

Youtube Assistant

David Palricas nº 108780 e Henrique Coelho nº 108342

Janeiro 2025

1 Introdução

O nosso assistente trata-se de uma assistente que controla o **YouTube** por voz e/ou por gestos pelo utilizador. O nosso assistente garante feedback via voz para cada ação feita pelo utilizador.

Este assistente foi desenvolvido em **Python**, com auxílio da biblioteca **Selenium**. Decidimos escolher esta linguagem de programação devido à sua simplicidade sintática e à biblioteca **Selenium** para controlar o navegador do utilizador.

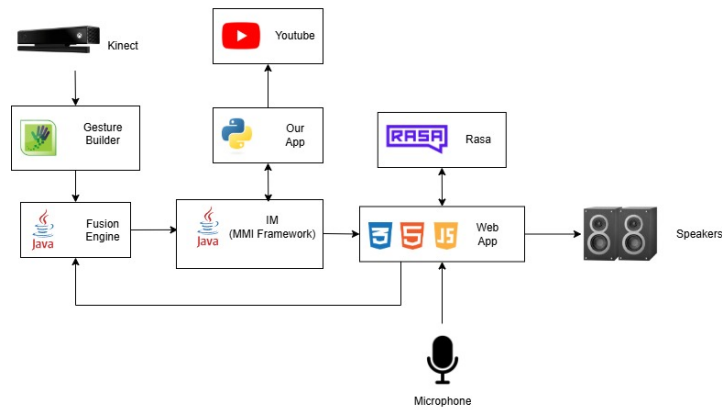


Figure 1: Arquitetura de Fusão

Para os módulos de **Speech Recognition**, **Text to Speech (TTS)**, **Modalidade de Gestos** e **Motor de fusão** foi utilizada a framework multi-modal do **IEETA**.

As funcionalidades escolhidas/implementadas, foram tendo em consideração o impacto destas no YouTube e também da nossa própria experiência de utilização da aplicação.

2 Requisitos

- Sistema operativo **Windows**
- A biblioteca **Selenium** integrada no **Python**
- **Google Chrome**
- Perfil criado no **Google Chrome**
- Conta de **YouTube** com sessão iniciada no **Chrome**
- Conta de **WhatsApp** com sessão iniciada no **Chrome**
- Versão 3.10 ou superior de **Python**
- Requisitos da framework multimodal do **IEETA**

3 Funcionalidades de Fusão

3.1 Redundantes

3.1.1 Meter um vídeo em ecrã cheio

Exemplo por Voz: Ecrã cheio por favor

Exemplo por Gestos: Junte as suas mão a frente do seu peito e estique os seus braços na diagonal mantendo as suas mão fechadas, quando não consegui -los esticar-los mais abra as suas mãos.

3.1.2 Meter um vídeo em ecrã normal

Exemplo por Voz: Ecrã normal por favor

Exemplo por Gestos : Estique os braços na diagonal até não conseguir mais, com as suas mãos abertas e lentamente aproxime os seus braços ao seu peito quando tiver as suas mão a frente do seu peito junte-as.

3.2 Complementares

Para as funcionalidades complementares, permitimos que o utilizador possa alterar o volume(aumentar/diminuir) de um vídeo utilizando expressões naturais, como "pouco", "um bocado", "mais um pouco", entre outras, ou simplesmente não acrescentar mais nenhuma expressão alterando o volume para o valor "padrão" definido por nós.

Consoante estas expressões, ajustamos os valores. Por exemplo, se o utilizador disser "aumente mais um pouco o volume", o volume aumenta em 25%. No caso de o utilizador não especificar nada, o volume do vídeo é alterado em 15% (valor padrão).

Adotamos o mesmo princípio para dar scroll na página para cima e para a baixo, por exemplo se utilizador disser "deslize um pouco para baixo" o valor de

scroll terá 200 unidades, no caso se disser "deslize mais um bocado para cima" o valor de scroll será de 600 unidades e se não especificar nenhum será de 400 unidades.

3.2.1 Aumentar o Volume de um vídeo do Youtube

Exemplo: Comece por estender um dos seus braços para o lado a nível da sua cintura e eleve-o até a sua cabeça. A seguir diga algo como "Aumenta um pouco o volume".

3.2.2 Diminuir o volume de um vídeo do Youtube

Exemplo: Estique o seu braço até a sua cabeça e baixe-o de lado, lentamente até altura da sua cintura. A seguir diga algo como "Diminui o volume do vídeo".

3.2.3 Scroll para cima

Exemplo: Estique um dos seus braços para frente e estenda o outro para baixo, com a palma da mão aberta. Eleve-o lentamente até o seu peito. A seguir diga algo como "Dá scroll um pouco para cima".

3.2.4 Scroll para baixo

Exemplo: Estique um dos seus braços para frente e estenda o outro para cima, com a palma da mão aberta. Baixe-o lentamente até o seu peito. A seguir diga algo como "Dá scroll mais um bocado para baixo".

4 Lista de funcionalidades das outras modalidades

4.1 Voz

1. Pesquisa de vídeos Exemplo: "Pesquise pelo vídeo Despacito."
2. Reproduzir/Parar um vídeo Exemplo: "Continue com o vídeo." ou "Pause o vídeo."
3. Ativar/Desativar o som do vídeo Exemplo: "Silencia o vídeo por favor." ou "Ative o som."
4. Ativar/Desativar Legendas de um vídeo Exemplo: "Ative as legendas." ou "Desative as legendas."
5. Alterar a velocidade de reprodução de um vídeo (alterar para um valor específico, diminuir ou aumentar a velocidade do vídeo para um valor específico ou não) Exemplo: "Altera a velocidade do vídeo para 1.5" "Aumente a velocidade do vídeo." ou "Reduza a velocidade do vídeo em 0.75."

6. Guardar um vídeo numa playlist Exemplo: "Guarde este vídeo na playlist favoritos."
7. Escrever um comentário no YouTube Exemplo: "Comenta Gostei muito deste vídeo"
8. Inscrever/cancelar inscrição do canal do vídeo selecionado (suporte a gestão de notificações no ato de inscrição do canal) O que eu faço: Inscrevo você no canal ou cancelo sua inscrição. Exemplo: "Subscreve o canal." ou "Cancela a subscrição."
9. Avançar/Recuar x tempo no vídeo Exemplo: "Avance 10 segundos." ou "Recua 15 segundos."
10. Partilha de vídeos via **Whatsapp** Exemplo: "Partilha o vídeo com o Artur."
11. Comando de ajuda a mostrar as funcionalidades das modalidades do sistema Exemplo: "Preciso de ajuda"
12. Desligar o assistente por voz Exemplo: "Adeus Assistente."

4.1.1 Notas da modalidade de voz

Para as funcionalidades que incluem intenções foi implementado um sistema de confirmação do assistente do valor destas. O assistente avança com a funcionalidade caso o grau de confiança seja superior a 80 %, caso contrário confirma com o utilizador o valor da entidade.

Esta confirmação tem duas fases distintas dependendo do valor do grau de confiança da entidade. Se o valor for inferior a 45 % o assistente irá reproduzir uma mensagem a indicar que não percebeu o que o utilizador queria. Se este valor for entre 45% e 79%, o assistente irá reproduzir uma mensagem a perguntar se era aquilo que o utilizador queria, e ativa o **speech recognition**, para este aceitar/negar a sua intenção.

Para estas ambas fases de confirmação, as mensagens de reprodução são adequadas para cada intenção e entidade. Estas mensagens estão definidas em ficheiros **JSON**.

Em ações mais críticas como partilha de vídeo para um contacto e desligar o assistente, o assistente pergunta ao utilizador ativando o **speech recognition** se deseja realizar esta ação ou não e só a conclui em caso afirmativo.

4.2 Gestos

1. Dar like em vídeos Exemplo: "Você faz um gesto com o pulgar virado para cima, com a mão a frente do seu peito"
2. Dar dislike em vídeos Exemplo: "Você faz um gesto com o pulgar virado para baixo, com a mão a frente do seu peito"

5 Instruções

Leia o ficheiro **README.MD**

Bibliografia

- [1] A. Teixeira, “Multimodal interaction fusion,” Apresentação de aula, Departamento de Eletrónica, Telecomunicações e Informática (DETI), Universidade de Aveiro, 2019.
- [2] A. Teixeira and M. O. e Silva, “Interacção multimodal: Fusion of input modalities,” Apresentação de aula, Departamento de Eletrónica, Telecomunicações e Informática (DETI), Universidade de Aveiro, 2024.
- [3] J. Barnett, “Introduction to scxml,” in *Multimodal Interaction with W3C Standards*, D. A. Dahl, Ed. Springer International Publishing, 2017, pp. 81–107.
- [4] A. Teixeira, “Im-2023-2024-dialog-systems-short,” PowerPoint presentation, 2024, accessed: Nov. 28, 2024.