

BƯỚC TIỀN MỚI CHO CÔNG CỤ DỊCH THUẬT CHUNG TOÀN CẦU

Nếu bạn từng sử dụng công cụ dịch thuật internet thì bạn chắc sẽ nhận thấy rằng công nghệ dịch thuật còn kém như thế nào so với các loại chương trình máy tính khác. Lỗi chấm câu, từ để không đúng chỗ và lỗi văn phạm có thể khiến các

Nếu bạn từng sử dụng công cụ dịch thuật internet thì bạn chắc sẽ nhận thấy rằng công nghệ dịch thuật còn kém như thế nào so với các loại chương trình máy tính khác. Lỗi chấm câu, từ để không đúng chỗ và lỗi văn phạm có thể khiến các câu dịch trở nên khó hiểu.

Dịch ngôn ngữ bằng lời thậm chí còn tệ hơn. Không ai có thể phát triển một hệ thống dịch tự động có thể đi đến bất kỳ đâu như khả năng của một dịch giả hoặc thông dịch viên. Các tiện ích dịch chỉ có thể dịch những từ xác định trước nhất định và các câu dịch chỉ đủ để một người khách du lịch đặt khách sạn nhưng lại vô dụng nếu bạn muốn hiểu được một bản tin.

Người Châu Âu rất muốn có một công cụ dịch thuật chung cho tất cả mọi người. Với 23 ngôn ngữ chính thức, các Viện ở Châu Âu tiêu tốn hơn một tỉ euro một năm để dịch thuật các văn bản và phiên dịch các bài phát biểu. Các công ty kinh doanh trên khắp Châu Âu tốn hơn cả hàng triệu chỉ để hiểu đối tác kinh doanh của mình.

“Rõ ràng là có nhu cầu cho loại công nghệ này ở Châu Âu và bất cứ nơi nào trên thế giới... nó tiết kiệm thời gian và chi phí cho việc dịch thuật,” nhà nghiên cứu Marcello Federico giải thích.

Ông Federico dẫn đầu nhóm nghiên cứu đang nỗ lực để đạt được một thứ gì đó còn hơn cả tham vọng. Làm việc trong dự án TC-STAR do Châu Âu tài trợ, họ đang nghiên cứu thứ có lẽ là thách thức công nghệ ngôn ngữ loài người lớn nhất từ trước đến nay: tiếp nhận lời nói từ một ngôn ngữ và cho ra câu nói bằng một ngôn ngữ khác.

(Ảnh: www.indiamart.com)

Dịch từ lời nói sang lời nói

“Đối với con người, dịch thuật rất khó. Chúng ta phải am hiểu cả ngôn ngữ gốc và ngôn ngữ đích, và dịch thuật bằng máy thì còn khó hơn rất nhiều” Federico ghi nhận. “TC-STAR là dự án đầu tiên

trên thế giới giải quyết dịch từ lời nói sang lời nói không giới hạn.”

Đối với một hệ thống như thế có khả năng dịch từ bất cứ lời nói nào bất kể đề tài và ngữ cảnh nào, thì có ba công nghệ được sử dụng và ba công nghệ này vẫn còn chưa hoàn thiện. Công nghệ nhận dạng lời nói tự động ARS được sử dụng để phiên âm lời nói thành văn bản. Công nghệ dịch ngôn ngữ nói SLT dịch ngôn ngữ gốc ra ngôn ngữ đích. Công nghệ Văn bản ra Lời nói TTS phân tích kết quả nói ra.

Các đối tác nghiên cứu dự án TC-STAR phát triển thêm các bộ phận để xử lý một trong những nhiệm vụ đó, tạo ra nền tảng đưa công nghệ dịch tiên tiến nhất xích lại gần hơn, giống với việc dịch thuật do con người thực hiện.

Một trong những đổi mới chính của họ là kết hợp kết quả của một số hệ thống ASR và SLT lại với nhau để làm cho các câu dịch và phiên âm chính xác hơn đáng kể.

Dựa trên phương pháp BLEU, phương pháp so sánh việc dịch giữa người và máy, chất lượng dịch đã cải thiện từ 40 đến 60% trong suốt dự án trong khi đến hơn 70% số từ được dịch chính xác, cho dù chúng không được đặt ở vị trí đúng trong câu.

Từ lời nói đến bản tin tiếng Trung Quốc

11 đối tác - gồm các công ty viễn thông và giải trí như Nokia, Siemens, IBM và Sony – đã làm việc với các bản thảo âm lời nói từ Quốc Hội Châu Âu và dịch chúng giữa tiếng Anh và tiếng Tây Ban Nha. Họ còn làm việc với các chương trình phát sóng tin tức qua radio và dịch chúng từ tiếng Trung Quốc qua tiếng Anh.

Mặc dù hệ thống vẫn không thể giống với độ chính xác do người dịch, nhưng ông Federico tin rằng, với nghiên cứu sâu hơn, một máy dịch lời nói sang lời nói tự động sẽ khả thi trong vòng vài năm tới, ít nhất là đối với một số cặp ngôn ngữ đơn giản hơn.

Trong thời gian này, các bộ phận do dự án TC-STAR phát triển đã sẵn sàng với sự cho phép mã nguồn mở. Dự án này ít nhất cũng đã cho ra đời một công ty spin-off và một sáng kiến tiếp theo. (Mô hình công ty spin-off là công ty tách một phần hoạt động của mình thành lập công ty mới độc lập)

Có tên PerVoice, công ty này đang cung cấp các dịch vụ phiên âm tự động từ xa cho các công ty và các cơ quan chính quyền.

“Nó tiết kiệm cho họ tiền bạc và thời gian để phiên âm tự động các biên bản cuộc họp hoặc phiên họp hội đồng thành phố,” ông Federico ghi nhận

Dự án tiếp theo, JUMAS, tập trung vào phát triển một hệ thống phiên âm tương tự để thu âm các vụ kiện xử án tại tòa.

