambridge Dictionary Translate

Phần mềm này được xuất bản bởi Đại học Cambridge.



Ưu điểm:

- Dịch được 23 ngôn ngữ.
- Có thể tự động phát hiện ngôn ngữ.
- Kết quả có độ chính xác cao.

Nhược điểm:

- Miễn phí tối đa 160 từ và 2000 từ trên một ngày.
- Cần phải qua một bước trung gian là ấn nút “Dịch” để thực hiện dịch.
- Cần có kết nối mạng có thiết bị.

Link tham khảo: <https://dictionary.cambridge.org/vi/translate>

1.3 HƯỚNG PHÁT TRIỂN CỦA LUẬN VĂN

Trong lịch sử phát triển dịch máy, có rất nhiều phương pháp đã được nghiên cứu và áp dụng đem lại thành công nhất định. Trong sự phát triển của dịch máy, có cách tiếp cận chủ yếu là dịch chuyển đổi, lịch liên ngữ và dịch dựa trên dữ liệu. Trong đó, dịch máy thống kê, một trong những phương pháp theo cách tiếp cận dựa trên dữ liệu, hiện đang là một hướng phát triển đầy tiềm năng, thu hút được nhiều sự quan tâm của các nhà nghiên cứu.

Ưu điểm vượt trội của phương pháp dịch máy thống kê là thay vì xây dựng các quy luật, từ điển được chuyển đổi bằng tay, nó tự động thiết lập các quy luật, từ điển dựa trên kết quả thống kê có được từ kho ngữ liệu. Chính vì thế nên dịch máy thống kê có tính linh hoạt cao có thể áp dụng được cho bất kì cho một cặp ngôn ngữ ngẫu nhiên. Dịch máy thống kê hiện nay có 3 hướng tiếp cận chính đó là: dịch máy thống kê theo đơn vị từ, dựa trên đơn vị cụm từ và dựa trên cú pháp.

Tuy nhiên các phương pháp này vẫn có những hạn chế do sự thiếu hụt về thông tin ngôn ngữ. Mô hình dịch thống kê vẫn chưa giải quyết được một số vấn đề còn sai sót như trật tự từ, khả năng chọn cụm từ phù hợp.

Luận văn sử dụng phương pháp dịch máy dựa thống kê để tiếp cận hoàn toàn dựa trên ngữ liệu nên nó hoàn toàn độc lập với ngôn ngữ. Những tham số thống kê thu được từ việc huấn luyện trên ngữ liệu song ngữ sẽ được sử dụng cho các lần dịch sau.

Để hoàn thành luận văn, nhóm sinh viên tiến hành xây dựng một mô hình dịch máy từ tiếng Anh sang tiếng Việt và có ứng dụng thử nghiệm để đánh giá mô hình.

1.4 MỤC TIÊU CỦA LUẬN VĂN

Để hoàn thành tốt đề tài luận văn, bản luận văn và sản phẩm cuối cùng của nhóm sinh viên sẽ đảm bảo các mục tiêu sau đây:

- ❖ Trình bày lý thuyết nền tảng và giải pháp để xử lý việc dịch một văn bản từ tiếng Anh sang tiếng Việt.
- ❖ Xây dựng, thu thập dữ liệu và đào tạo mô hình để dịch một văn bản từ tiếng Anh sang tiếng Việt.
- ❖ Xây dựng trang web demo việc sử dụng mô hình để dịch một văn bản từ tiếng Anh sang tiếng Việt
- ❖ Viết 120 trang luận văn trình bày các nội dung liên quan theo đúng chuẩn nhà trường yêu cầu và có trích dẫn tài liệu tham khảo một cách chi tiết và đầy đủ.

1.5 PHẠM VI ĐỀ TÀI

- Sản phẩm được tạo ra hướng đến những người có mong muốn sử dụng các công cụ dịch.
- Sản phẩm chỉ cung cấp dịch từ tiếng Anh sang tiếng Việt.
- Sản phẩm yêu cầu kết nối internet để sử dụng.
- Sản phẩm luận văn sẽ được áp dụng và mở rộng các thư viện có sẵn cũng như các cắt thô để đáp ứng yêu cầu nhằm đạt được các mục tiêu đã đặt ra.