

Theo quy định của cuộc thi Vietnamese Text-To-Speech VLSP 2020, các đội đăng ký cần tham gia việc thẩm định dữ liệu để chất lượng dữ liệu của cuộc thi được tốt hơn.

Để đủ điều kiện nhận dữ liệu các đội cần hoàn thành việc thẩm định dữ liệu VLSP VIETNAMESE TEXT-TO-SPEECH

Thời gian thực hiện từ 19/10/2020 đến 25/10/2020. Mỗi đội dự kiến thẩm định tổng 30 phút audio.

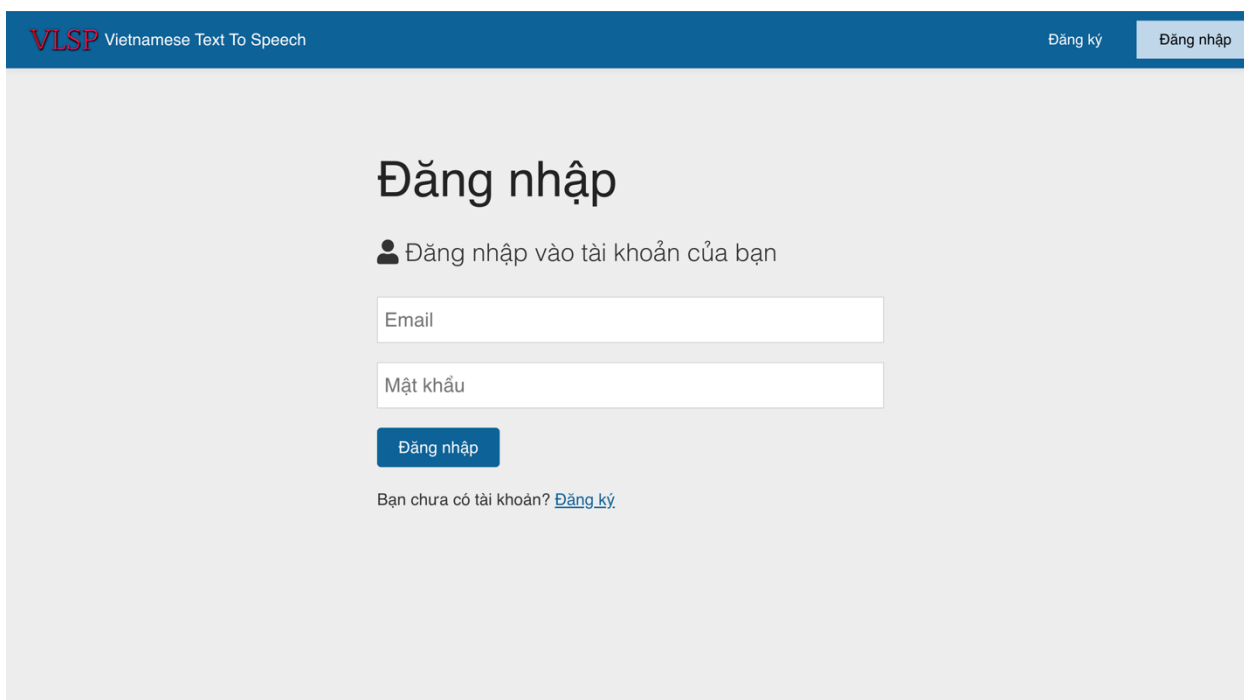
Theo kế hoạch đã thông báo, tập dữ liệu huấn luyện sẽ được công bố vào ngày 31/10/2020 cho các đội đã hoàn thành việc thẩm định

Dưới đây là tài liệu hướng dẫn thẩm định dữ liệu.

TÀI LIỆU HƯỚNG DẪN THẨM ĐỊNH DỮ LIỆU

1. Đăng nhập

- Sau khi được cấp tài khoản, click vào đường link <https://tts.vlsp.org.vn/login> để tiến hành đăng nhập.



The screenshot shows the login interface of the VLSP Vietnamese Text To Speech website. At the top, there is a blue header bar with the VLSP logo and the text 'Vietnamese Text To Speech' on the left, and two buttons labeled 'Đăng ký' (Register) and 'Đăng nhập' (Login) on the right. The main content area has a light gray background. In the center, the title 'Đăng nhập' (Login) is displayed in a large, bold font. Below the title, there is a subtitle 'Đăng nhập vào tài khoản của bạn' (Login to your account) preceded by a user icon. There are two input fields: 'Email' and 'Mật khẩu' (Password). Below these fields is a blue button labeled 'Đăng nhập'. At the bottom, there is a link that says 'Bạn chưa có tài khoản? [Đăng ký](#)' (You don't have an account? [Register](#)).


Sau đăng nhập thành công, bạn cần thay đổi mật khẩu để tiếp tục sử dụng hệ thống

VLSP Vietnamese Text To Speech

Xin chào, **TEAM1**

Đổi mật khẩu

Đăng xuất

 **Đổi mật khẩu**

Xác nhận

Sau khi đổi mật khẩu thành công, bạn sẽ được chuyển đến trang <https://tts.vlsp.org.vn/team/competitions>. Tại đây, sẽ xuất hiện 2 vùng:

- **Công việc bạn cần hoàn thành**, ở đây là tham gia thẩm định dữ liệu cho cuộc thi VLSP VIETNAMESE TEXT-TO-SPEECH
- **Tài nguyên được chia sẻ**, đây là phần dữ liệu bạn sẽ được chia sẻ sau khi bạn hoàn thành việc thẩm định dữ liệu

VLSP Vietnamese Text To SpeechXin chào, **TEAM1**Đổi mật khẩuĐăng xuất

Công việc cần hoàn thành

VLSP VIETNAMESE TEXT-TO-SPEECH	17 - 10 - 2020	Thực hiện tiếp
--------------------------------	----------------	----------------

< 1 >

Tài nguyên được chia sẻ

DỮ LIỆU HUẤN LUYỆN CHO CUỘC THI VLSP TTS 2020	Download
---	--------------------------

< 1 >

2. Trạng thái công việc


- Đối với các “Công việc cần hoàn thành”, bạn có thể:
 - “Tham gia”, công việc hiển thị nút này nếu bạn chưa bắt đầu tham gia xây dựng dữ liệu cho cuộc thi. Khi nhấn vào nút này, bạn sẽ bắt đầu tiến hành xây dựng dữ liệu.
 - “Thực hiện tiếp”: Công việc hiển thị nút này nếu bạn đang thực hiện xây dựng dữ liệu nhưng chưa xong. Bạn có thể nhấn vào nút “Thực hiện tiếp” để tiếp tục xây dựng dữ liệu. Hệ thống sẽ chuyển sang câu đầu tiên bạn chưa làm tính từ đầu danh sách.
 - Với những công việc bạn đã thực hiện xong, bạn sẽ không thể thực hiện thao tác gì trên đó. Vì vậy, công việc này mang tính chất liệt kê và có trạng thái là “Đã hoàn thành”.

3. Thẩm định dữ liệu

VLSP Vietnamese Text To Speech

Xin chào, **TEAM1** [Đổi mật khẩu](#) [Đăng xuất](#)

VLSP VIETNAMESE TEXT-TO-SPEECH

 **Hướng dẫn chỉnh sửa nội dung audio**

- Số, ngày tháng, giờ được viết dạng chữ số
VD: *mười lăm* → 15, *ngày hai mươi tháng ba* → ngày 20/3
- Từ nước ngoài được viết theo từ gốc
VD: *Hemingway, Jordan* (xem danh sách các từ tham khảo [tại đây](#))
- Các loại khác được viết về dạng nguyên bản
VD: *xen-li-mét* → cm
- Các trường hợp khó tìm dạng nguyên bản có thể viết dạng phiên âm tiếng Việt
VD: *phát-xít, xi-măng*

Bắt đầu nào


0:00 / 0:06

No Data


Nhập nội dung audio

- Bên trái là phần Hướng dẫn chỉnh sửa nội dung audio, bên phải là phần thẩm định dữ liệu. Bạn có thể đóng phần hướng dẫn để tập trung thẩm định dữ liệu bằng cách click vào button collapse



VLSP Vietnamese Text To Speech



Xin chào, **TEAM1** [Đổi mật khẩu](#) [Đăng xuất](#) 

VLSP VIETNAMESE TEXT-TO-SPEECH

 **Bạn đã hoàn thành 4 / 20 câu**

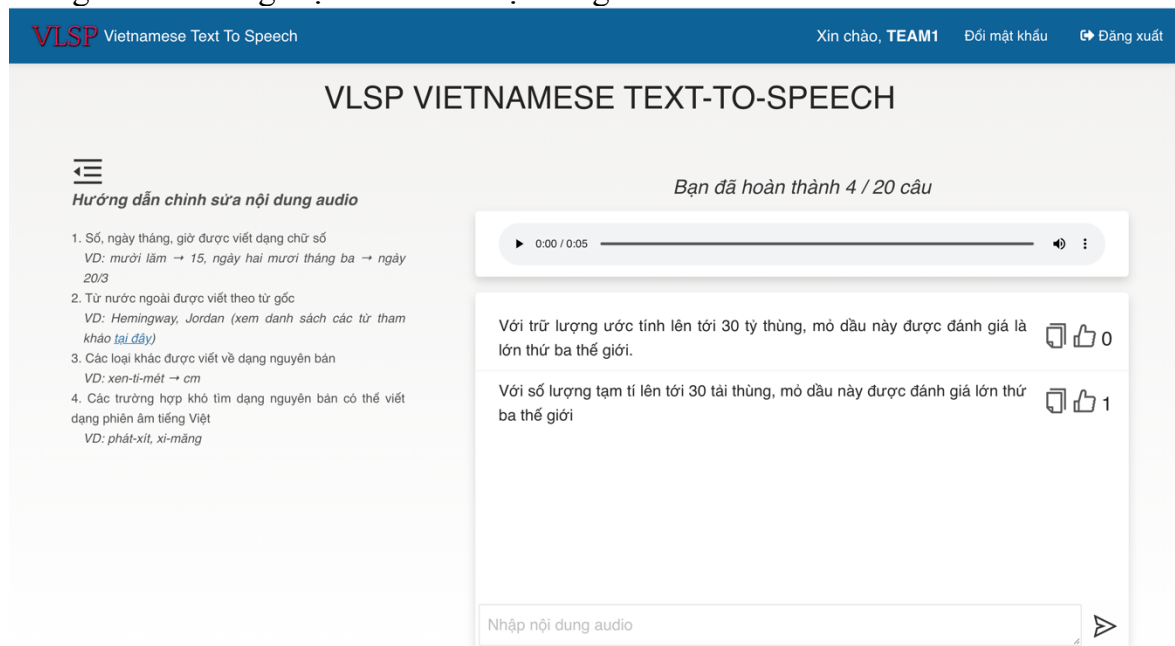
0:00 / 0:05

Với trữ lượng ước tính lên tới 30 tỷ thùng, mỏ dầu này được đánh giá là lớn thứ ba thế giới.   0

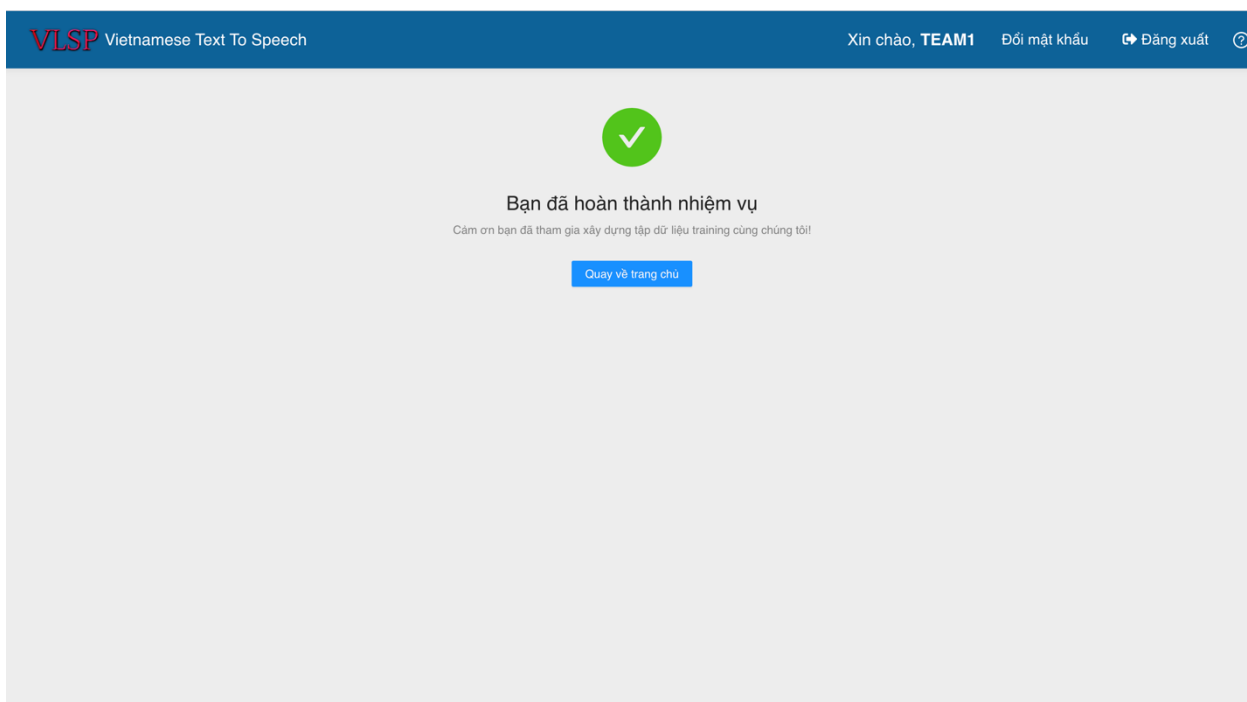
Với số lượng tạm tí lên tới 30 tài thùng, mỏ dầu này được đánh giá lớn thứ ba thế giới   1

Nhập nội dung audio

- Đối với phần thẩm định dữ liệu, sẽ có audio để nghe và nhiệm vụ của bạn sẽ là nghe audio và đưa ra nội dung của audio.
- Sẽ có các trường hợp xảy ra:
 - o TH1: Nếu audio không có bất cứ text nào đi kèm (như hình trên), bạn sẽ cần nghe audio và gõ lại chính xác nội dung audio



- o TH2: Nếu audio đã có text đi kèm (như hình trên), bạn sẽ cần nghe audio và kiểm tra xem có đoạn text nào đúng với nội dung audio mà bạn nghe không? Đoạn text nào đúng, bạn hãy click vào “nút like” tương ứng với đoạn text để vote cho đoạn text đó và chuyển sang câu tiếp theo. Nếu trong tất cả các đoạn text đi kèm mà không có đoạn text nào đúng, bạn cần gõ lại nội dung audio đúng nhất với những gì mà bạn nghe được.
- Trong quá trình chỉnh sửa, bạn có thể sử dụng tính năng copy nhanh text thông qua “icon copy” (cạnh nút like) để tiến hành copy và sửa trực tiếp trên bản copy đó.
- Sau khi hoàn thành số lượng câu được yêu cầu, bạn sẽ nhận được thông báo hoàn thành nhiệm vụ:



4. Hướng dẫn chỉnh sửa dữ liệu

- Số, ngày tháng, giờ được viết dạng chữ số
 - o VD: mười lăm → 15, ngày hai mươi tháng ba → ngày 20/3 2.
- Từ nước ngoài được viết theo từ gốc
 - o VD: Hemingway, Jordan (xem danh sách các từ tham khảo tại đây)
- Các loại khác được viết về dạng nguyên bản
 - o VD: xen-ti-mét → cm
- Các trường hợp khó tìm dạng nguyên bản có thể viết dạng phiên âm tiếng Việt
 - o VD: phát-xít, xi-măng