BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH**

**BÁO CÁO TỔNG KẾT**

**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN**

**ĐỀ TÀI: 475**

**ỨNG DỤNG SOẠN THẢO VĂN BẢN**

**CHO NGƯỜI KHIẾM THỊ**

**Hồ Chí Minh, Tháng 4 Năm 2022**

BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH**

**BÁO CÁO TỔNG KẾT**

**ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN**

**ĐỀ TÀI: 475**

**ỨNG DỤNG SOẠN THẢO VĂN BẢN**

**CHO NGƯỜI KHIẾM THỊ**

**Chủ nhiệm đề tài: Phạm Nhật Khoa - 1951012053**

**Khoa: Công Nghệ Thông Tin**

**Các Thành Viên:**

**Đặng Lê Đăng Khoa - 1951012051**

**Lưu Gia Huy - 1951012038**

**Người hướng dẫn: ThS Nguyễn Thị Mai Trang**

**Hồ Chí Minh, Tháng 4 Năm 2022**

|  |  |
| --- | --- |
| BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC MỞ**  **THÀNH PHỐ HỒ CHÍ MINH** | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc** |

**THÔNG TIN KẾT QUẢ NGHIÊN CỨU CỦA ĐỀ TÀI**

**1. Thông tin chung:**

**-** Tên đề tài: ỨNG DỤNG SOẠN THẢO VĂN BẢN CHO NGƯỜI KHIẾM THỊ

- Sinh viên thực hiện: Phạm Nhật Khoa

- Lớp: DH19CS02 Khoa: Công Nghệ Thông Tin Năm thứ: 3 Số năm đào tạo: 3

- Người hướng dẫn: ThS Nguyễn Thị Mai Trang

**2. Mục tiêu đề tài:**

Xây dựng một công cụ hỗ trợ soạn thảo văn bản tiếng Việt thông minh giúp người dùng có thể sử dụng máy tính soạn thảo văn bản bằng giọng nói, đặc biệt rất hữu dụng đối với người khiếm thị, nhà văn, soạn giả,.... Ứng dụng có khả năng có thể được phát triển và cải thiện để tăng độ chính xác áp dụng mở rộng cho nhiều đối tượng người dùng và có khả năng tích hợp vào các công cụ, dịch vụ khác.

Nhóm nghiên cứu hi vọng công trình này sẽ giúp đỡ nâng cao hiệu suất công việc, góp phần vào việc giúp người Việt có một công cụ để tham gia, kết nối với nhiều tính năng, thiết bị trên toàn cầu. Mặt khác, công cụ cũng hỗ trợ rất nhiều việc soạn thảo văn bản của người dùng là người khiếm thị, tạo điều kiện để họ có thể dễ dàng học tập, làm việc và gắn kết với cộng đồng.

**3. Tính mới và sáng tạo:**

Xây dựng một bộ Speech-To-Text với thanh âm – thanh điệu hỗ trợ tối đa và phù hợp với quy chuẩn hóa Tiếng Việt, kết hợp các khẩu lệnh hỗ trợ thao tác soạn thảo như xuống dòng, đặt dấu cuối câu, ngắt câu,… chưa có trong các ứng dụng tương tự.

Ứng dụng soạn thảo văn bản bằng giọng nói với chi phí giá tiết kiệm cho người dùng, không yêu cầu thêm bất kỳ một thiết bị phần cứng nào ngoài máy tính. Tích hợp các thao tác xử lý văn bản thông qua giọng nói, hỗ trợ tốt cho người dùng tìm kiếm và soạn thảo văn bản mà không cần sự trợ giúp của người khác

**4. Kết quả nghiên cứu:**

Hoàn thành phiên bản đầu tiên của phần mềm Soạn Thảo Văn Bản Bằng Giọng Nói Speech-To-Text với các chức năng cơ bản như: soạn thảo văn bản bằng giọng nói, soạn thảo văn bản bằng file âm thanh.

Tích hợp các tiện ích giúp cho người dùng, đặc biệt là người khiếm thị soạn thảo văn bản một cách dễ dàng hơn.

**5. Đóng góp về mặt kinh tế - xã hội,** **giáo dục và đào tạo, an ninh, quốc phòng và khả năng áp dụng của đề tài:**

Trải qua quá trình tìm hiểu cũng như triển khai dự án, nhóm nghiên cứu mong muốn ứng dụng này sẽ trở nên phổ biến cũng như rộng rãi dành cho nhiều người dùng. Ứng dụng giúp rút ngắn được thời gian soạn thảo văn bản cũng như giúp việc nhập liệu, soạn thảo văn bản trở nên dễ dàng và nhanh chóng. Ứng dụng sớm được tích hợp vào nhiều hệ thống, dịch vụ khác.

Nghiên cứu này cũng có thể dùng làm tài liệu tham khảo cho các bạn sinh viên khóa sau.

**6.** **Công bố khoa học của sinh viên từ kết quả nghiên cứu của đề tài** *(ghi rõ tên tạp chí nếu có)* hoặc nhận xét, đánh giá của cơ sở đã áp dụng các kết quả nghiên cứu *(nếu có)*:

*Ngày        tháng        năm*

|  |  |
| --- | --- |
|  | **Sinh viên chịu trách nhiệm chính**  **thực hiện đề tài**  *(ký, họ và tên)* |

**Nhận xét của người hướng dẫn về những đóng góp khoa học của sinh viên thực hiện đề tài** *(phần này do người hướng dẫn ghi):*

*Ngày        tháng        năm*

|  |  |
| --- | --- |
| **Xác nhận của đơn vị**  *(ký tên và đóng dấu)* | **Người hướng dẫn**  *(ký, họ và tên)* |
|  |  |

|  |  |
| --- | --- |
| BỘ GIÁO DỤC VÀ ĐÀO TẠO  **TRƯỜNG ĐẠI HỌC MỞ**  **THÀNH PHỐ HỒ CHÍ MINH** | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc** |

**THÔNG TIN VỀ SINH VIÊN**

**CHỊU TRÁCH NHIỆM CHÍNH THỰC HIỆN ĐỀ TÀI**

**I. SƠ LƯỢC VỀ SINH VIÊN:**

Họ và tên: Phạm Nhật Khoa

Sinh ngày:        19      tháng      8      năm 2001

Nơi sinh: Cần Thơ

Lớp:   DH19CS02                                  Khóa: 2019

Khoa: Công nghệ thông tin

Địa chỉ liên hệ: 88/42 Bạch Đằng, p24 Bình Thạnh, TPHCM

Điện thoại:  0762868381                Email: 1951012053khoa@ou.edu.vn

**II. QUÁ TRÌNH HỌC TẬP** (kê khai thành tích của sinh viên từ năm thứ 1 đến năm đang học):

***\* Năm thứ 1:***

    Ngành học:    Khoa học máy tính        Khoa: Công nghệ thông tin

    Kết quả xếp loại học tập: Điểm tích luỹ hệ số 4: 2,89

    Sơ lược thành tích: Đạt học bổng khuyến khích học tập HK3 2019-2020

***\* Năm thứ 2:***

    Ngành học:    Khoa học máy tính        Khoa: Công nghệ thông tin

Kết quả xếp loại học tập: Điểm tích luỹ hệ số 4: 2,65

   Sơ lược thành tích:

*Ngày 03 tháng 04 năm 2022*

|  |  |
| --- | --- |
| **Xác nhận của đơn vị**  *(ký tên và đóng dấu)* | **Sinh viên chịu trách nhiệm chính**  **thực hiện đề tài**  *(ký, họ và tên)* |

(đã kí)

**Phạm Nhật Khoa**

**MỤC LỤC**

[I. PHẦN MỞ ĐẦU 1](#_Toc100310517)

[1.1. Tổng quan đề tài 1](#_Toc100310518)

[1.2. Lý do chọn đề tài 2](#_Toc100310519)

[1.3. Mục tiêu đề tài 3](#_Toc100310520)

[1.4. Cách tiếp cận 3](#_Toc100310521)

[1.5. Phương pháp nghiên cứu 4](#_Toc100310522)

[1.6. Đối tượng và phạm vi nghiên cứu 4](#_Toc100310523)

[II. KẾT QUẢ NGHIÊN CỨU VÀ PHÂN TÍCH 4](#_Toc100310524)

[2.1. Các nghiên cứu liên quan 4](#_Toc100310525)

[2.2. Tổng quan về Google Speech-To-Text API 5](#_Toc100310526)

[2.3. Cách đăng ký Google Speech-To-Text API 5](#_Toc100310527)

[2.4. Ứng dụng API vào dự án Speech-To-Text 13](#_Toc100310528)

[2.5. Giao diện ứng dụng Speech-To-Text 18](#_Toc100310529)

[*2.5.1.* Các *trang chính của ứng dụng* 18](#_Toc100310530)

[*2.5.2.* *Cách sử dụng ứng dụng Speech-To-Text* 19](#_Toc100310531)

[III. KẾT LUẬN VÀ KIẾN NGHỊ 22](#_Toc100310532)

[3.1. Kết luận 22](#_Toc100310533)

[*3.1.1.* *Kết quả đạt được:* 22](#_Toc100310534)

[*3.1.2.* *Những vấn đề còn tồn tại:* 22](#_Toc100310535)

[3.2. Kiến nghị và hướng phát triển 22](#_Toc100310536)

[IV. TÀI LIỆU THAM KHẢO 24](#_Toc100310537)

1. **PHẦN MỞ ĐẦU**
   1. **Tổng quan đề tài**

Trong 10 năm trở lại đây, việc ứng dụng các Công Nghệ mới mà đặc biệt có thể kể đến là trí tuệ nhân tạo (Intelligent Artificial) không còn bị giới hạn trong khuôn khổ các khối ngành kỹ thuật mà còn được áp dụng trong cuộc sống thường ngày của mỗi chúng ta. Công nghệ nhận dạng tiếng nói chuyển thành văn bản (Speech-To-Text) cũng vì thế mà phát triển nhanh chóng.

Hệ thống Speech to Text (STT) là một hệ thống dựa trên phần mềm AI tự động chuyển đổi lời nói của con người thành văn bản thuần túy. Hệ thống Spechc-To-Text chuyển đổi ngôn ngữ bình thường thành văn bản. Người dùng có thể sử dụng trực tiếp từ các thiết bị thu âm thanh cũng như từ tệp âm thanh để chuyển thành tệp văn bản.

Speech-To-Text có thể nói là công nghệ quan trọng nhất trong cầu nối tương tác thông qua giọng nói giữa người dùng với các thiết bị - hệ thống điện tử như: Google Assistant, Alexa, Smart Tivi hay các ứng dụng Smart Home như Echo, Alexa, Google Home. Thay vì bạn phải thao tác với các remote hay điện thoại, giờ đây bạn chỉ việc đọc các câu lệnh bằng giọng nói, Speech-To-Text sẽ xử lý câu lệnh và hiện thực nó cho bạn. Đối với các công ty lớn như Google hay Microsoft, việc phát triển một hệ thống Speech-To-Text mạnh mẽ là một trong những hướng phát triển trọng tâm.

Việc phát triển các hệ thống Speech-To-Text cho các thị trường khác nhau là một điều không hề đơn giản, khi mà mỗi thị trường Quốc Gia khác nhau lại có một ngôn ngữ riêng biệt, không quốc gia nào giống quốc gia nào. Vì vậy việc đa dạng hóa các hệ thống Speech-To-Text sao cho phù hợp với các Quốc Gia khác nhau là một thách thức không nhỏ cho nhà phát triển.

Ngày nay, các công ty và nhà phát triển trên khắp thế giới phát triển nhiều hệ thống Text-To-Speech - Hệ thống gắn liền với Speech-To-Text. Hai hệ thống này hỗ trợ lẫn nhau trong xử lý. Mỗi hệ thống được phát triển theo cách riêng của nhà phát triển, có những ưu nhược điểm riêng nhưng tất cả đều nhằm mục đích có thể chuyển văn bản thành giọng nói gần giống với giọng nói thực tế của con người nhất có thể. Qua đó hỗ trợ việc xử lý giọng nói đầu vào của hệ thống Speech-To-Text chính xác hơn.

Trên các thiết bị Android tiêu biểu là ứng dụng Voice Text cho phép bạn gửi và nhận tin nhắn bằng giọng nói. Ứng dụng này còn được nhà phát triển tích hợp thêm nhiều câu lệnh tiện ích giúp cho bạn có thể điều khiển điện thoại một cách dễ dàng hơn. Còn đối với các thiết bị IOS Dictate – Speech to text là một ứng dụng nổi bật nhưng đi kèm với đó là một mức giá khá đắt.

Đối với các máy tính chạy hệ điều hành Window th ứng dụng này còn nhiều mặt hạn chế. Microsoft cũng đã tích hợp một ứng dụng Speech-To-Text vào phần mềm Microsoft Word của mình. Tuy nhiên việc nhập văn bản bằng giọng nói chưa được tối ưu cho Tiếng Việt, cũng như các thao tác soạn thảo văn bản chưa có, người dùng hoàn toàn phải làm các thao tác với bàn phím như xuống hàng hay xóa từ đã nhập.

Các hệ thống Speech-To-Text được phát triển hiện nay chưa được tối ưu hoàn toàn cho Tiếng Việt. Theo các nhà nghiên cứu ngôn ngữ học, Tiếng Việt có khoảng 6 thanh điệu và hơn 6500 âm tiết. Các thanh điệu và âm tiết lại thay đổi theo vùng miền đi cùng với giọng nói, vì vậy tối ưu hóa hoàn toàn cho Tiếng Việt là một thách thức lớn.

Công nghệ phát triển, các thiết bị phần mềm để hỗ trợ người khuyết tật cũng được phát triển nhiều hơn. Ngày càng nhiều các công ty công nghệ hướng đến đối tượng người dùng là người khuyết tật. Điển hình là các ứng dụng như TapTap See, RightHear…. Hệ thống Speech-To-Text đã có nhiều công ty phát triển, điển hình là hệ thống Kurzweii 1000 là một thiết bị hỗ trợ tiêu biểu đã từng đạt các giải thưởng cao trên thế giới.

Tại Việt Nam, Công ty CP công nghệ I-Tek Việt Nam cũng đã nghiên cứu sản xuất ra thiết bị “Chuột thông minh” để hỗ trợ nhu cầu này, nhưng ở thời điểm hiện tại giá của thiết bị vẫn khá cao. Với giá thành xấp xỉ một triệu không phải là một mức giá dễ tiếp cận đối với mọi người, chưa kể đó là một thiết bị gắn ngoài và cần phải có thiết bị phần cứng – phần mềm phù hợp mới có thể hoạt động hiệu quả. Việc cài đặt phần mềm và phần cứng cho chuột I- Tek là điều khó khăn với những người bị khiếm thị.

Còn về các ứng dụng hỗ trợ người khiếm thị, hiện nay ở Việt Nam cũng đã có nhiều công ty phát triển các ứng dụng đọc màn hình hỗ trợ người khiếm thị như NVDA (NonVisual Desktop Access) với một mức chi phí khá cao. Tuy nhiên vấn đề được đặt ra ở đây là: làm sao người khiếm thị có thể tìm được một vấn đề mà họ cần để ứng dụng đọc màn hình hỗ trợ đọc kết quả? Hầu như để sử dụng được phần mềm đọc màn hình để đọc kết quả tìm kiếm hay trang web, người khiếm thị phải có một người khác hỗ trợ cho họ, họ hoàn toàn không thể chủ động trong việc tìm kiếm tài liệu hỗ trợ học tập và làm việc, cũng như soạn thảo văn bản.

Vấn đề đặt ra của đề tài là giải quyết vấn đề soạn thảo văn bản hỗ trợ người khiếm thị chỉ với một phần mềm ứng dụng cài đặt trên máy tính mà không phụ thuộc vào bất kỳ thiết bị khác. Dễ dàng tiếp cận và thân thiện với người khiếm thị.

* 1. **Lý do chọn đề tài**

Với sự phát triển của Internet và công nghệ, các công cụ hỗ trợ công việc ngày càng phổ biến, đòi hỏi sự cải thiện không ngừng để đáp ứng được yêu cầu về mặt thời gian. Chính vì thế, việc tiết kiệm thời gian và thuận tiện cho việc biên soạn văn bản sẽ góp phần to lớn trong việc nâng cao và tối ưu hoá chất lượng công việc như viết sách, làm báo, công việc văn phòng nói chung và hỗ trợ tối đa việc nhập văn bản cho người khiếm thị nói riêng. Được sử dụng máy tính và soạn thảo văn bản đối với người khiếm thị là một cơ hội to lớn, có thể giúp cho họ có được nhiều cơ hội việc làm hơn trong tương lai.

Được truyền cảm hứng từ các chủ đề nghiên cứu và tài liệu nghiên cứu toàn cầu về "Speech to text", nhóm nghiên cứu nhận thấy công cụ này rất hữu ích, tuy nhiên phần nội dung chưa được tối ưu hoá triệt để cho ngôn ngữ Tiếng Việt, các kết quả trả về chỉ mang tính tương đối. Và để hỗ trợ việc soạn thảo văn bản cho người dùng là người khiếm thị nên nhóm quyết định nghiên cứu ứng dụng soạn thảo văn bản Tiếng Việt bằng giọng nói nhằm hỗ trợ đối tượng người dùng là người khiếm thị.

* 1. **Mục tiêu đề tài**

Xây dựng một công cụ hỗ trợ soạn thảo văn bản tiếng Việt thông minh giúp người dùng có thể sử dụng máy tính soạn thảo văn bản bằng giọng nói, đặc biệt rất hữu dụng đối với người khiếm thị, nhà văn, soạn giả, ... Ứng dụng có khả năng có thể được phát triển và cải thiện để tăng độ chính xác áp dụng mở rộng cho nhiều đối tượng người dùng và có khả năng tích hợp vào các công cụ, dịch vụ khác.

Nhóm nghiên cứu cũng đặt mục tiêu nghiên cứu phát triển ứng dụng lâu dài, cập nhập những tính năng cũng như công nghệ mới định kỳ giúp cho người dùng có được trải nghiệm tốt nhất sau này. Nghiên cứu trải nghiệm người dùng khiếm thính từ đó cải thiện ứng dụng.

Nhóm nghiên cứu hi vọng công trình này sẽ giúp đỡ nâng cao hiệu suất công việc, góp phần vào việc giúp người Việt có một công cụ để tham gia, kết nối với nhiều tính năng, thiết bị trên toàn cầu. Mặt khác, công cụ cũng hỗ trợ đắc lực cho việc soạn thảo văn bản của đối tượng người dùng là người khiếm thị. Từ đó mở ra cơ hội học tập và việc làm cho người khiếm thị.

* 1. **Cách tiếp cận**

Để xây dựng một ứng dụng Speech-To-Text hoàn chỉnh, đầu tiên nhóm nghiên cứu phải tiến hành lựa chọn API Speech-To-Text. Các chức năng chính của ứng dụng sẽ bao gồm:

Soạn thảo văn bản bằng giọng nói.

Soạn thảo văn bản bằng File Âm Thanh.

Nhóm đối tượng được hướng tới trong đề tài này là những người có nhu cầu soạn thảo văn bản bằng giọng nói. Vì vậy các chức năng cần đảm bảo dễ dàng tiếp cận cũng như dễ dàng sử dụng đối với những người có nhu cầu. Nhóm nghiên cứu đã thảo luận thêm những chức năng sau:

Giao diện dễ sử dụng và thân thiện với mọi đối tượng người dùng.

Sử dụng phím tắt thông dụng để bắt đầu, tạm dừng chương trình.

Nghiên cứu phát triển ứng dụng nhận lệnh giọng nói trực tiếp từ người dùng.

Phát triển các bộ công cụ soạn thảo văn bản thông minh.

Nghiên cứu phát triển một bộ phần mềm đa nền tảng, đa chức năng xoay quanh chức năng cơ bản là soạn thảo văn bản bằng giọng nói.

* 1. **Phương pháp nghiên cứu**

Nhóm nghiên cứu sử dụng Phương pháp phân tích và tổng hợp lý thuyết làm phương pháp nghiên cứu chính của đề tài, cụ thể như sau:

Thu thập và tổng hợp tài liệu về cơ sở lý thuyết của hệ thống Speech-To-Text.

Nghiên cứu và tối ưu hóa bộ nhận dạng Tiếng Việt cho hệ thống.

Xây dựng phần mềm soạn thảo văn bản Speech-To-Text.

* 1. **Đối tượng và phạm vi nghiên cứu**

Hệ thống Speech-To-Text dành cho người Việt là đối tượng nghiên cứu chính của đề tài.

Trong nghiên cứu hệ thống Speech-To-Text dành cho người Việt, nhóm nghiên cứu xác định những người dùng có nhu cầu soạn thảo văn bản Tiếng Việt nói chung và người khiếm thị nói riêng là phạm vi nghiên cứu của đề tài.

1. **KẾT QUẢ NGHIÊN CỨU VÀ PHÂN TÍCH**
   1. **Các nghiên cứu liên quan**

Công nghệ Speech-To-Text hiện nay được tích hợp rất nhiều trong các lĩnh vực, đặc biệt trong các hệ điều hành dành cho các thiết bị smartphone như Android, IOS. Tuy nhiên, các ứng dụng này chỉ đơn thuần nhận vào giọng nói và xuất ra văn bản trên các text editor. Việc bỏ dấu, ngắt câu, ngắt dòng phải do người dùng tự hiệu chỉnh nên chỉ phù hợp trong việc tìm kiếm từ khóa và những tin nhắn ngắn gọn.

Có khá nhiều Speech-To-Text Software như: Google Speech To Text API, AI Speech To Text FPT, Microsoft Speech To Text, Speech Recognition API, ….Trong đó, Google là công ty tiên phong trong công nghệ Speech-To-Text, với nền tảng hệ thống ngôn ngữ có sẵn của Google Translate nên Google Speech-To-Text API có thể hỗ trợ tối ưu cho Tiếng Việt. Nhận thấy sự mạnh mẽ của Google Speech-To-Text API, nhóm nghiên cứu quyết định sử dụng Google Speech-To-Text API để có được sự tối ưu hóa tốt nhất cho đề tài.

* 1. **Tổng quan về Google Speech-To-Text API**

Google Cloud Platform là một dịch vụ điện toán đám mây được cung cấp bởi Google. Nó cung cấp một loạt các dịch vụ đám mây mô-đun bao gồm điện toán, lưu trữ dữ liệu, phân tích dữ liệu và máy học. Trong số đó còn có cả AI trên đám mây bao gồm các dịch vụ Cloud AutoML, Cloud Machine Learning Engine, Cloud Natural Language, Cloud Speech-to-Text, Cloud Translation API, …Và ở đề tài nghiên cứu khoa học này nhóm nghiên cứu đã sử dụng dịch vụ Cloud Speech-To-Text để chuyển văn bản thành giọng nói vậy nên giờ chúng ta sẽ đi vào tìm hiểu về dịch vụ Speech To Text của Google và cách đăng ký dịch vụ đó.

Google Speech-To-Text được ra mắt chính thức vào năm 2018. Một trong những lý do Google Speech-To-Text có độ chính xác cao đó là do khả năng lựa chọn giữa các mô hình máy học khác nhau, tùy thuộc vào ứng dụng mà bạn đang sử dụng, chính điều đó đã làm cho Google Speech-To-Text trở thành một giải pháp phù hợp cho nhiều dự án khác nhau. Ngoài ra, Google Speech-To-Text còn có các tùy chọn chấm câu mở rộng, điều đó giúp vấn đề lỗi chấm câu được cải thiện đáng kể.

Google Speech-To-Text không miễn phí. Với thời lượng âm thanh đầu vào dưới 60 phút thì sẽ được miễn phí còn trên 60 phút thì Google sẽ tính phí 0.006 USD mỗi 15 giây. Còn đối với âm thanh từ video thì sẽ tính phí 0.006 USD/15 giây với video dưới 60 phút còn trên 60 phút thì sẽ có giá 0.012 USD mỗi 15 giây.

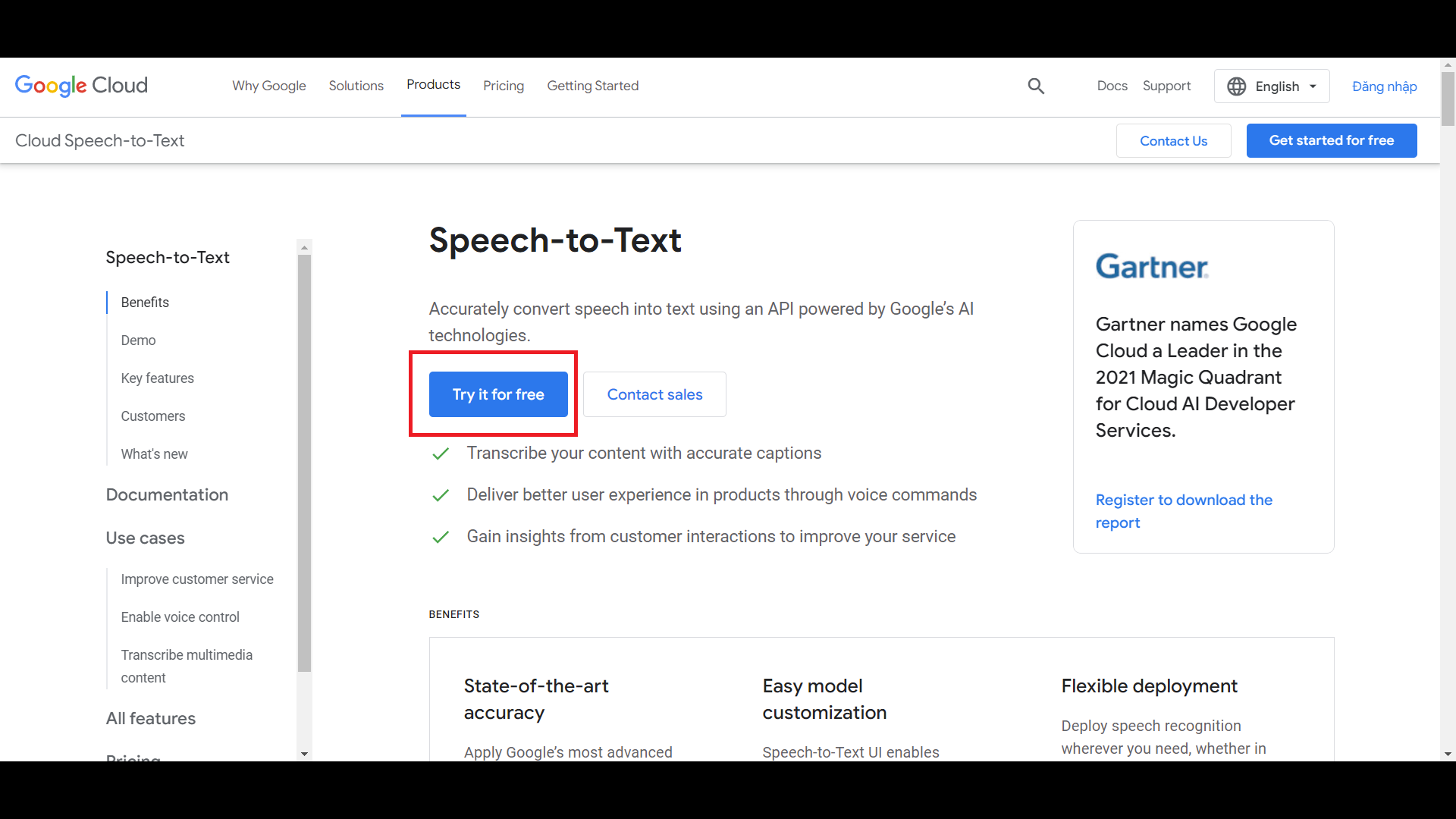
Ưu điểm:

* Hỗ trợ hơn 120 ngôn ngữ.
* Có nhiều mô hình máy học để tăng độ chính xác đầu ra văn bản.
* Quyền riêng tư dữ liệu.
* Có thể khử được tiếng ồn từ cuộc gọi và video
* Có nền tảng mạnh mẽ

Nhược điểm:

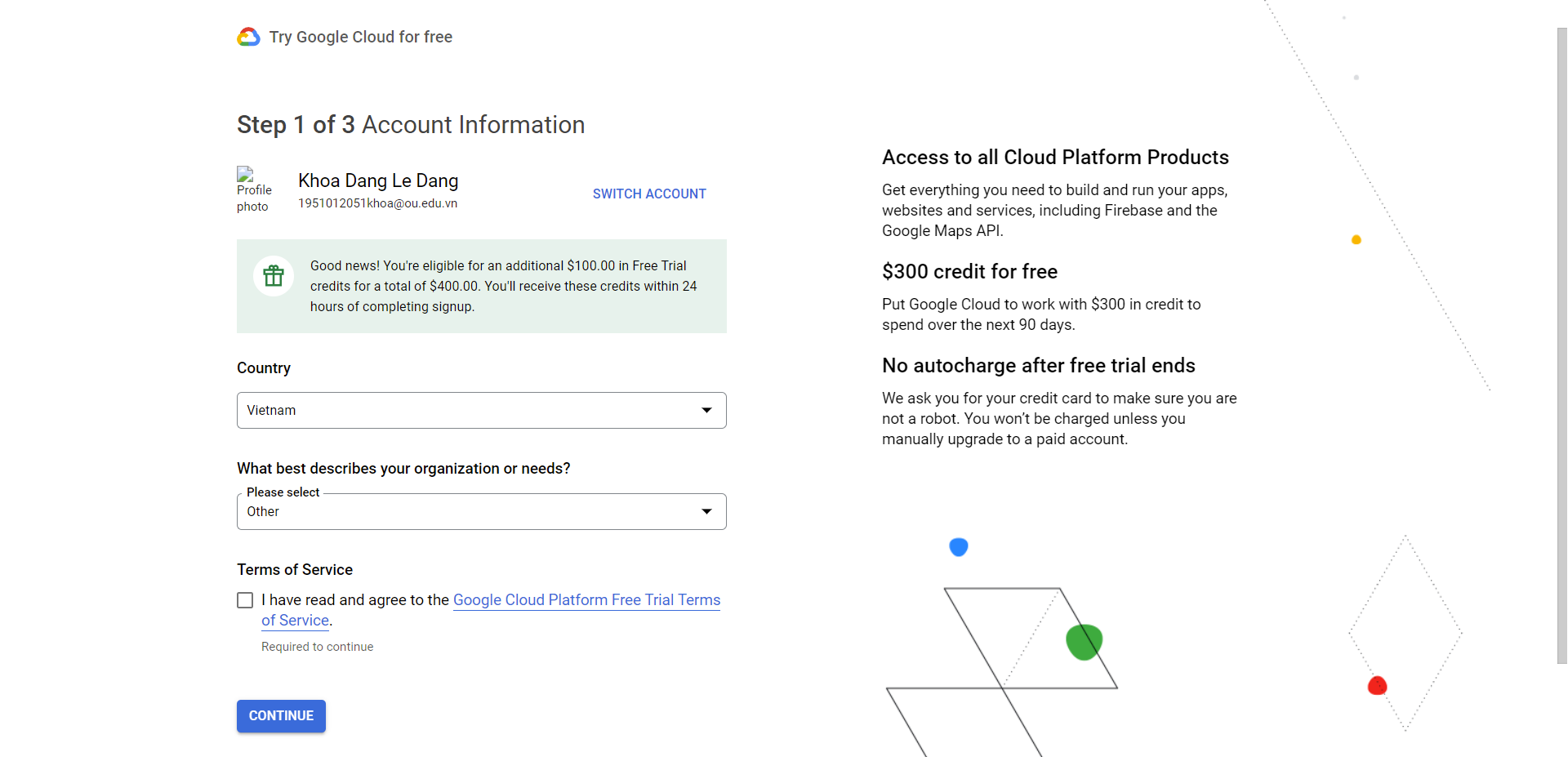
* Không miễn phí
* Nguồn từ vựng Tiếng Việt còn hạn chế
  1. **Cách đăng ký Google Speech-To-Text API**

Đầu tiên là chúng ta truy cập vào trang https://cloud.google.com/speech-to-text/ Lựa chọn Try it for free và tiến hành đăng nhập bằng tài khoản Google



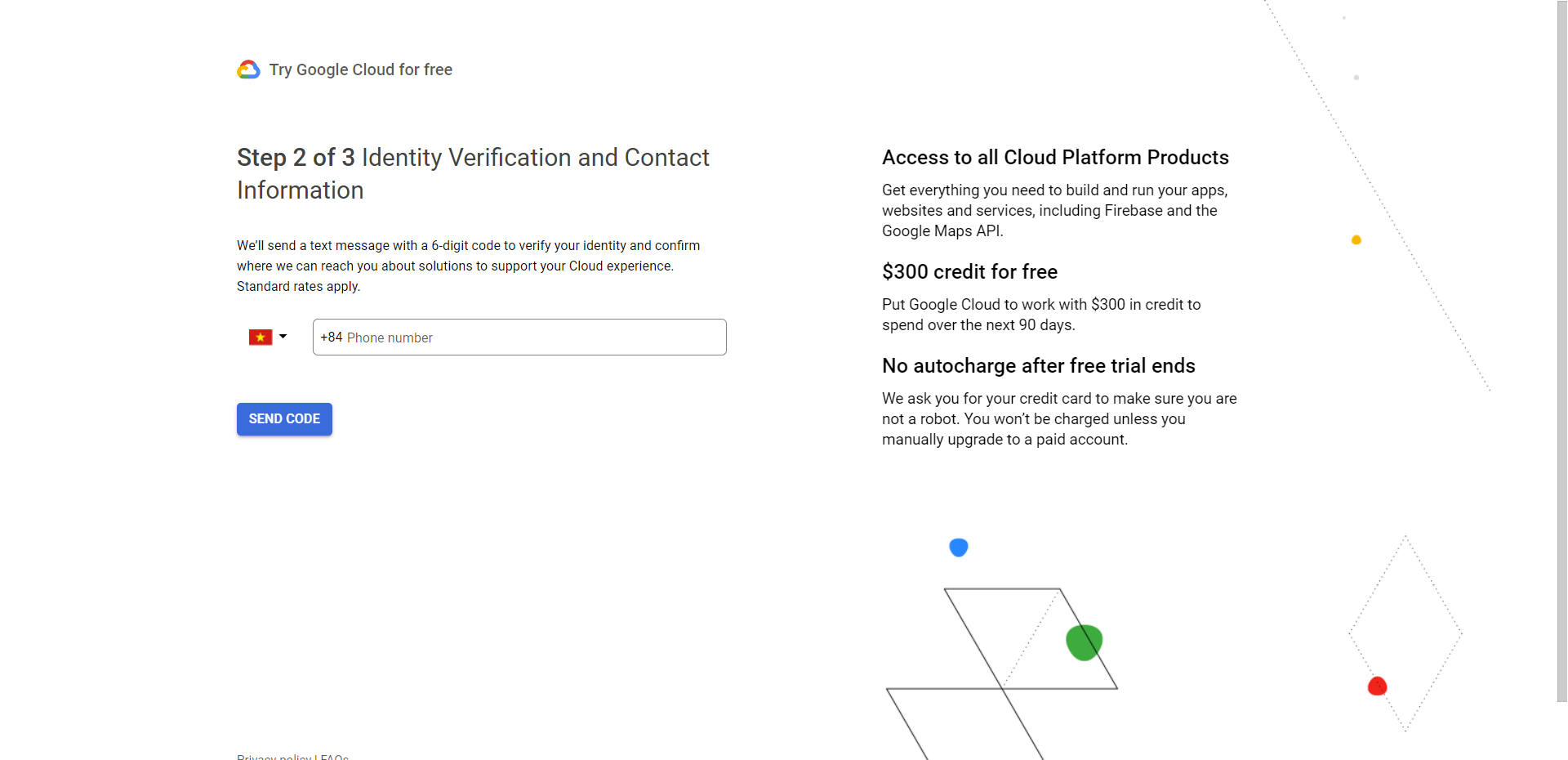
Trang chủ API Speech-To-Text

Sau đó chúng ta chọn Country, tiếp đó tích chọn mô tả tổ chức hoặc nhu cầu của cá nhân cuối cùng là tích vào ô Terms of Service và nhấn continue



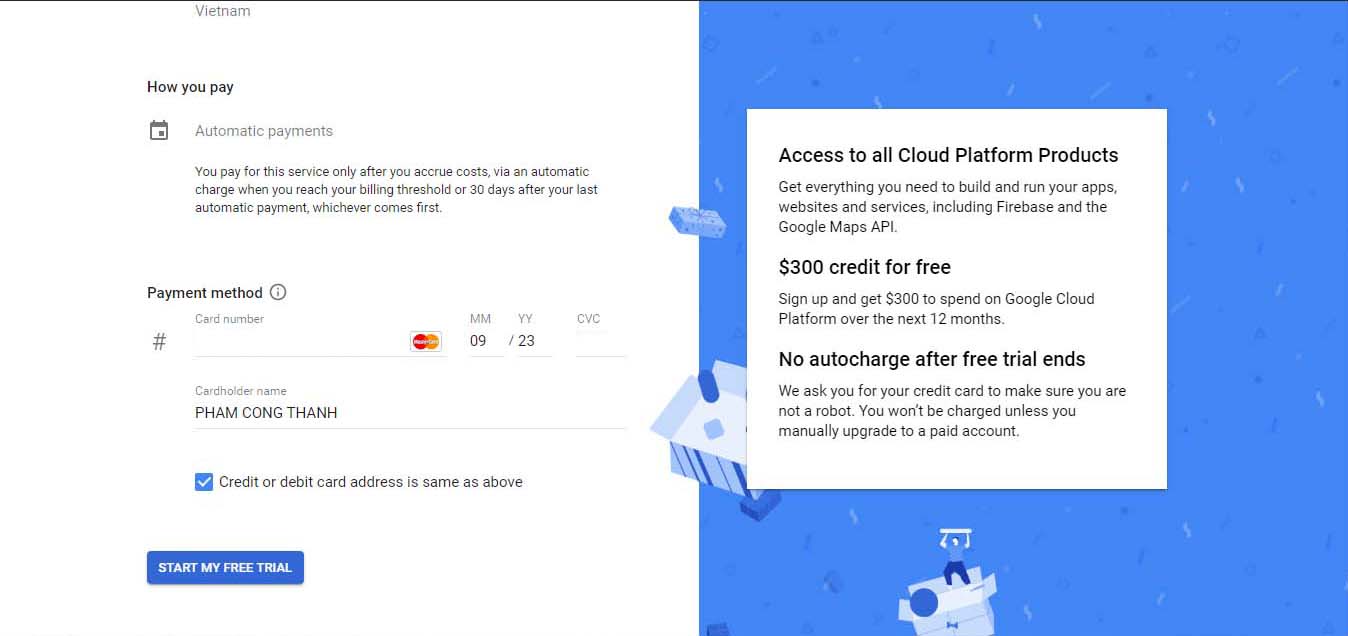
Điền thông tin cá nhân

Nhập số điện thoại đăng ký và Google sẽ gửi mã xác minh để bạn xác nhận số điện thoại đó



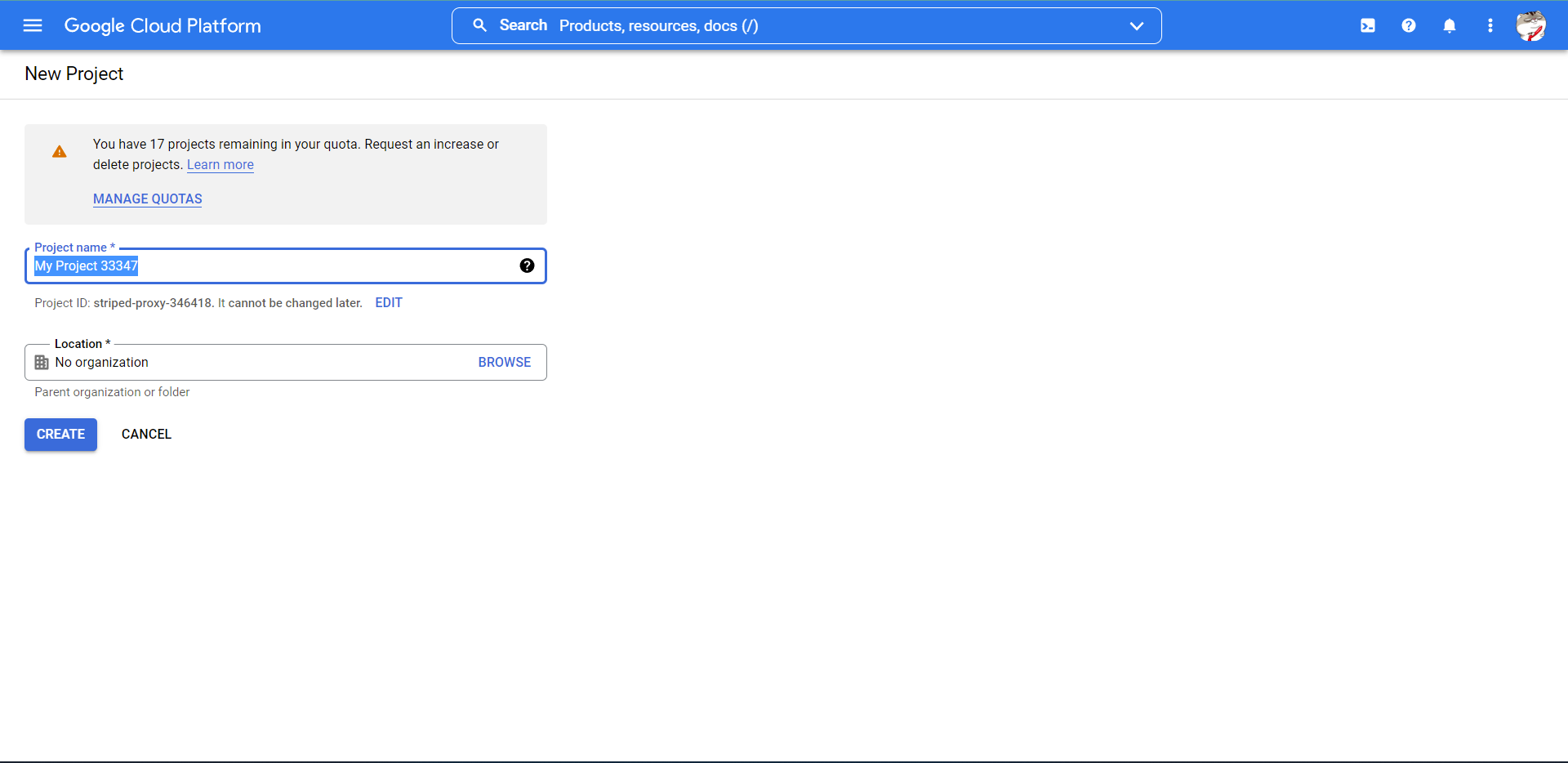
Nhập số điện thoại và xác minh

Bước tiếp theo, chúng ta điền thông tin cá nhân cũng như thông tin thẻ thanh toán quốc tế để có thể hoàn tất đăng ký và có thể sử dụng được các dịch vụ của Google Cloud. Với lần đầu tiên mở tài khoản thì Google - Cloud cho phép sử dụng 300$ lưu lượng miễn phí để sử dụng các dịch vụ trong vòng 3 tháng. Khi sử dụng hết 300$ lưu lượng miễn phí hoặc hết 3 tháng thì các phí dịch vụ sẽ được trừ vào tài khoản thanh toán quốc tế đã đăng ký

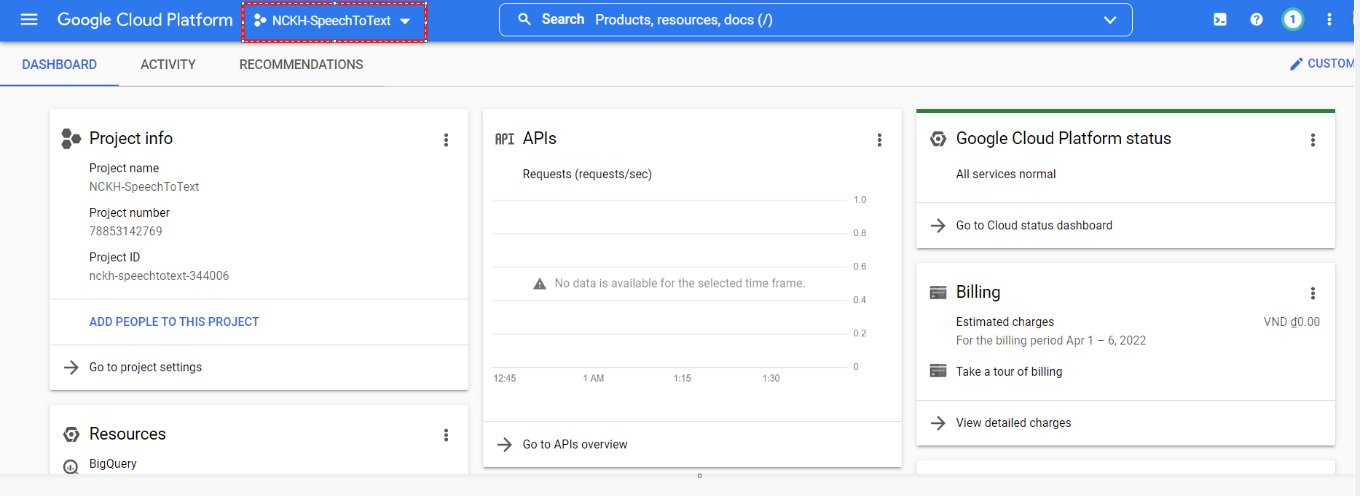


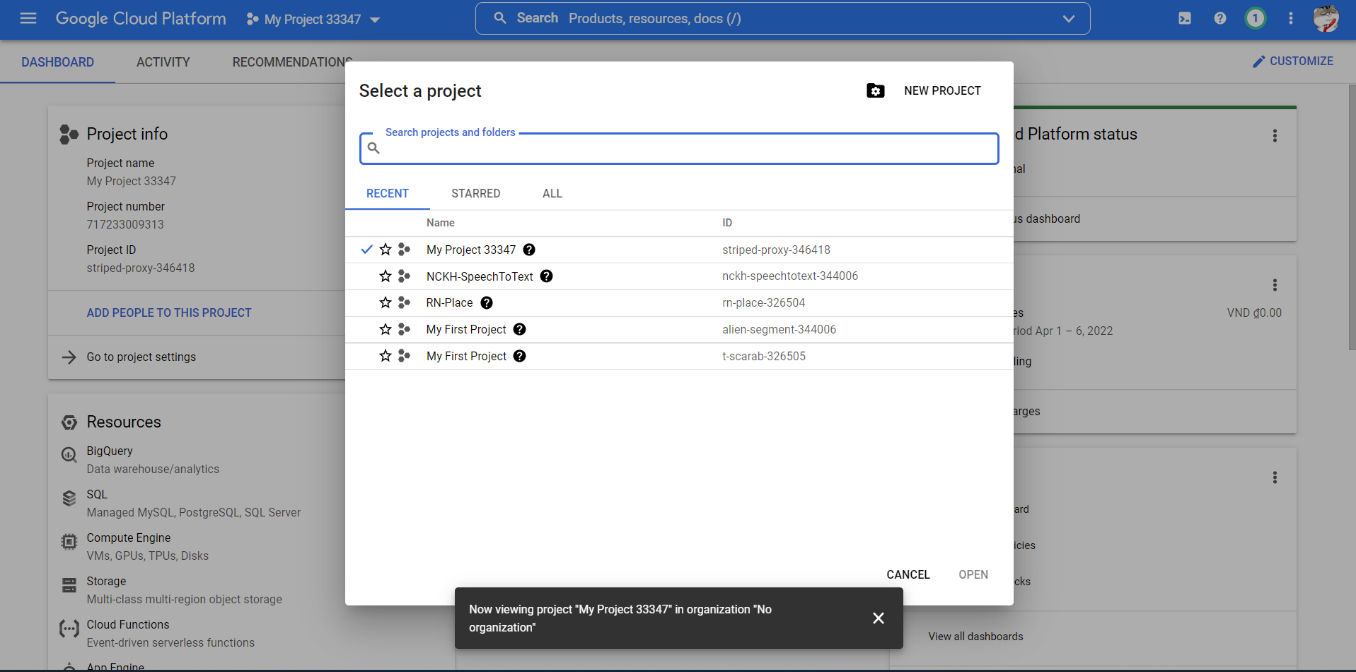
Tiến hành thanh toán qua thẻ master card/ visa

Sau khi hoàn tất quá trình đăng ký Google Cloud thì chúng ta sẽ tiến hành tạo project

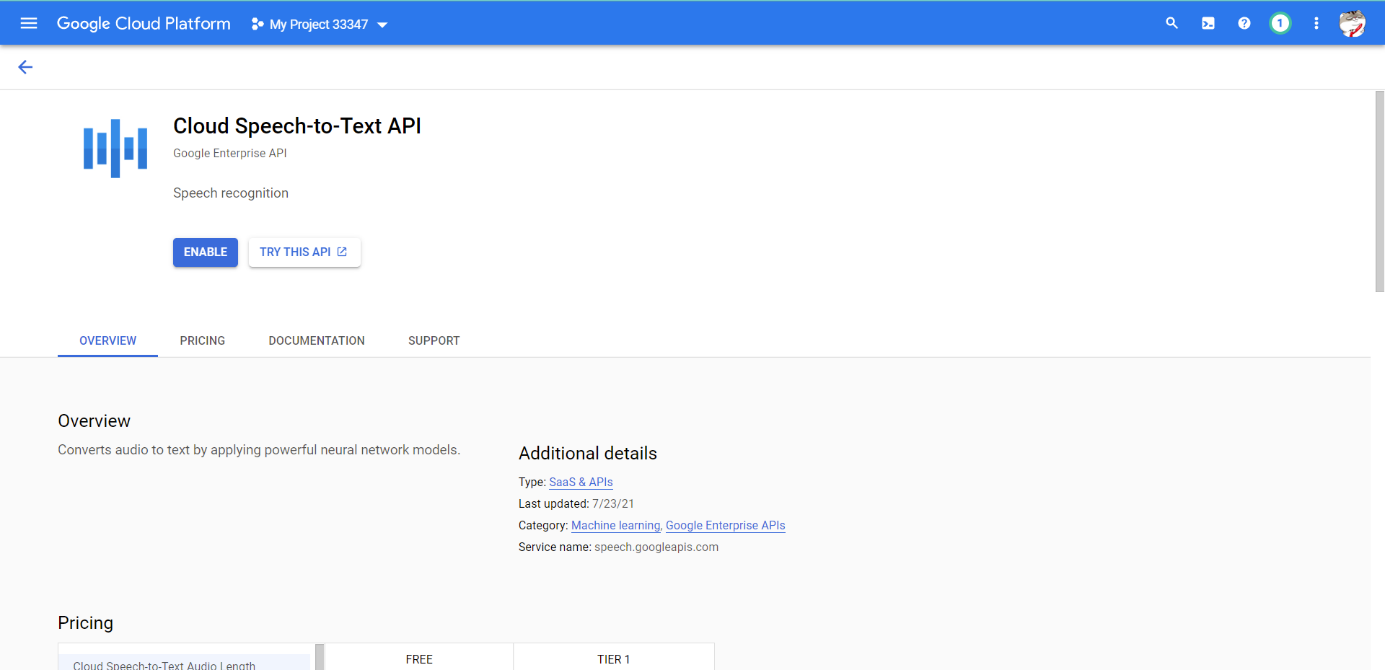


Tạo project

Chọn project vừa tạo

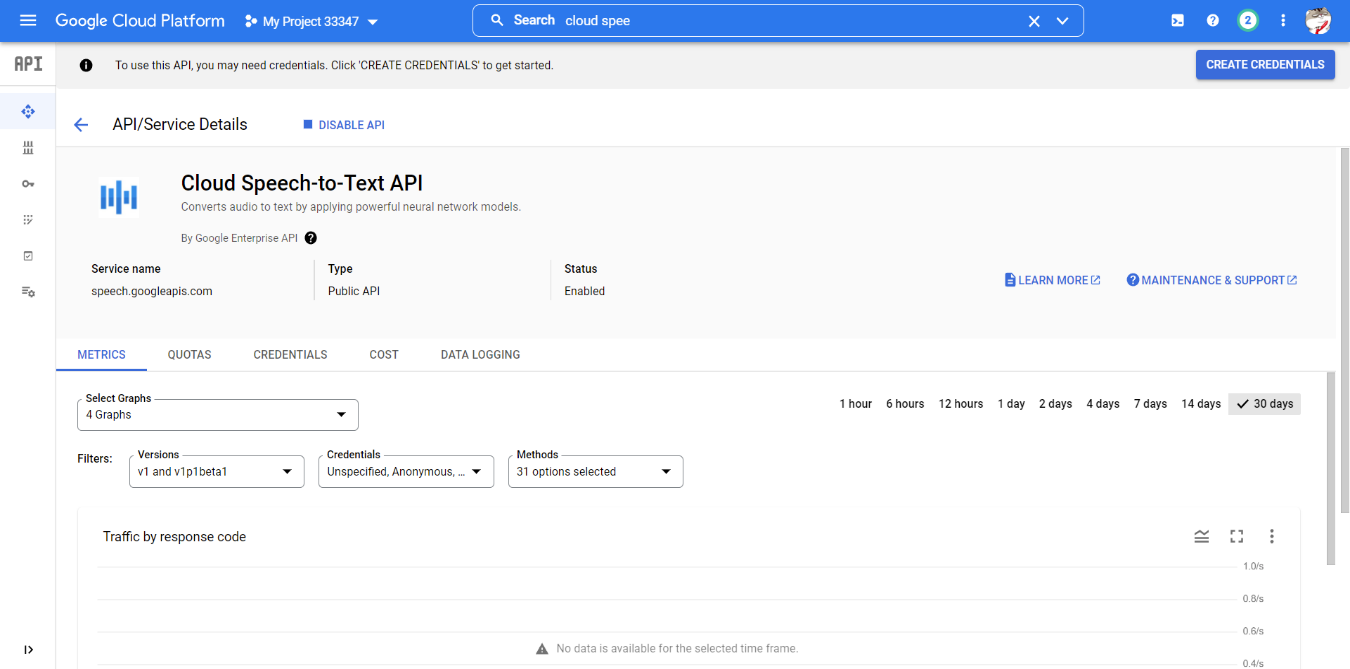
Chọn Project vừa tạo

Chọn Project vừa tạo

Ở thanh tìm kiếm, tìm kiếm Cloud Speech-to-text API

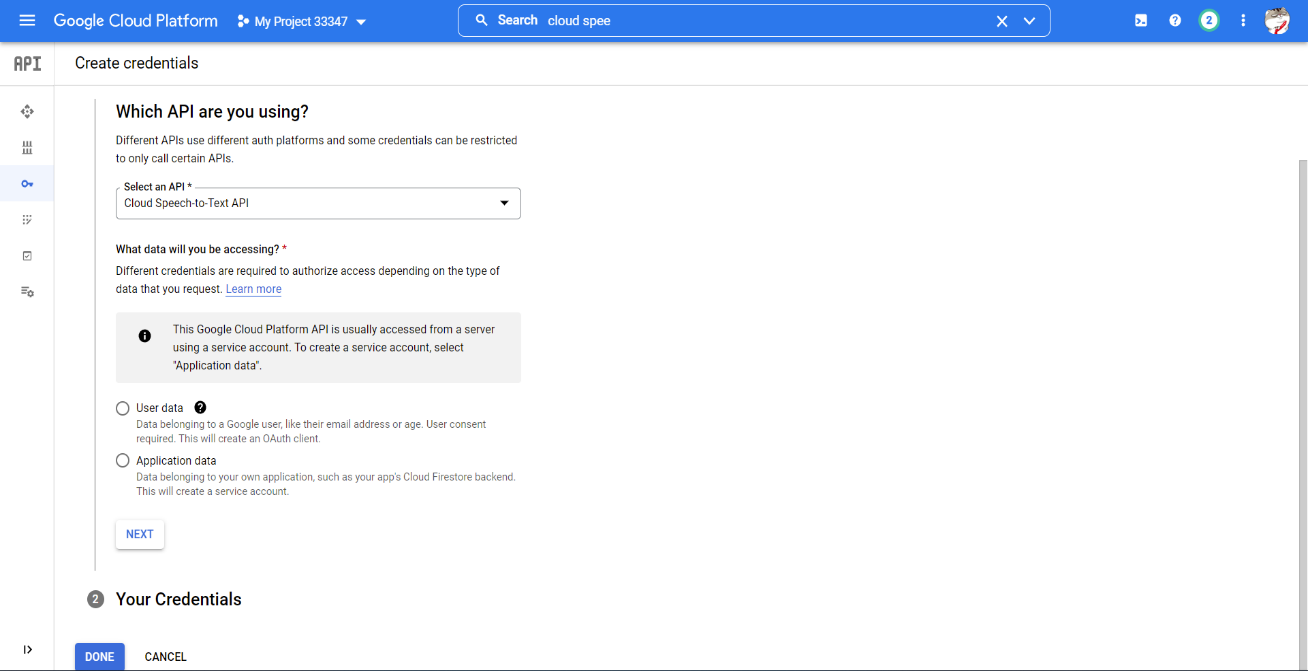
Cloud Speech-To-Text API

Click nút Enable để có thể kích hoạt API, sau đó tạo Credential

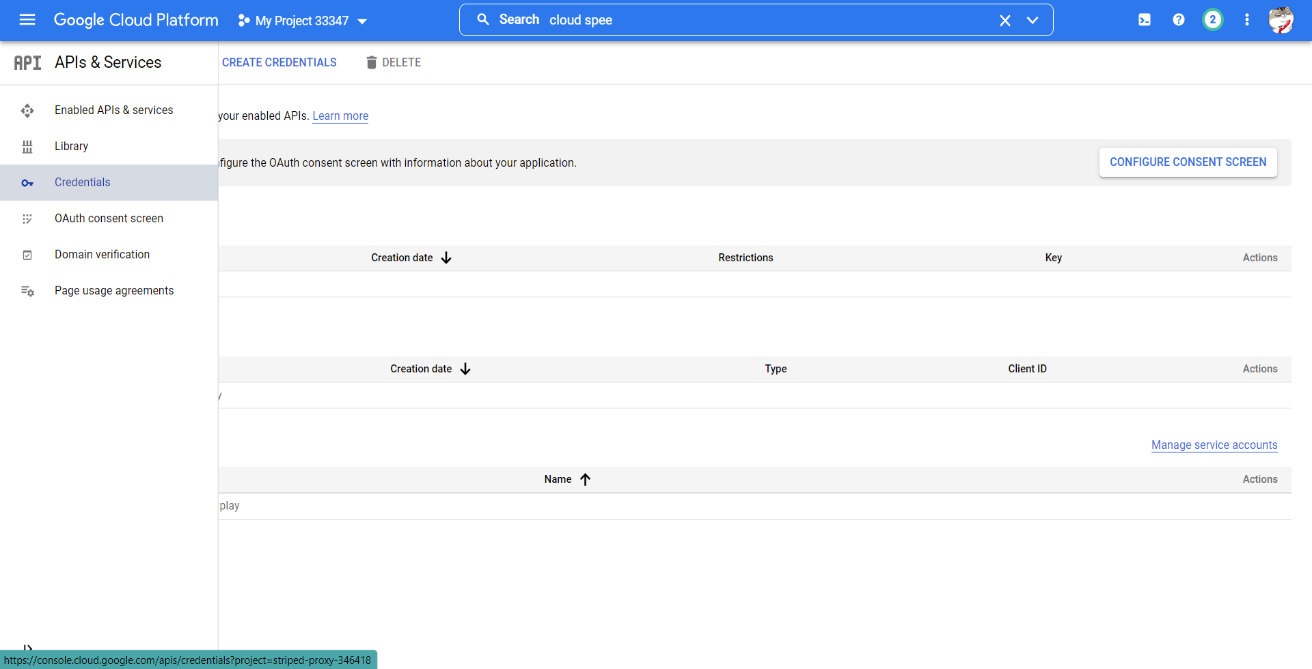


Kích hoạt API và sử dụng

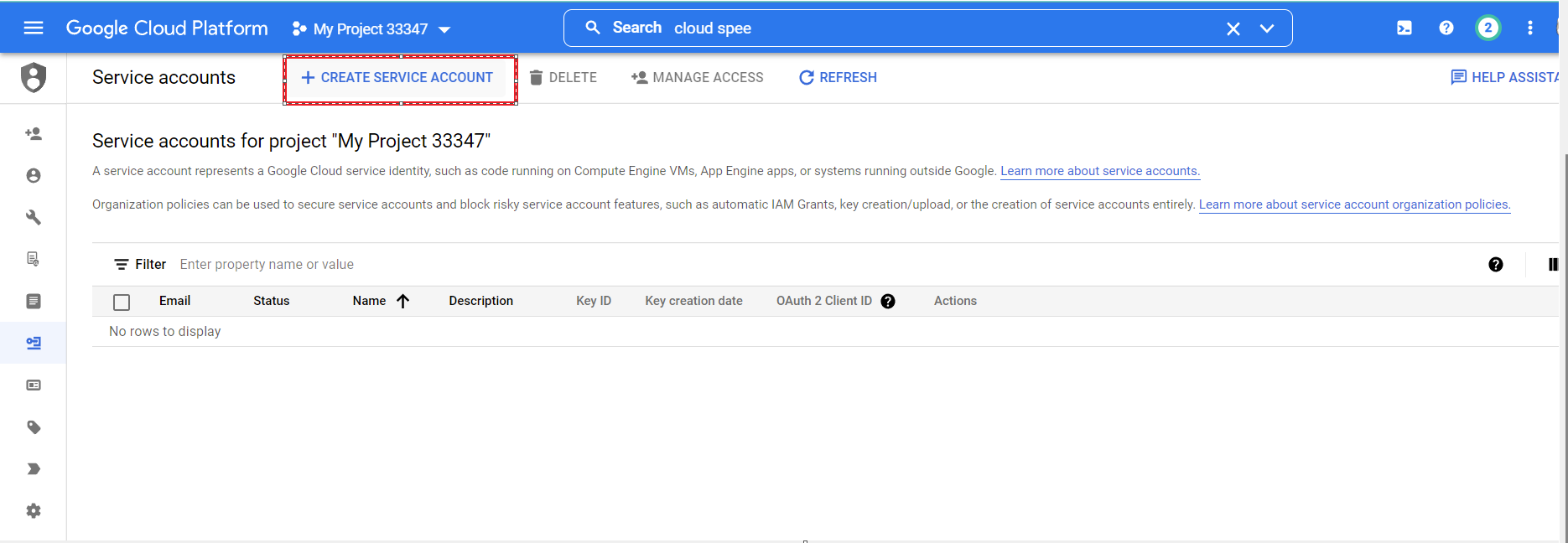
Chọn kiểu truy cập dữ liệu



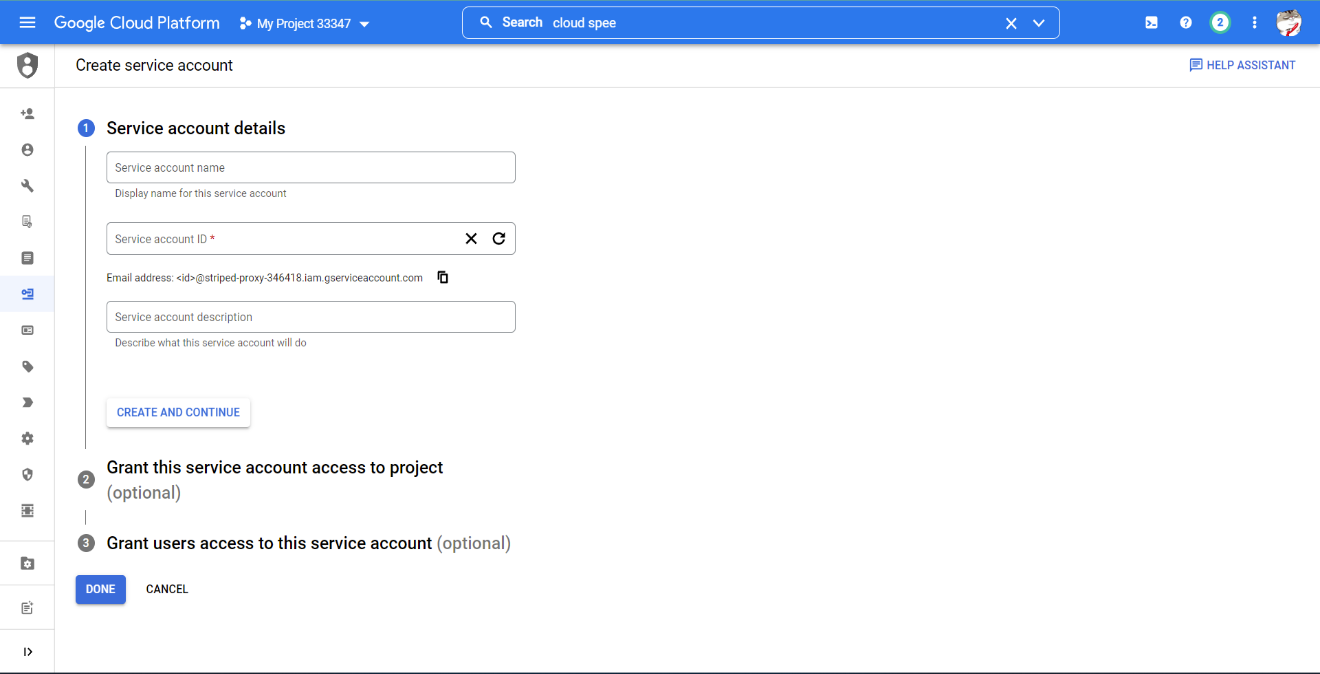
Chọn kiểu dữ liệu

Bước tiếp theo, ta chọn phần Credential và chọn Manage service accounts

Chọn Manage service accounts

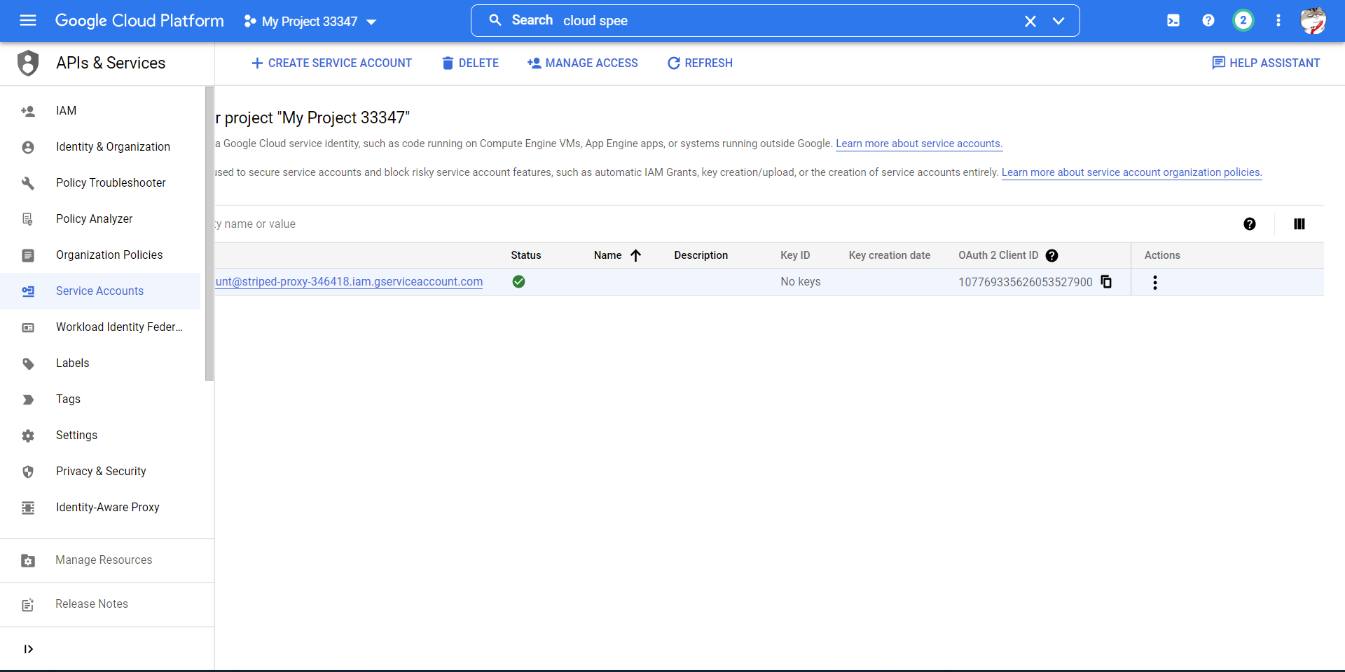
Chọn Create service account

Crearte service account

Nhập các thông tin cần thiết và nhấn Done

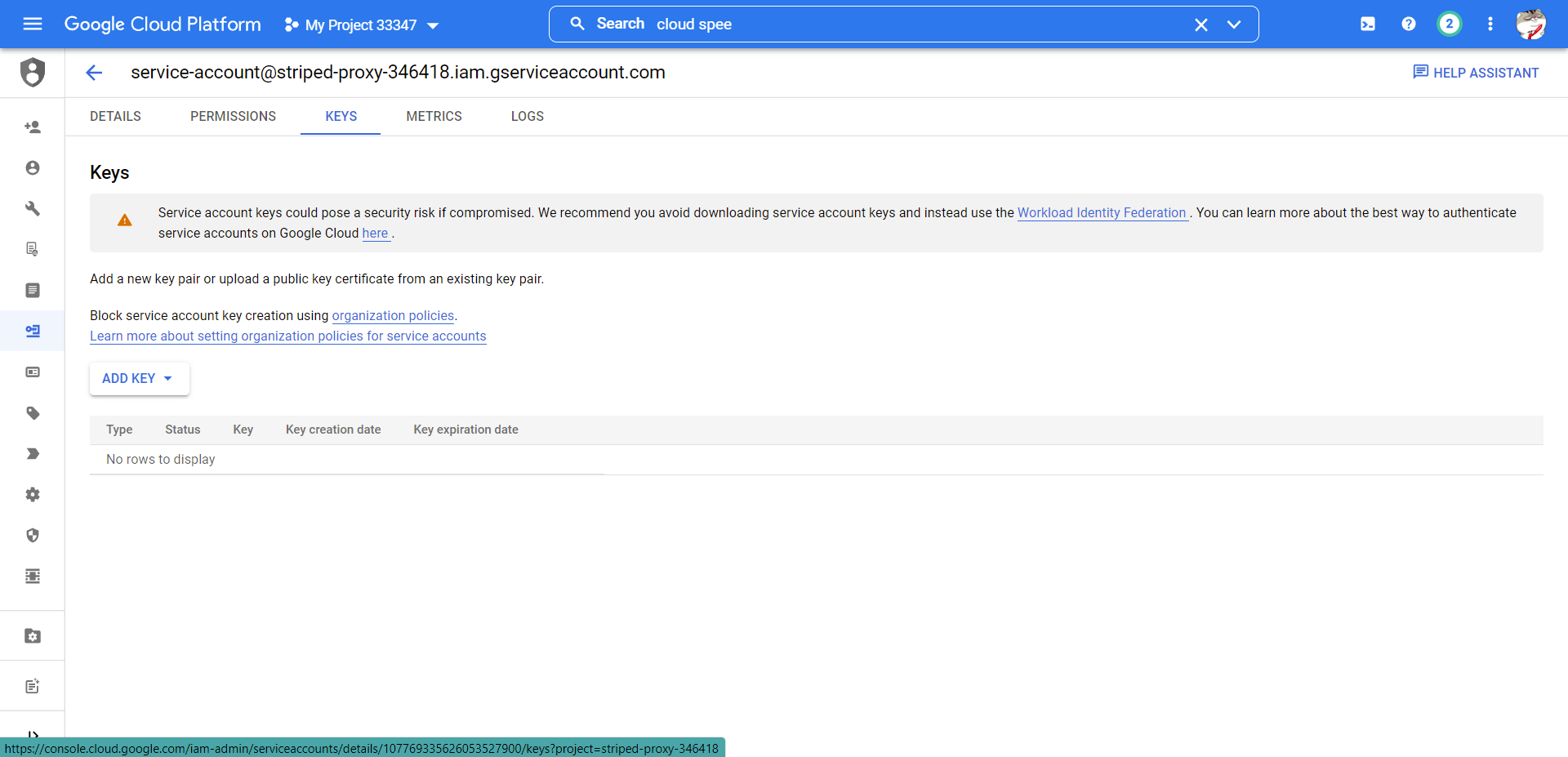
Nhập các thông tin cần thiết

Sau khi tạo xong ta sẽ được tài khoản dịch vụ

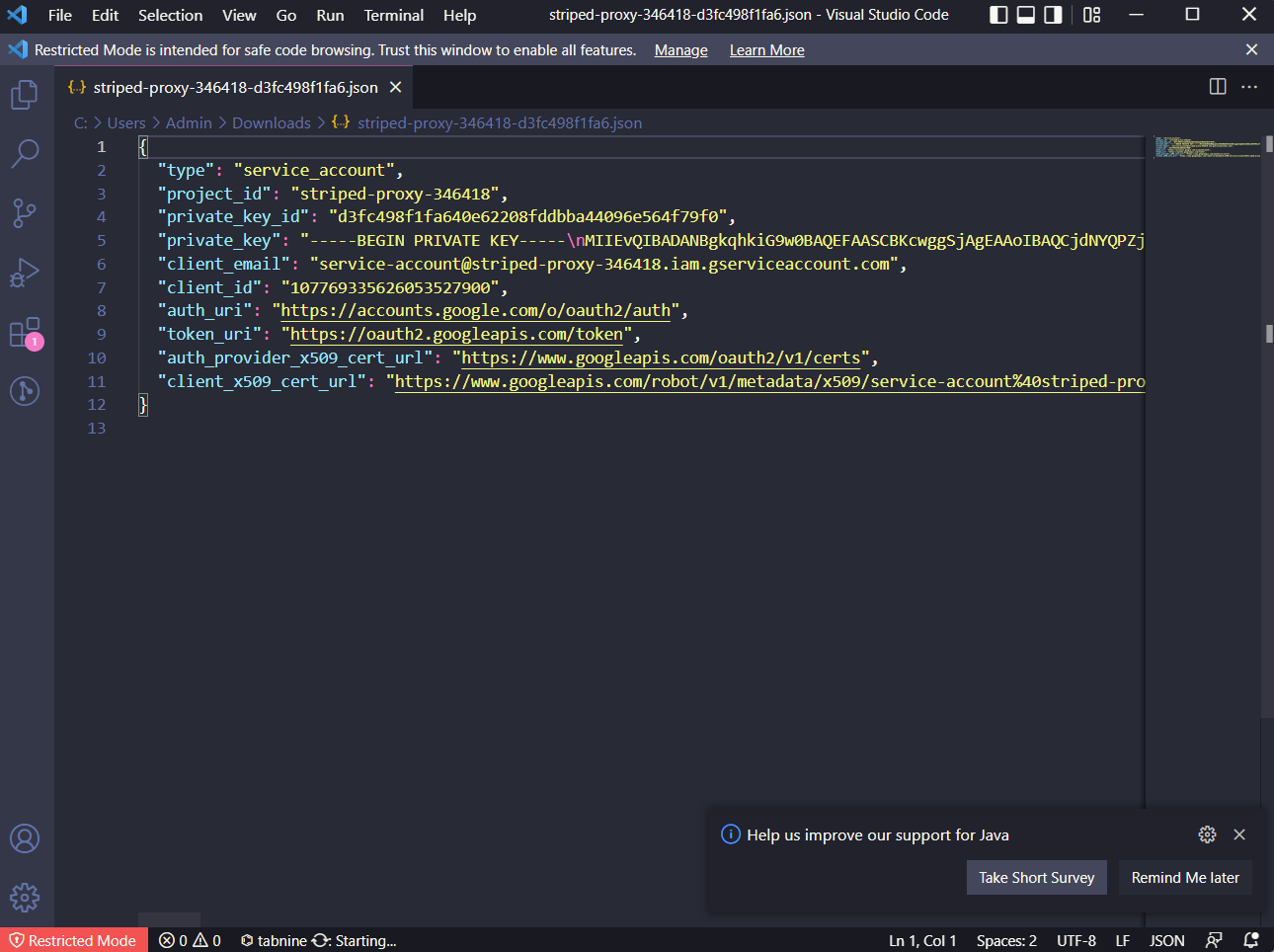


Tài khoản dịch vụ

Để tạo key thì ta nhấn vào tài khoản dịch vụ ta vừa tạo và chọn phần Key sau đó ấn add key chọn create new key và chọn Json



Tải file JSON từ trang chủ

Và khi đó ta sẽ nhận được file Json chứa key

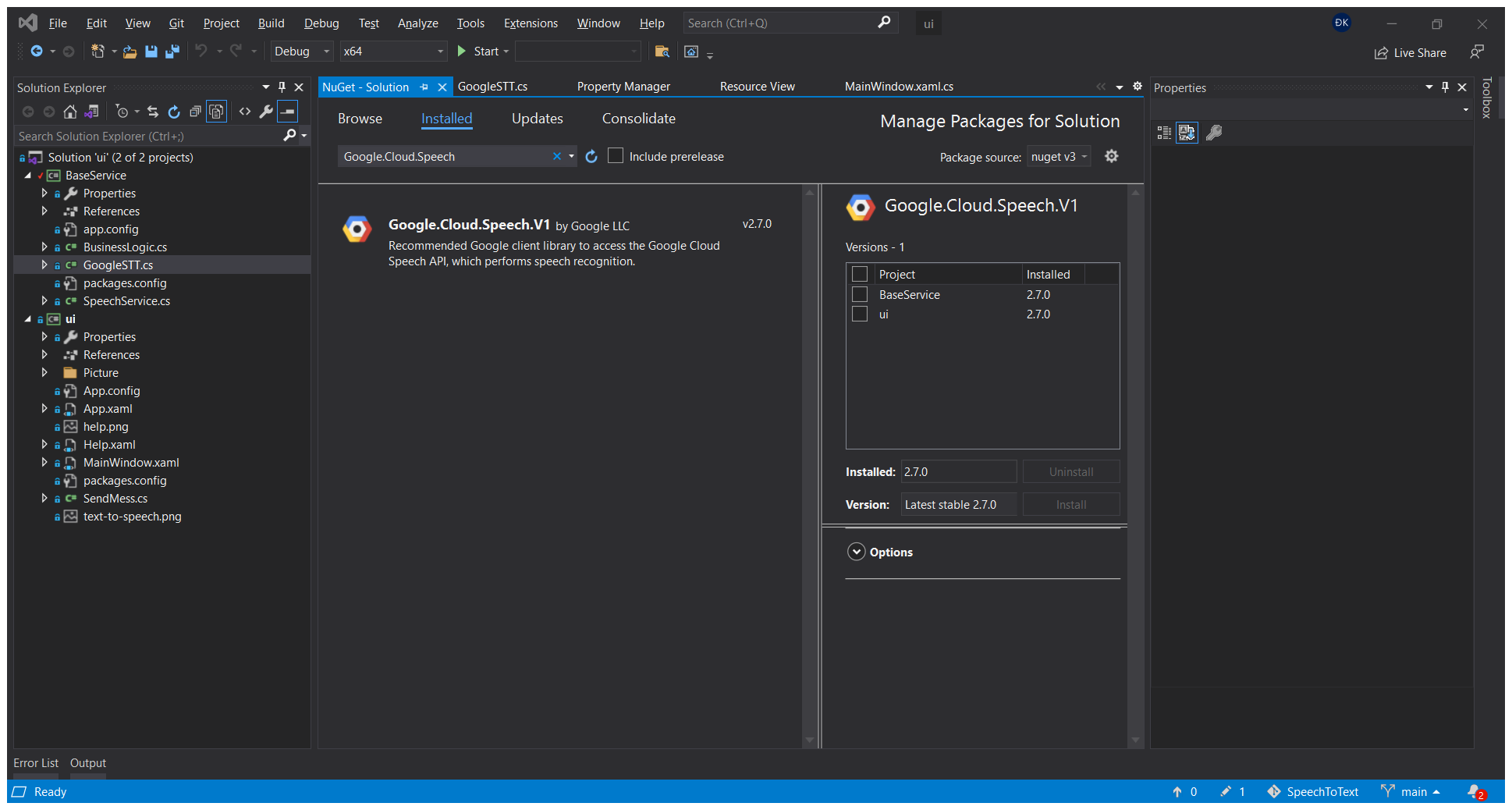
Cấu trúc file Json

Và cuối cùng ta chỉ cần cấu hình project của chúng ta với file json đó thì ta có thể sử dụng được Speech-To-Text api của google cloud.

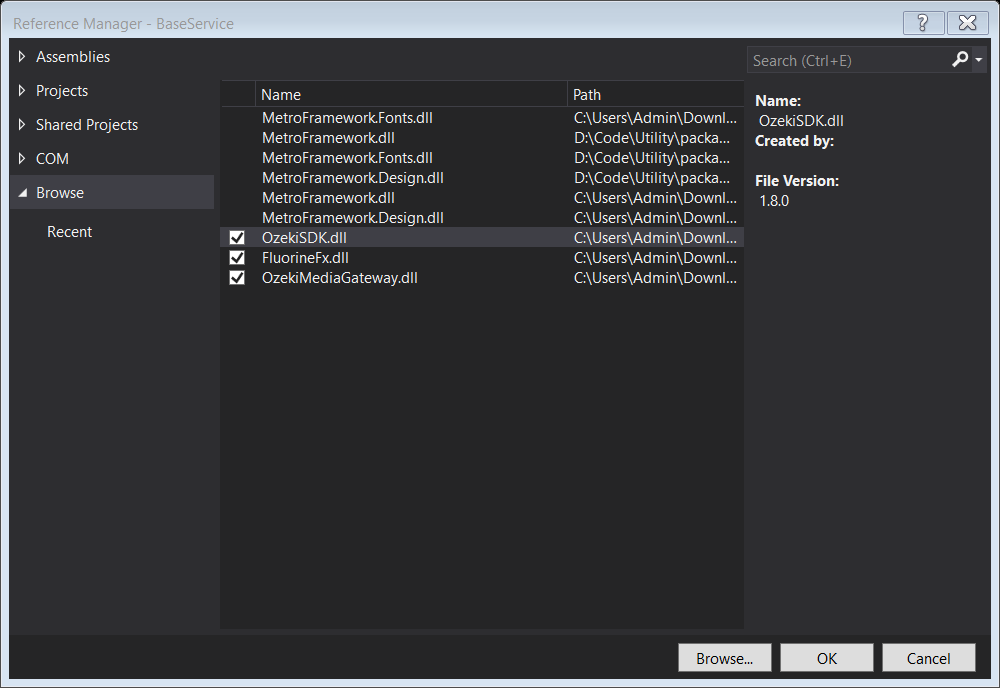
* 1. **Ứng dụng API vào dự án Speech-To-Text**

Sau khi thực hiện đăng ký Speech-To-Text API của Google Cloud, chúng ta ứng dụng API vào dự án Speech-To-Text.

Đầu tiên ta add Nuget “Google.Cloud.Speech.V1” để có thể sử dụng được thư viện hỗ trợ speech-to-text của Google Cloud

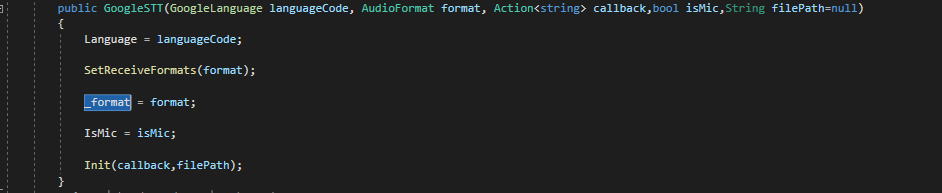


Sau đó ta sẽ tải thư viện Ozeki và thêm 3 file dll vào References để có thể xử lý âm thanh nhận từ micro

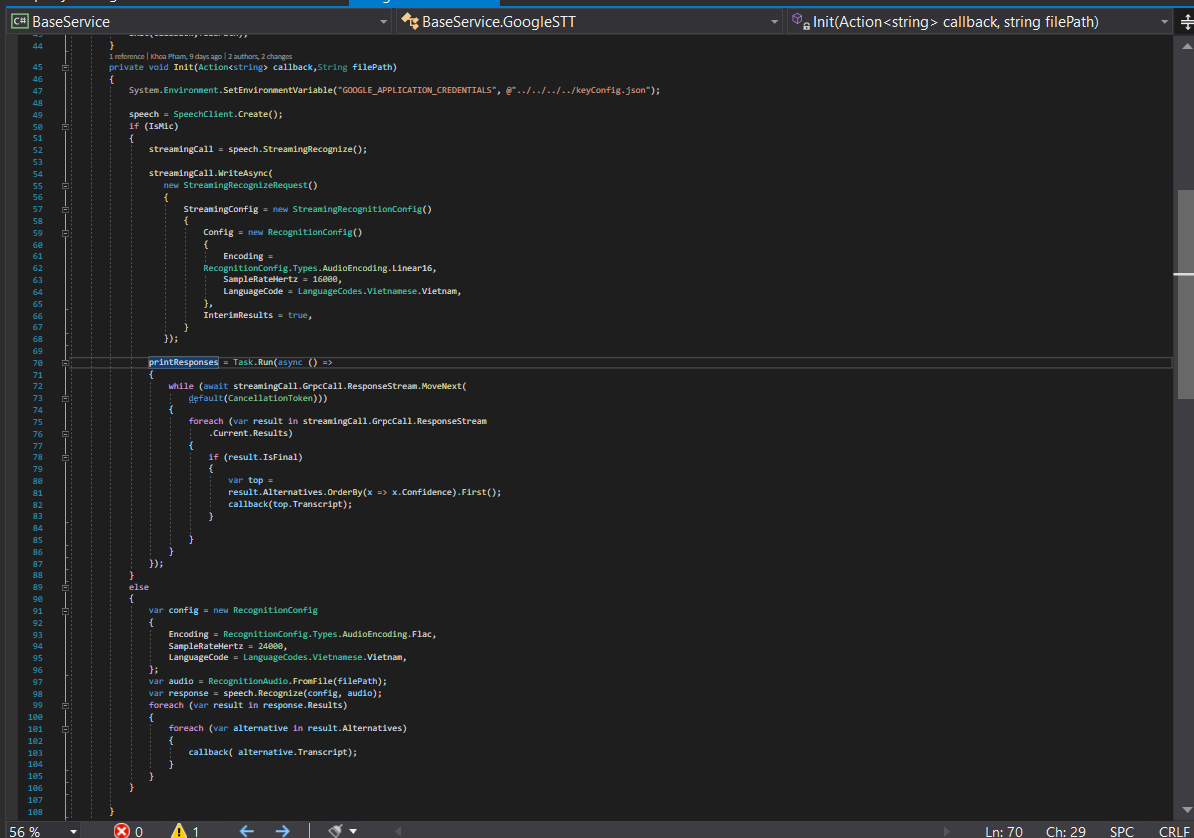


Tiếp đó ta tạo một class để chứa những hàm để xử lý file âm thanh và chuyển nó thành văn bản. Ở đây ta sẽ đặt tên class là GoogleSTT

Hàm khởi tạo của class nhận vào những tham số đó là ngôn ngữ đang dùng, cấu hình âm thanh, một call back, một kiểu dữ liệu dạng bool để xem người dùng chọn nhập âm thanh từ micro hay file và cuối cùng là đường dẫn tới file âm thanh nếu người dùng chọn nhập từ âm thanh (mặc định là null)



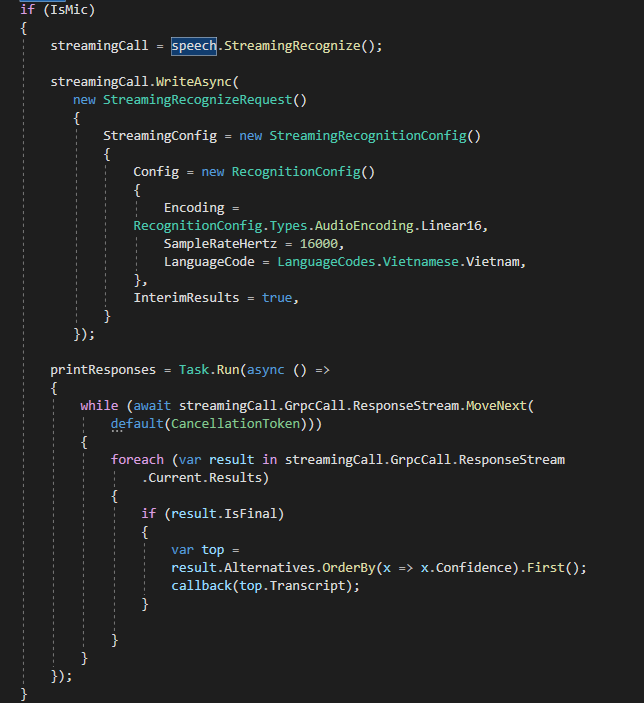
Đầu tiên, ta tạo hàm init với tham số truyền vào là một call back, hàm này tự động thực thi sau khi thực hiện xong việc chuyển giọng nói thành văn bản, tham số thứ 2 là một đường dẫn dẫn đến file chứa âm thanh đầu vào



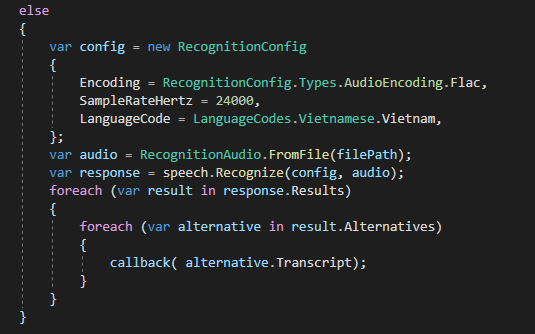
Hàm SetEnvironmentVariable của lớp Enviroment tạo biến môi trường để có thể sử dụng được api của google thông qua file json đã tạo



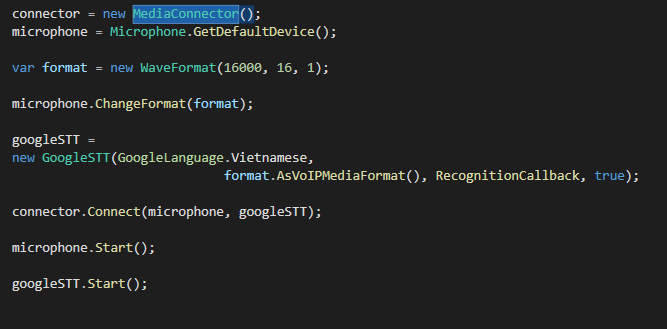
Bước tiếp theo, ta sẽ kiểm tra xem là người dùng sẽ nhập file âm thanh từ micro hay từ file. Nếu người dùng nhập từ micro thì ta sẽ cấu hình Stream Recognition phù hợp với âm thanh từ micro ta chuyền vào



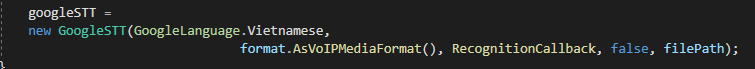
Còn nếu người dùng chọn nhập âm thanh từ file thì ta sẽ cấu hình Recognition phù hợp với file âm thanh

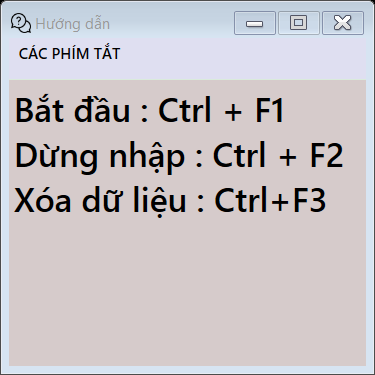


Để có thể nhận được âm thanh từ mic ta phải sử dụng thư viện “Ozeki.Media”. Ta dùng hàm Microphone.GetDefaultDevice để có thể lấy được micro của máy và sẽ cấu hình cho micro đó. Ta sẽ khởi tạo đối tượng của lớp GoogleSTT (với các tham số là ngôn ngữ, cấu hình âm thanh, hàm call back và true vì sử dụng micro để truyền âm thanh) để có thể dùng api chuyển file âm thanh thành văn bản. Cuối cùng ta sẽ dùng hàm Connect để kết nối âm thanh từ micro đến api và gọi hàm Start cho cả microphone và GoogleSTT



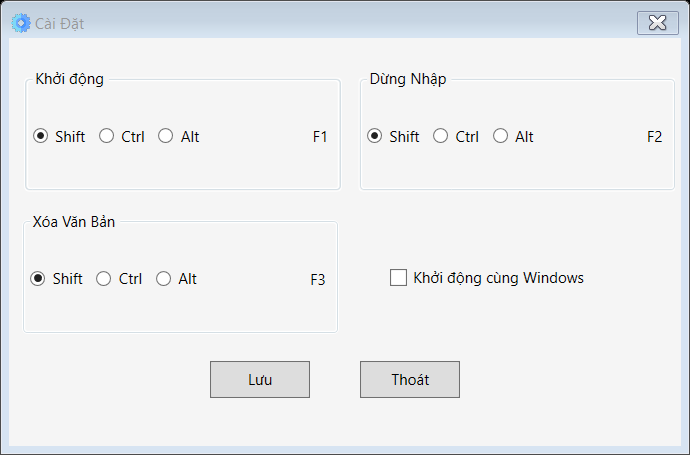
Còn đối với việc nhận âm thanh từ file thì ta sẽ đổi true thành false và có 1 tham số nữa đó là đường dẫn của file âm thanh đó



* 1. **Giao diện ứng dụng Speech-To-Text**
     1. ****Các *trang chính của ứng dụng*

Giao diện chính ứng dụng

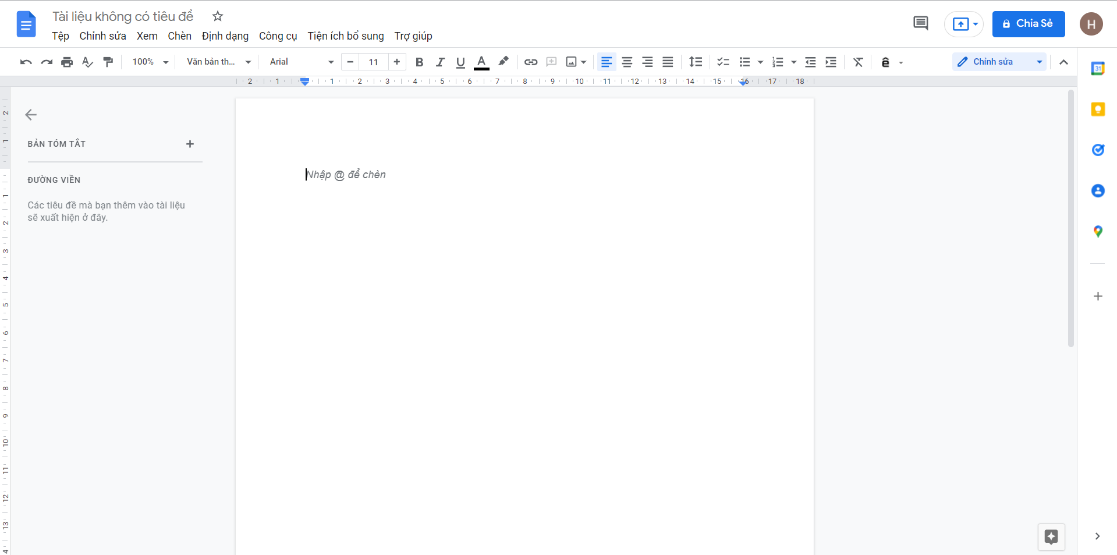
Hướng dẫn sử dụng ứng dụng



Giao diện cài đặt ứng dụng

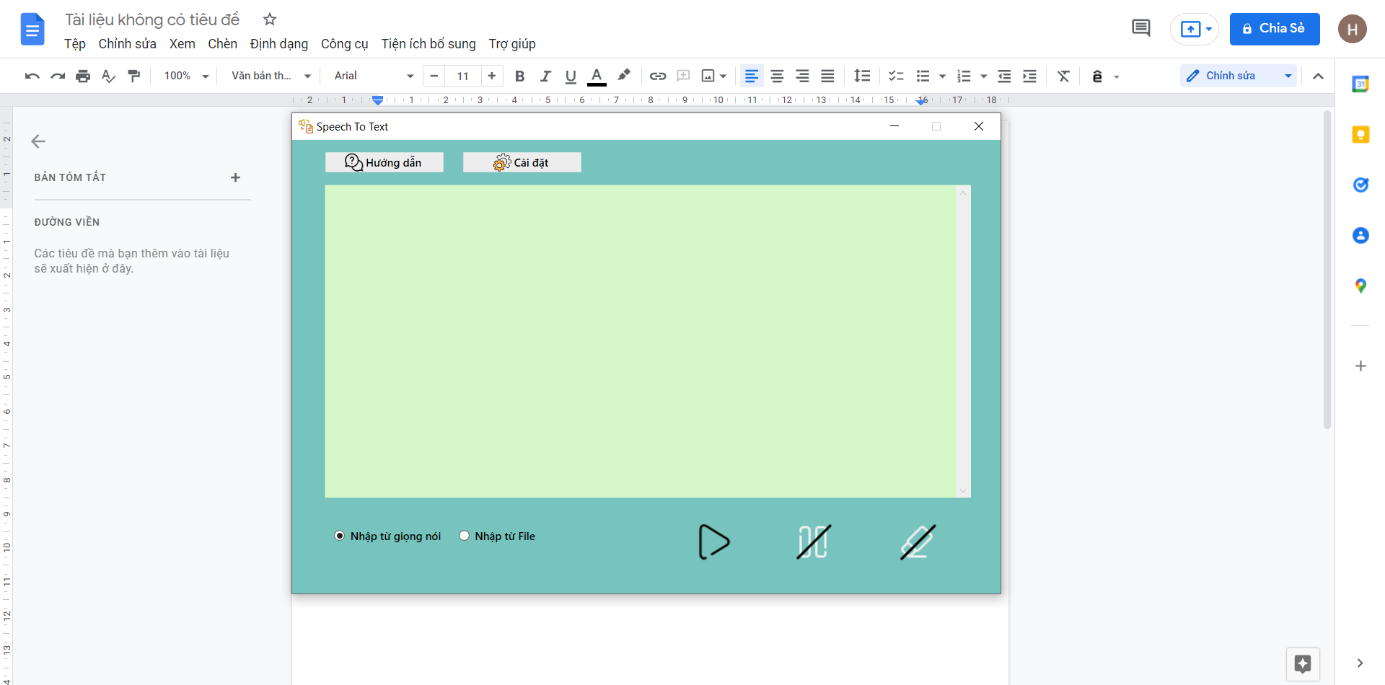
* + 1. *Cách sử dụng ứng dụng Speech-To-Text*

Đầu tiên chúng ta khởi chạy một bộ soạn thảo văn bản

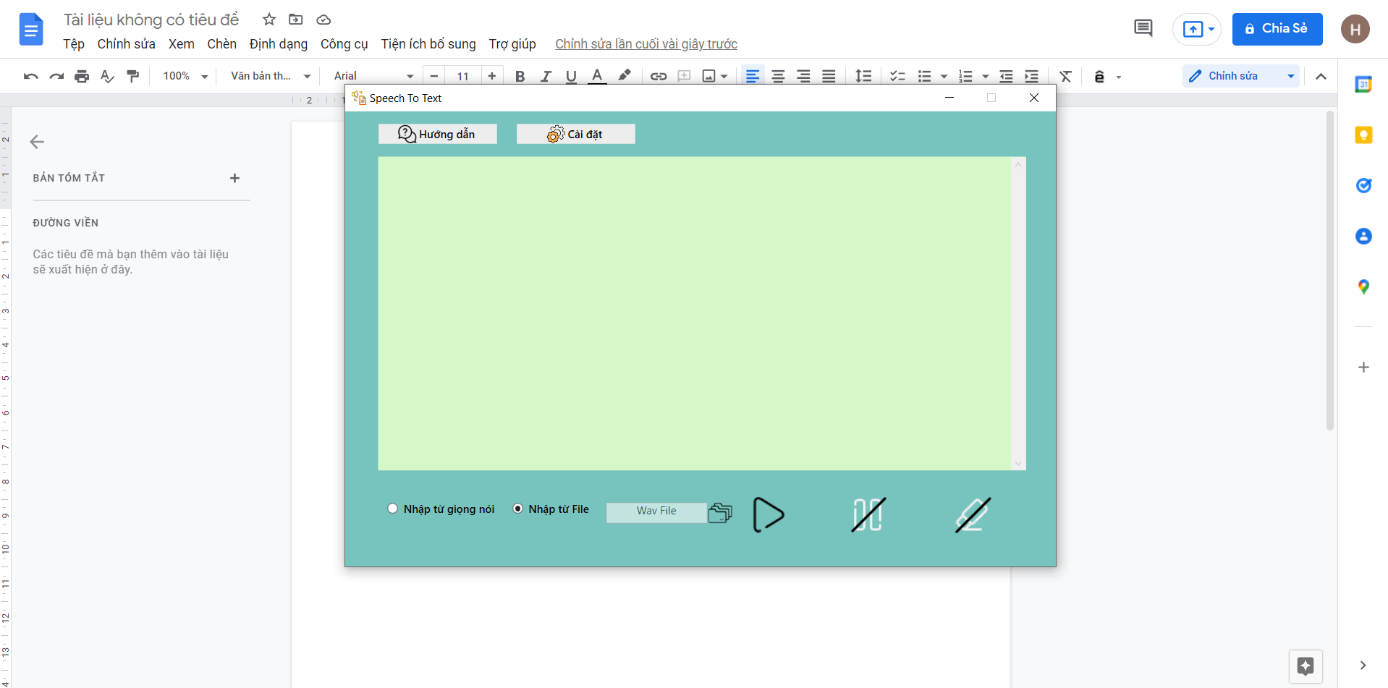


Bắt đầu soạn thảo văn bản

Khởi động ứng dụng để sẵn sàng soạn thảo



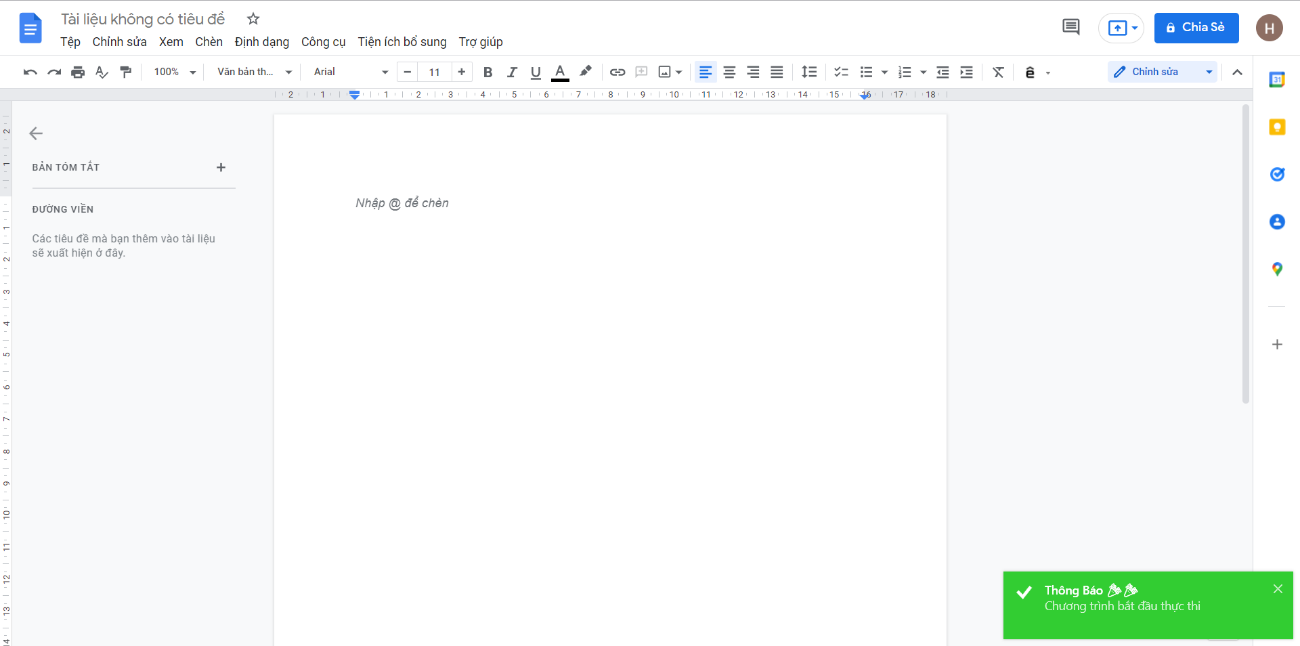
Khởi động ứng dụng

Có 2 chế độ soạn thảo chính: Soạn thảo văn bản bằng giọng nói và soạn thảo văn bản từ file âm thanh (định dạng \*.flac)

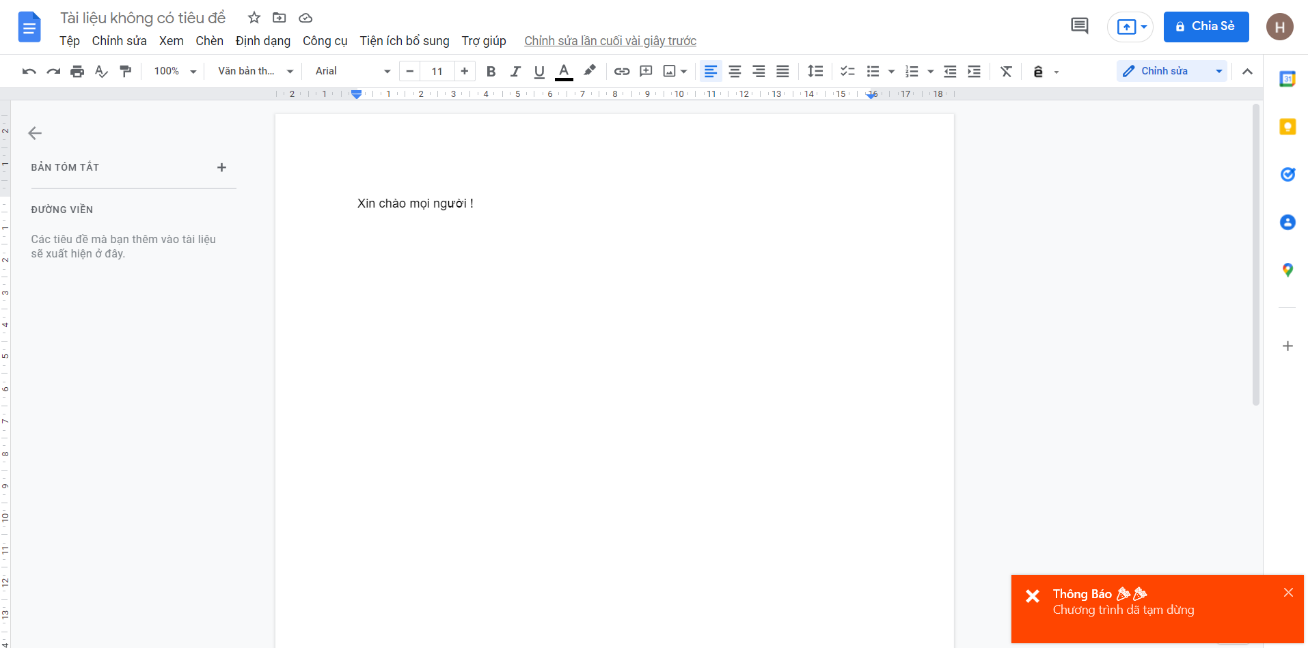
Soạn thảo văn bản từ file âm thanh (\*.flac)

Sau khi chọn xong chế độ soạn thảo văn bản, có thể bắt đầu nhập bằng 2 cách. Thứ nhất là ấn biểu tượng bắt đầu, thứ hai là nhấn phím Hot Key bạn đã cài đặt ở trong phần cài đặt chương trình

Sau khi bắt đầu thì chương trình, thông báo bắt đầu sẽ được hiện ra và chương trình tự động thu gọn lại là một chương trình chạy ẩn, bạn có thể bắt đầu soạn thảo văn bản bằng giọng nói hay file âm thanh



Bắt đầu soạn thảo

Để dừng lại, vui lòng sử dụng phím Hot Key dừng lại đã được cài đặt trong bộ cài đặt chương trình. Thông báo dừng lại sẽ hiện ra khi bạn dừng lại chương trình

Kết thúc soạn thảo

Để hiển thị lại giao diện chương trình, bạn hãy vào góc trái của thanh Taskbar và chọn biểu tượng Speech-To-Text để hiển thị lại giao diện chương trình.

1. **KẾT LUẬN VÀ KIẾN NGHỊ**
   1. **Kết luận**
      1. *Kết quả đạt được:*

Nhóm nghiên cứu đã hoàn thành phiên bản phần mềm đầu tiên của đề tài: Ứng Dụng Soạn Thảo Văn Bản Bằng Giọng Nói (Speech-To-Text). Ứng dụng đã đạt được mục tiêu đề ra là hỗ trợ người dùng soạn thảo văn bản bằng giọng nói tiếng Việt.

Đặc biệt, phần mềm nà có hỗ trợ các khẩu lệnh điều khiển trong quá trình soạn thảo như bỏ dấu cuối câu, dấu chấm, phẩy, xuống dòng và danh sách các khẩu lệnh này cho phép người dùng có thể chỉnh sửa và tùy chọn trong quá trình sử dụng.

Phần mềm hiện hoạt động tốt trên nền tảng hệ điều hành Windows và các phần mềm soạn thảo văn bản và text editor phổ biến.

Đây là một phần mềm có đóng góp không nhỏ giúp hỗ trợ người dùng, đặc biệt là người khiếm thị có thể tìm kiếm thông tin, soạn thảo văn bảng bằng giọng nói và là công cụ hữu ích đối với những đối tượng người dùng khác có như cầu soạn thảo văn bản nhiều như nhà văn, nhà báo, thư ký văn phòng, …. Phần mềm được nhóm nghiên cứu đánh giá có tiềm năng rất lớn khi ứng dụng vào thực tiễn và sớm trở thành công cụ hỗ trợ đắc lực cho những người có nhu cầu nói chung và người khiếm thị nói riêng.

* + 1. *Những vấn đề còn tồn tại:*

Việc phát âm tiếng Việt theo từng vùng miền còn gây ra một số sai sót khi chuyển giọng nói thành văn bản. Nhóm đã nhận thấy được điều này và đang tìm hướng khắc phục để nâng cao chất lượng của phần mềm, giúp nhận dạng giọng nói và chuyển thanh văn bản tốt hơn.

Do thời gian hạn chế và ảnh hưởng của dịch bệnh, hiện tạo giao diện phần mềm chưa thật sự hoàn chỉnh, nhóm sẽ khắc phục và sẽ đưa khẩu lệnh vào việc điều khiển đóng/mở, bật/tắt các chức năng của phần mềm hoàn toàn bằng khẩu lệnh.

* 1. **Kiến nghị và hướng phát triển**

Sau quá trình tổng hợp tài liệu, nghiên cứu và phát triển. Nhóm nghiên cứu đã rút ra được nhiều kinh nghiệm qua đó đúc kết và có những đề xuất những hướng nghiên cứu phát triển tiếp theo như sau:

* Nghiên cứu bổ sung các bộ giọng nói hỗ trợ người khiếm thị tiếp cận sử dụng máy tính.
* Nghiên cứu phát triển hệ thống Speech-To-Text dành riêng cho người Việt.
* Tiếp tục nghiên cứu phát triển các bộ AI thông minh giúp người dùng thao tác xử lý văn bản.
* Tích hợp phần mềm hệ thống điều khiển máy tính bằng giọng nói hỗ trợ người dùng.
* Xử lý các vấn đề Service liên quan đến tham chiếu DLL được sử dụng trong đề tài
* Sử dụng Web Service/ Web API để có thể phát triển phần mềm thành một công cụ đa nền tảng, chạy được trên nhiều hệ điều hành, kể cả các hệ điều hành dành cho SmartPhone.
* Nghiên cứu phát triển phần mềm kết hợp bộ công cụ xử lý văn bản Tiếng Việt thông minh tự động sửa lỗi khi người dùng phát âm sai

1. **TÀI LIỆU THAM KHẢO**
2. Google. (n.d.). *Speech-to-text basics | cloud speech-to-text documentation | google cloud*. Google. Retrieved April 7, 2022, from https://cloud.google.com/speech-to-text/docs/basics
3. Fagerberg, J. (2017). *Asp.Net Core 2.0 Mvc & Razor Pages for Beginners: Tactical Guidebook*. CreateSpace Independent Publishing Platform.
4. LAWRENCE, H. O. N. G. V. I. C. T. O. R. I. A. F. I. S. C. H. E. R. (2017). *Visual Basic.Net: An introduction to computer programming*. KENDALL HUNT.
5. Moroney, L. (2008). *Beginning web development, Silverlight, and ASP.NET ajax: From novice to professional*. Amazon. Retrieved April 7, 2022, from [https://www.amazon.com/Development-Windows-Presentation-Foundation-Information-ebook/dp/B08ZG1TMR7](https://www.amazon.com/Development-Windows-Presentation-Foundation-Information-ebook/dp/B08ZG1TMR7%20)
6. *Windows Presentation Foundation: WPF & .NET*. Visual Studio. (2021, December 10). Retrieved April 7, 2022, from https://visualstudio.microsoft.com/vs/features/wpf/
7. Admin. (2019, March 6). *Lợi ích Của việc chuyển đổi Giọng Nói Thành Văn Bản*. https://upsharework.com. Retrieved April 7, 2022, from [https://www.upsharework.com/vi/blogs/the-benefits-of-speech-to-text-speech2text](https://www.upsharework.com/vi/blogs/the-benefits-of-speech-to-text-speech2text%20)
8. *Speech to text API*. Ozeki SMS Server. (n.d.). Retrieved April 7, 2022, from [https://ozeki.hu/p\_5919-speech-to-text-api.html](https://ozeki.hu/p_5919-speech-to-text-api.html%20)
9. Shows how to use speechtotext google API. (n.d.). Retrieved April 7, 2022, from [https://voip-sip-sdk.com/p\_7604-shows-how-to-use-speechtotext-google-api.html](https://voip-sip-sdk.com/p_7604-shows-how-to-use-speechtotext-google-api.html%20)
10. *ĐÁNH Giá Các Hệ Thống Nhận Dạng giọng Nói Tiếng Việt (vais ...* (n.d.). Retrieved April 7, 2022, from [https://www.researchgate.net/publication/352567298\_DANH\_GIA\_CAC\_HE\_THONG\_NHAN\_DANG\_GIONG\_NOI\_TIENG\_VIET\_VAIS\_VIETTEL\_ZALO\_FPT\_VA\_GOOGLE\_TRONG\_BAN\_TIN-Tap\_chi\_khoa\_hoc\_giao\_duc\_ky\_thuat](https://www.researchgate.net/publication/352567298_DANH_GIA_CAC_HE_THONG_NHAN_DANG_GIONG_NOI_TIENG_VIET_VAIS_VIETTEL_ZALO_FPT_VA_GOOGLE_TRONG_BAN_TIN-Tap_chi_khoa_hoc_giao_duc_ky_thuat%20)
11. *How does speech recognition technology work?* Summa Linguae. (2022, March 7). Retrieved April 7, 2022, from [https://summalinguae.com/language-technology/how-does-speech-recognition-technology-work/](https://summalinguae.com/language-technology/how-does-speech-recognition-technology-work/%20)