TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO NHẬP MÔN TRÍ TUỆ NHÂN TẠO**

**ĐỀ TÀI:**

**LẬP TRÌNH NHẬN DIỆN GIỌNG NÓI**

|  |  |
| --- | --- |
|  |  |
| |  |  |  | | --- | --- | --- | | **Sinh viên thực hiện** | **: BÙI QUANG ĐIỆP**  **NGUYỄN DUY TRƯỜNG**  **HOÀNG MINH PHƯƠNG** | | | **Giảng viên hướng dẫn** | **: PHAN THỊ KIM DUNG** | | **Ngành** | **: CÔNG NGHỆ THÔNG TIN** | | **Chuyên ngành** | **: CÔNG NGHỆ PHẦN MỀM** | | **Lớp** | **: D14CNPM3** | | **Khóa** | **: 2019-2023** | |  |
| **Hà Nội ,tháng 12 năm 2021** |  |

# MỤC LỤC

[MỤC LỤC 1](#_Toc91279247)

[LỜI MỞ ĐẦU 3](#_Toc91279248)

[I, GIỚI THIỆU BÀI TOÁN 4](#_Toc91279249)

[1. Bối cảnh lý do thực hiện đề tài 4](#_Toc91279250)

[2. Mục đích của đề tài 4](#_Toc91279251)

[3. Phương pháp thực thiện đề tài 4](#_Toc91279252)

[II, CƠ SỞ LÝ THUYẾT 4](#_Toc91279253)

[1. Trợ lý ảo 4](#_Toc91279254)

[2. Định nghĩa API 6](#_Toc91279255)

[III, GIẢI QUYẾT BÀI TOÁN 9](#_Toc91279256)

[1. Sử dụng Web Speech API 9](#_Toc91279257)

[IV, Cài đặt và demo chương trình 10](#_Toc91279258)

[1. Cài đặt 10](#_Toc91279259)

[2. Demo chương trình 10](#_Toc91279260)

[V, KẾT LUẬN ĐÁNH GIÁ 11](#_Toc91279261)

[1. Hướng phát triển của đề tài 11](#_Toc91279262)

[2. Đánh giá 11](#_Toc91279263)

# LỜI MỞ ĐẦU

Ngày nay, trí tuệ nhân tạo đang ngày càng phát triển mạnh mẽ. Các hãng lớn như Apple, Microsoft, Google đều đưa ra các công nghệ tương tác trực tiếp với người dùng. Trí tuệ nhân tạo đang ngày càng được nâng cấp hoàn thiện giúp người dùng dễ dàng tương tác, dễ sử dụng và giảm quá trình thực hiện. Trí tuệ nhân tạo đang là một lĩnh vực mới mẻ và được sự quan tâm rất lớn từ các hãng công nghệ hàng đầu. Với công nghệ đang ngày càng được áp dụng trong đời sống giúp con người làm việc hiệu quả hơn tiết kiệm thời gian và sức lực, trí tuệ nhân tạo như một hệ thống được xây dựng để phục vụ cho điều đó. Hệ thống trợ lý ảo là một hệ thống giúp con người giao tiếp với máy thực hiện các yêu cầu ý muốn của người dùng. Hiện nay trên thế giới hệ thống trợ lý ảo đang được nghiên cứu và phát triển mạnh mẽ ở các hãng công nghệ hàng đầu thế giới. Với mong muốn hiểu sâu về trí tuệ nhân tạo, em quyết định chọn đề tài “Lập trình nhận diện giọng nói” làm đề tài môn trí tuệ nhân tạo.

# I, GIỚI THIỆU BÀI TOÁN

**1. Bối cảnh lý do thực hiện đề tài**

Hiện nay, trên thế giới việc sử dụng các phần mềm về âm thanh không còn là một vấn đề quá mới mẻ. Tuy nhiên mỗi quốc gia đều có một đặc thù ngôn ngữ khác nhau nên không có các quy tắc, các luật chung trong việc xử lý từ. Vì thế chúng em chọn và thực hiện lập trình nhận diện giọng nói.

**2. Mục đích của đề tài**

Lập trình xây dựng website nhận diện giọng nói.

**3. Phương pháp thực thiện đề tài**

Sử dựng ngôn ngữ lập trình Javascript, các thư viện hỗ trợ nhận dạng giọng nói API.

# II, CƠ SỞ LÝ THUYẾT

**1. Trợ lý ảo**

Trợ lý ảo là phần mềm được xây dựng dựa trên nền tảng trí thông minh nhân tạo (AI). Nó được các các hãng công nghệ tích hợp sâu vào trong hệ điều hành với mục đích chính là hỗ trợ người dùng thiết bị thực hiện các thao tác dễ dàng hơn thông qua giọng nói của chính họ. Hiện nay có những trợ lý ảo cực kì nổi tiếng là Google Assistant, Siri của Apple, Bixby của Samsung.

**Khả năng nhận diện giọng nói**

Bằng cách sử dụng trí thông minh nhân tạo, những trợ lý ảo ngày nay có thể nhận diện giọng nói rất chính xác. Ngoài khả năng nhận dạng giọng nói bằng tiếng Anh, những trợ lý ảo đã bắt đầu có thể nhận dạng nhiều thứ tiếng khác.

Bằng cách sử dụng trí thông minh nhân tạo, những trợ lý ảo ngày nay có thể nhận diện giọng nói rất chính xác, từ đó cho phép người dùng có thể tương tác với máy tính bằng giọng nói rất dễ dàng mà không cần chạm tay vào máy.

Ví dụ, trợ lý ảo Siri của Apple sẽ bắt đầu khởi chạy ngay khi bạn nói “Hey Siri” thay vì phải nhấn giữ nút Home trên iPhone, iPad, sau đó có thể tiếp tục ra lệnh cho Siri thực hiện những điều bạn muốn.

**Khả năng xử lý ngôn ngữ**

Một khả năng khác cũng mạnh mẽ không kém so với nhận diện giọng nói chính là khả năng xử lý ngôn ngữ. Ngoài việc nhận diện giọng nói để thực hiện tác vụ mà người dùng yêu cầu, trợ lý ảo còn có thể xử lý ngôn ngữ và đưa ra những câu trả lời phản hồi người dùng để tạo ra sự tương tác chân thực như người với người. Công nghệ xử lý ngôn ngữ của các trợ lý ảo hiện nay chủ yếu dựa trên cơ sở dữ liệu khổng lồ bao gồm rất nhiều câu nói theo cả ngôn ngữ chung lẫn ngôn ngữ địa phương. Khi đó, hệ thống sẽ xử lý và chọn ra câu trả lời chính xác nhất với ngữ cảnh.

Ngoài việc nhận diện giọng nói để thực hiện tác vụ mà người dùng yêu cầu, trợ lý ảo còn có thể xử lý ngôn ngữ và đưa ra những câu trả lời phản hồi người dùng, tạo nên sự tương tác giữa người dùng và thiết bị giống hệt cuộc nói chuyện giữa người với người.

**Khả năng tìm kiếm thông tin**

Khả năng tìm kiếm là một trong nhiều ưu điểm nổi bật của trợ lý ảo. Các công cụ tìm kiếm như Bing, Google Search, thư viện bách khoa toàn thư Wikipedia hay Wolfram Alpha là những bộ máy chứa một lượng thông tin vô cùng lớn để các trợ lý ảo có thể thực hiện thao tác tìm kiếm. Với những kho dữ liệu khổng lồ như vậy, rõ ràng những trợ lý ảo này sẽ có khả năng tìm kiếm cực kỳ mạnh mẽ cả về tốc độ lẫn sự chính xác trong kết quả tìm kiếm. Mỗi trợ lý ảo lại có cách tìm kiếm khác nhau, chẳng hạn như Siri sẽ liên tục đưa ra các câu hỏi cho đến khi bạn tìm được kết quả như ý trong khi Google Now sẽ dùng công nghệ Knowledge Graph để tìm ra kết quả chính xác nhất bằng cách phân tích ý muốn của bạn thông qua các từ khoá tìm kiểm.

Ngay khi đặt câu hỏi nền tảng tìm kiếm của các trợ lý ảo là gì, bạn sẽ phải tỏ ra bất ngờ khi “đứng sau” trợ giúp cho tính năng tìm kiếm của các trợ lý ảo chính là những công cụ đình đám như Bing, Google Search, thư viện bách khoa toàn thư Wikipedia hay Wolfram Alpha. Với những kho dữ liệu khổng lồ như vậy, rõ ràng những trợ lý ảo sẽ có khả năng tìm kiếm cực kỳ mạnh mẽ cả về tốc độ lẫn sự chính xác trong kết quả tìm kiếm.

**Khả năng học hỏi người dùng**

Khả năng “học” từ thói quen sử dụng của người dùng được xem như ưu điểm cốt lõi của các trợ lý ảo. Như chúng ta đã biết, các trợ lý ảo được phát triển dựa trên trí thông minh nhân tạo (AI) nên chúng hoàn toàn có thể thu thập dữ liệu từ các thói quen sử dụng, tìm kiếm dữ liệu của người dùng và đưa ra những gợi ý chính xác nhất. Theo thời gian, lượng dữ liệu có được từ các hoạt động sử dụng của người dùng sẽ ngày một gia tăng và các trợ lý ảo sẽ trở nên thông minh hơn.

Theo thời gian và lượng dữ liệu có được từ các hoạt động sử dụng máy của người dùng, các trợ lý ảo sẽ ngày càng trở nên thông minh hơn, có thể tự thay người dùng thực hiện một số công việc nhờ công nghệ AI.

**Khả năng liên kết thông tin từ nhiều nguồn khác nhau**

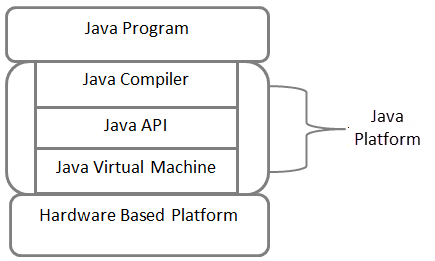
Một điểm mạnh khác của trợ lý ảo là khả năng liên kết thông tin giữa nhiều nguồn khác nhau. Những thông tin được liên kết có thể là lịch trình bay của bạn trong email, đồng thời nó cũng là ghi chú trong ứng dụng báo thức để nhắc nhở bạn khi sắp bay. Như vậy, thay vì phải ghi nhớ và nhập thông tin trong từng ứng dụng thì trợ lý ảo đã tự động làm hết những điều này cho bạn. Thông qua những điểm mạnh ở trên, chúng ta có thể thấy được sự tiện lợi mà trợ lý ảo mang lại cho mỗi chúng ta. Với trợ lý ảo, chắc chắn những trải nghiêm của bạn khi sử dụng smartphone, máy tính bảng,… sẽ trở nên tuyệt vời hơn.

Những thông tin được liên kết có thể ví dụ như lịch trình bay của bạn trong email sẽ được ghi chú trong ứng dụng Calendar (Lịch), cùng lúc là ghi chú trong ứng dụng báo thức để nhắc nhở bạn khi sắp bay. Như vậy, thay vì phải ghi nhớ và nhập thông tin trong từng ứng dụng, nay trợ lý ảo đã tự động làm hết những điều này cho bạn.

**2. Định nghĩa API**

API – Application Programming Interface – được định nghĩa là phương thức trung gian kết nối các ứng dụng và thư viện khác nhau. Một API gồm các lớp (classes), giao diện (interfaces), gói (packages), các phương thức (methods), các trường (fields) và cả các hàm tạo (constructors).

Cụ thể, API nằm trong bộ công cụ phát triển Java (JDK – Java Development Kit). JDK thuộc nền tảng Java. JDK gồm có 3 thành phần chính: Java compiler, Java Virtual Machine (JVM) và cuối cùng là Java API.



Nhiệm vụ của API là giải thích các chức năng của phần tử. Trong Java đã có một số lượng lớn các thành phần lập trình đã được tạo sẵn và sử dụng rộng rãi. Nhờ Java API mà các lập trình viên có thể sử dụng các chương trình viết sẵn này để rút ngắn số dòng lệnh phải viết. Cụ thể, việc sử dụng API sẽ giúp các coders lập tức tạo ra các lớp (classes) và gói (packages) lệnh bắt buộc của chương trình thay vì phải tự viết các câu lệnh.

Để nói về vai trò của API là gì trong Java thì nó là bộ công cụ giúp các nhà lập trình tối ưu hóa câu lệnh và tiết kiệm thời gian trong suốt quá trình lập trình. Java API là nhân tố quan trọng của JDK và của nền tảng Java.

**Ưu điểm của API**

Kết nối mọi lúc mọi nơi nhờ vào Internet

Cung cấp giải pháp rút ngắn thời gian để phát triển phần mềm, giờ đây chỉ cần API để sử dụng phần mềm được dựng sẵn.

Tạo ra trải nghiệm tốt và thân thiện đối với người dùng

Giao tiếp 2 chiều nên quá trình thực hiện khá đáng tin cậy

Tích hợp: Các API cho phép nhúng nội dung bởi bất kỳ trang web hoặc phần mềm nào một cách dễ dàng hơn. Điều này đảm bảo cung cấp dữ liệu linh hoạt bổ sung và trải nghiệm người dùng được tích hợp sẵn. Từ đó tối ưu hóa trải nghiệm người dùng.

Các mạng xã hội (Facebook, Twitter,…) là một minh chứng cho câu hỏi ưu điểm của API là gì. Nhờ có Java API với khả năng tích hợp mà người dùng có thể sử dụng nhiều tính năng cùng lúc như vừa trả lời tin nhắn, vừa cập nhật news feed,… cùng một lúc thay vì lần lượt.

Hiệu quả: Sau khi cung cấp quyền truy cập cho Java API, nội dung được tạo ra có thể được phát hành ngay lập tức và nó có sẵn cho mọi kênh. Điều này cho phép nó được phân phối cũng như gửi đi một cách nhanh chóng.

Java API là một công cụ vô cùng hữu ích giúp các nhà lập trình rút ngắn thời gian hoàn thành sản phẩm. Đồng thời nó giúp tăng hiệu năng cho phần mềm với khả năng tự động hóa, tích hợp, giúp ứng dụng hoạt động trơn tru và hiệu quả hơn. Hy vọng bài viết về vai trò của API là gì trong Java cũng như những ưu điểm của nó sẽ giúp bạn hiểu thêm về API nói riêng và Java platform nói chung.

**Nhược điểm của API**

Bất kể thứ gì cũng sẽ có một vài nhược điểm nhất định. Không nằm ngoài ngoại lệ, API cũng có một và nhược điểm khá đáng kể như:

Đòi hỏi người thực hiện phải có kiến thức chuyên sâu

Tốn rất nhiều chi phí để phát triển, vận hành cũng như chỉnh sửa API

Web API không hoàn toàn là RESTful service, chỉ hỗ trợ trợ cho POST và GET là chủ yếu.

Cần phải có kiến thức về bảo mật nếu không hệ thống sẽ rất dễ bị tấn công.

# 

# III, GIẢI QUYẾT BÀI TOÁN

**1. Sử dụng Web Speech API**

Web Speech API được định nghĩa là một API cho phép chúng kết hợp nhận dạng và tổng hợp giọng nói vào trang web. Nó cho phép chúng ta sử dụng script để chuyển text-to-speech và sử dụng nhận dạng giọng nói làm đầu vào. API cho phép trang web của chúng ta kiểm soát được việc kích hoạt, thời gian và xử lý kết quả. PI được thiết kế cho phép cả đầu vào giọng nói ngắn(một lần) và đầu vào giọng nói liên tục. Kết quả nhận dạng giọng nói được cung cấp cho trang web dưới dạng danh sách hypotheses cùng với các thông tin liên quan cho mỗi hypotheses

Speech synthesis: tổng hợp giọng nói hay chuyển văn bản thành giọng nói (text to speech)

Speech recognition: nhận dạng giọng nói hay chuyển giọng nói thành văn bản (speech to text)

**Cơ chế hoạt động**

Sau khi cấp quyền cho microphone recognition service bắt đầu lắng nghe và chúng ta nói điều gì đó, tùy theo cài đặt thuộc tính của chúng ta việc ngừng nói để gọi đến phương thứcstop() hay nút bấm để thông báo với recognition service sẽ dừng việc lắng nghe, sau đó:

* Đối với Cốc Cốc: Cốc Cốc lấy âm thanh và gửi đến máy chủ của Google để thực hiện quá trình phiên âm. Hiện tại speech recognition chỉ được hỗ trợ trong Cốc Cốc và một vài trình duyệt dựa trên Chromium

# 

# IV, Cài đặt và demo chương trình

**1. Cài đặt**

Chương trình được viết bằng : Javascript

Sử dụng Web Speech API.

**2. Demo chương trình**

Các chức năng chính của chương trình : lắng nghe và thực hiện yêu cầu của người nói



Nhấn Icon voice để bắt đầu lắng nghe yêu cầu của người dùng .

# 

# V, KẾT LUẬN ĐÁNH GIÁ

**1. Hướng phát triển của đề tài**

Tìm hiểu xây dựng một ứng dụng có thể nhận dạng và tìm kiếm bằng giọng nói thay cho việc phải viết .Có thể lấy được thông tin ngay bằng cách sử dụng các API giúp thuận tiện và nhanh chóng cho người dùng.

**2. Đánh giá**

Do sử dụng API nên bắt buộc phải có internet mới có thể sử dụng,

Giao diện còn đơn giản.

**Tài liệu tham khảo**

<https://freetuts.net/>

<https://www.javascript.com/>