

# YouTube Voice Assistant

David Palricas nº 108780 e Henrique Coelho nº 108342

Novembro 2024

## 1 Introdução

Para a entrega de interação por voz, decidimos optar por desenvolver um assistente de voz que controlasse o **YouTube** 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.

Para os módulos de **Speech Recognition** e **Text to Speech (TTS)**, foi utilizada a framework multimodal 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**

Por favor, leia o ficheiro **README.MD**, para instalação destes requisitos.

### 3 Funcionalidades

O nosso sistema, inclui um total de 12 funcionalidades, sendo elas:

1. Pesquisa de Vídeos
2. Reproduzir/Parar um vídeo
3. Ativar/Desativar o som do vídeo
4. Ativar/Desativar Legendas de um vídeo
5. Alterar a velocidade de reprodução de um vídeo (alterar para um valor específico, diminuir ou aumentar a velocidade do vídeo)
6. Guardar um vídeo numa playlist
7. Escrever um comentário no YouTube
8. Inscrever/cancelar inscrição do canal do vídeo selecionado (suporte a gestão de notificações no ato de inscrição do canal)
9. Avançar/Recuar x tempo no vídeo
10. Partilha de vídeos via **Whatsapp**
11. Comando de ajuda a mostrar as funcionalidades do sistema
12. Desligar o assistente por voz

A maior parte das funcionalidades têm um comportamento simples, ou seja, fazem praticamente o que está no nome, exceto para as funcionalidades nº **1**, **8** e **10**. O comportamento destas irá ser discutido mais a frente nesta secção.

#### 3.1 Feedback

O assistente em todas estas funcionalidades fornece feedback por voz sobre se intenção do utilizador foi feita com sucesso ou não. Caso não seja, avisa o assistente e diz o porquê, normalmente deve-se a falhas na busca de elementos do browser. Além disso, o assistente avisa o utilizador de outro tipo de falhas que deve-se ao caso da intenção do utilizador não poder ser executada naquele preciso momento, pois os pré requisitos desta ainda não foram cumpridos. A tabela seguinte mostrará exemplos dessas falhas.

#### 3.2 Grau de Confiança de Entidades

Para as funcionalidades que incluem intenções (funcionalidades nº **1**, **5 a 7** e **10**), 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.

Table 1: Tabela de Intenções e Falhas

Falha/Intent	Pausar/Parar	Alterar a velocidade de reprodução
Vídeo não selecionado	x	x
Parar um vídeo pausado	x	
Alterar a velocidade de um short		x
Velocidade já no máximo		x

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**.

Estes valores dos graus de entidades para confirmação destas, foram retirados do powerpoint **IM-2023-2024-DIALOG-SYSTEMS-SHORT** do professor António Teixeira [1]

### 3.3 Funcionalidades Complexas

#### 3.3.1 Pesquisa de vídeos(nº1) e partilha de vídeos de Whatsapp(nº10)

Quando o utilizador pede ao assistente para pesquisar um vídeo ou partilhar com um contacto, o assistente informará o utilizador dos três primeiros resultados da sua pesquisa. Este também informa o utilizador para escolher por ordem qual dos resultados quer (primeiro, segundo ou terceiro). Após isto, o **speech recognition** é ativo, para o utilizador indicar a sua escolha. Depois da conclusão desta escolha o assistente concluirá com a sua funcionalidade.

Na funcionalidade de pesquisa de vídeo, queríamos que o assistente informasse caso um anúncio estava a ser reproduzido e indicasse se dava para ignorá-lo ou não. Caso desse para ignorá-lo, confirmava com o utilizador se queria ignorar ou não. Mas, devido ao mecanismo de segurança do **YouTube**, para evitar o uso de **AdBlockers**, não conseguimos fazer a deteção de anúncios, por isso aconselhamos que este sistema seja usado com uma conta premium do **Youtube**, para evitar anúncios.

#### 3.4 Inscrever/cancelar inscrição do canal do vídeo selecionado(nº8)

Nesta funcionalidade, a parte mais complexa é no ato de inscrição, pois cancelar a inscrição é direta. Quando o utilizador pede ao assistente para inscrever um

canal, o assistente pergunta se o utilizador quer desativar as notificações, ativar todas as notificações ou manter as preferências atuais (notificações padrão). Após esta pergunta o **speech recognition** é ativo para o utilizador indicar a sua escolha. Depois da escolha do utilizador o assistente concluirá com a sua funcionalidade, selecionando a opção escolhida.

## 4 Instruções

Leia o ficheiro **README.MD**

## Bibliografia

- [1] A. Teixeira, “Im-2023-2024-dialog-systems-short,” PowerPoint presentation, 2024, accessed: Nov. 28, 2024.
- [2] D. Palricas, “Kitchen Assistant for Active Ageing at Home,” <https://github.com/DavidPalricas/KitchenAssistant>, 2024, accessed: Nov. 27, 2024.
- [3] Rasa, “Conversational ai with rasa open source 3.x,” <http://www.youtube.com/playlist?list=PL75e0qA87dIEjGAc9j9v3a5h1mxI2Z9fi>, 2024, accessed: Nov. 27, 2024.
- [4] A. Teixeira, “Im-2023-2024-nlu-rasa-new-2023,” PowerPoint presentation, 2024, accessed: Nov. 28, 2024.