YouTube by Voice

Speech-based input and output

David Palricas nº 108780
Henrique Coelho, nº 108342

Multimodal Interaction - TP1
Master's in Digital Game Development
Master's in Informatics Engineering
29/11/2024

universidade
de aveiro

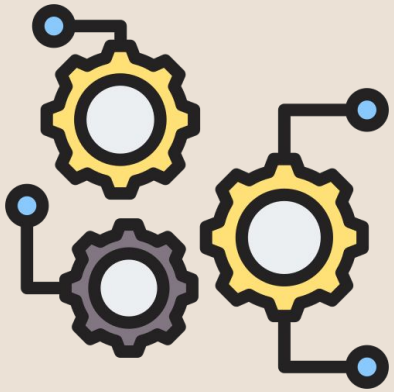
Introduction

- **Voice-Controlled YouTube Assistant** has the objective to enhance the user experience on YouTube through speech processing technologies. Our assistant integrates **Rasa** for natural language understanding and the **IEETA Multimodal Framework** for speech processing, combined with **Selenium/WebDriver** to automate various YouTube functionalities, allowing hands-free control of the platform.

- Through this presentation, we will talk about:
  - **Functionalities**
  - **Entities**
  - **Regex**
  - **Demo**

Functionalities

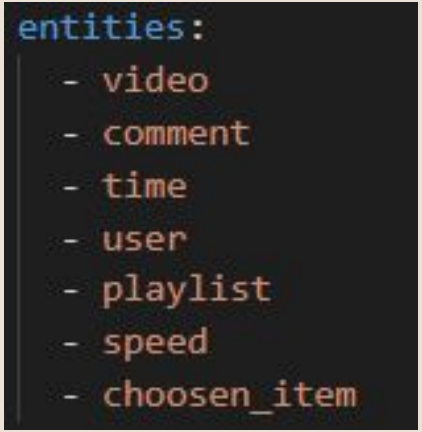
1. Subscribe Channels
2. Save Videos in playlists
3. Share Video (Whatsapp)
4. Activate/Deactivate Subtitles
5. Play/Stop Videos
6. Activate/Deactivate Sound
7. Seek Forward and Seek Backward
8. Search Videos
9. Video Speed Adjustment
10. Write Comments
11. Help Command
12. Shutdown the Assistant



Entities

We have **7 entities**:

- **video:** Used to Search Videos
- **comment:** Used for Writing Comments
- **time:** Used to set the time for the Seek Forward and Seek Backward
- **user:** Used to Share Video (through Whatsapp) with a person
- **playlist:** Used to Save Videos in the chosen playlist
- **speed:** Used for Video Speed Adjustment
- **choosen_video:** Used to Choose one of the Top 3 listed items (it can be a list of Videos or Contacts)



Regex

```
- regex: time
  examples: |
    - [0-9]+ (segundos|segundo|minutos|minuto|hora|horas)
```

This regular expression identifies time patterns in the format:

- [0-9]+ (seconds|second|minutes|minute|hour|hours)

What does it do?

- [0-9]+: Matches a sequence of one or more digits.
- (seconds|second|minutes|minute|hour|hours): Ensures the digits are followed by a time-related word.

Examples of matches:

- "10 seconds"
- "1 minute"
- "3 hours"

DEMO

Any Question?