

Universidade de Aveiro

Departamento de Eletrónica, Telecomunicações e
Informática

Interação Multimodal

Ano 2019/2020

Trabalho Prático nº 2

Realizado Por:

André Almeida N°: 62058

João Coelho N°: 80335

Aveiro, 15 de Novembro de 2019



Índice

Introdução	2
Identificação da aplicação	2
Cenário	3
Funcionalidades	4
Instruções de utilização	4
Implementação - Comandos por voz implementados	5
Exemplo de interação	8
Níveis de confiança	9
Feedback	10
Gramática	11
Referências	11

Introdução

Para este trabalho esperamos desenvolver um sistema que tenha uma interação natural entre o utilizador e o computador de modo a simplificar um pedido na Uber Eats, usando mecanismos de voz.

Será feita uma comunicação entre um reconhecedor de voz com uma gramática criada para os casos de interação, um “Interaction Manager” (IM) que envia e recebe os dados de informação e a API da aplicação. Como feedback, o sistema terá um sintetizador de voz que responde ao utilizador de acordo com os pedidos efetuados.

Identificação da aplicação

A aplicação escolhida para o desenvolvimento deste trabalho foi o website da plataforma Uber Eats. Esta plataforma permite a uma pessoa, encomendar comida de uma seleção de vários restaurantes até sua casa, numa forma fácil e conveniente. A plataforma está “dividida” por várias etapas, como: escolher o restaurante, escolher a secção, o produto e por último os detalhes do produto.

Adicionalmente, o utilizador poderá modificar ou adicionar novos produtos ao carrinho.

Cenário

O Uber Eats é uma plataforma que pode facilitar a tarefa de encomendar comida de um certo restaurante. Neste contexto, um cenário de aplicação é demonstrado abaixo.

O utilizador começa por dizer a frase que ativa o sistema (“Uber Eats Procura”) e, de seguida, pede ao programa para lhe apresentar o restaurante desejado. Enquanto isto acontece, o sistema vai interagindo com o utilizador por voz.

O sistema faz uma procura de acordo com o restaurante pretendido e mostra os produtos/opções a escolher. O utilizador escolhe o produto que quer e o sistema seleciona-o, perguntado se quer alterar algum detalhe do produto.

Pode, depois, adicionar o produto ao carrinho de compras e proceder ao pagamento manualmente.

Funcionalidades

As funcionalidades do sistema são:

1	Procurar por restaurantes.
2	Procurar por produtos dinamicamente.
3	Adicionar produtos ao carrinho.
4	Alterar detalhes do produto.
5	
6	Fazer scroll.

Instruções de utilização

As instruções de utilização são bastante simples: basta abrir os projetos “AppGui” e “speechModality”. É preciso também executar o módulo mmiframeworkV2.

Passados alguns segundos, o browser “chrome” irá estar aberto e pronto para os comandos de voz serem executados.

O sistema só inicia a escuta dos comandos quando ouve pela primeira vez a frase: “Uber Eats Procura”.

Exemplo de interação

U - Utilizador; S - Sistema.

U	Uber Eats Procura
S	Olá! O que gostaria de fazer?
U	Gostaria de encomendar do McDonald's
S	De qual?
U	Do McDonald's do Fórum
S	Que produto quer adquirir?
U	Um Signature Classic
S	Ok, quer alterar algum detalhe?
U	Sem bacon
S	Parece que estou a ficar surda, importa-se de repetir?
U	Sem bacon
U	Adicionar ao carrinho
S	Adicionado com sucesso!
U	Uber Eats Sair
S	Ok, até breve!

Níveis de confiança

Sendo que o reconhecimento de voz nem sempre é exato, este funciona com um nível de confiança. Nós usamos o nível de confiança para fazer a distinção em dois níveis e, por falta de tempo, não usamos em mais níveis (como por exemplo, se a percentagem de confiança estivesse entre 60% e 75%, perguntar ao utilizador se está correto o que foi percebido e, em caso positivo, usar um buffer para recuperar o que foi percebido no ciclo anterior).

- Quando o nível de confiança é inferior a 65%, o comando é ignorado pois não é suficientemente preciso para reconhecer um comando válido. O Sintetizador de Voz pede ao utilizador para repetir o comando.
- Quando o nível de confiança é superior a 65% o comando é executado e o Speech Recognition tem bastante certeza do que o utilizador lhe pediu para executar. O Sintetizador de Voz responde confirmando a operação pedida.

Feedback

O feedback do sistema é efetuado pelo Sintetizador de Voz da Hélia e é inicializado logo no início do programa assim que a frase de ativação é ouvida. O *sistema* começa por dar as boas vindas ao utilizador e fica pronto para ouvir.

De acordo com os níveis de confiança do Speech Recognition do discurso do utilizador, o Sintetizador de Voz dá-lhe um feedback da execução dos comandos pedidos, pedindo o próximo passo. Quando o Speech Recognition não percebe a maior parte do que o utilizador diz, o feedback que este tem é de que não foi entendido o seu discurso e se poderia repetir o comando.

Para cada comando disponível existem vários “templates” de frases que o Sintetizador de Voz responde ao utilizador, permitindo uma interação mais natural.

Gramática

Para que o utilizador tenha uma maior liberdade para dizer os comandos que quer que sejam executados, a gramática escolhida passa por ter o primeiro argumento obrigatório e os restantes opcionais podendo variar de acordo com o que o utilizador pretende realizar.

Algumas das palavras escolhidas na gramática:

KeyWord	Uber Eats	Uber Eats Sair	Uber Eats Tchau
Actions	scroll	Voltar	encomendar
Restaurants	McDonald's	100 montaditos	pizza hut
Places	Fórum	Glicínias	Universidade
Options	Novidades	McMenus	Sobremesas
Food	Signature Classic	Chicken Delights	Menu CBO
Modify	Sem Bacon	Sem Queijo	Sem Ketchup

Referências