

UNIVERSIDADE DE AVEIRO

DEPARTAMENTO DE ELETRÓNICA, TELECOMUNICAÇÕES E
INFORMÁTICA

MESTRADO EM ENGENHARIA INFORMÁTICA

Trabalho Prático nº2



Autores:

Andreia MACHADO, Nº 76501

Pedro MATOS, Nº 73941

INTERAÇÃO MULTIMODAL

20 de Abril de 2018

Conteúdo

Apresentação do Cenário	3
Funcionalidades	3
Instruções de instalação e uso	4
Escolhas de implementação	5
Comandos implementados	6
Gramática	6
Text to Speech	7
Pipe entre Speech Recognizer e TTS	7
Exemplo de Interação	8
Feedback	9

Apresentação do Cenário

Um utilizador da plataforma Discord [1], responsável por fazer a administração de dois servidores, cada um com vários canais de conversação, está a fazer as suas tarefas ao mesmo tempo que faz a gestão de moderação dos utilizadores em cada um dos servidores.

Olhando para a última mensagem presente no canal de conversação repara que esta tem conteúdo impróprio de mal gosto, de maneira a não parar o trabalho que está a fazer pede ao *bot wally* que apague essa última mensagem.

Depois de olhar mais atentamente para todo o conteúdo de um canal repara que está repleto de conteúdo inapropriado, decide apagar o canal, por isso pede ao *wally* para apagar esse canal.

O utilizador que anteriormente publicou uma mensagem com conteúdo impróprio, voltou a repetir a ação. Desta vez o utilizador com permissões de administração decide expulsá-lo do servidor e pede ao *wally* que o faça. Apesar de ter sido expulso o utilizador pode voltar ao servidor e caso volte a comportar-se de maneira inadequada pode ser banido do servidor.

O administrador está farto de ouvir os outros utilizadores, então pede ao *wally* que corte as comunicações com outros utilizadores, de maneira a que não os ouça e eles também não sejam capazes de o ouvir.

Funcionalidades

O sistema implementado serve para gerir a plataforma *Discord* de um utilizador, usando apenas comandos de voz, no qual o sistema reconhece o comando, e com o uso de um *bot* chamado *Wally* executa os comandos que o utilizador deseja.

As funcionalidades do sistema são:

1. Remover utilizadores dos servidores no qual o utilizador é administrador.
2. Banir utilizadores dos servidores no qual o utilizador é administrador.

3. remover banimentos aos utilizadores banidos.
4. Eliminar a última mensagem dos canais.
5. Eliminar canais dos servidores no qual o utilizador é administrador.
6. Sair de um servidor.
7. Saber o estado dos utilizadores (ex: online, offline).
8. Ativar ou desativar o microfone dos utilizadores.
9. Ativar ou desativar o áudio dos utilizadores.
10. Ativar ou desativar o áudio e o microfone dos utilizadores, ao mesmo tempo.

Instruções de instalação e uso

O primeiro passo para utilizar a aplicação é abrir a plataforma Discord (<https://discordapp.com/>) num browser e clicar em *Abrir Discord*. De seguida introduzir as seguintes credenciais:

- E-Mail: *IMStudent2018@gmail.com*
- Senha: *ambrosio*

Caso seja pedido para confirmar a autenticação por e-mail, a palavra passe do mesmo é igual à anterior.

Depois da autenticação irá ser apresentado com uma página ilustrada na Figura 1. Os dois servidores em que o utilizador tem permissões de administração estão localizados no canto superior esquerdo, clique num deles para entrar e de seguida clique no canal *teste* ou no canal *sueca*. A Figura 2 exemplifica a página que deverá ser apresentada.

Depois de entrar na plataforma é necessário iniciar a aplicação de *speechModality* e executar o jar *mmiframeworkV2*. Para executar a aplicação *DiscordControler* é necessário antes instalar uma dependência externa *Discord.Net* [2]. Para tal, siga os seguintes passos:

1. Abrir a solução.
2. Clicar com o botão direito do rato no símbolo da solução.
3. Selecionar a opção *Manage NuGet Packages*.
4. Nas caixa de pesquisa inserir “discord”.
5. O pacote *Discord.Net*[2] deverá aparecer entre os primeiros pacotes, clique na opção para o instalar.

A voz utilizada para o *Text To Speech* é a voz do Nuno, disponibilizada pelos docentes desta unidade curricular, no qual é necessário instalar antes de executar a aplicação *DiscordControler*.

Depois de completar estes passos a aplicação pode ser executada. Depois de receber a saudação do *wally* pode começar a dizer comandos.

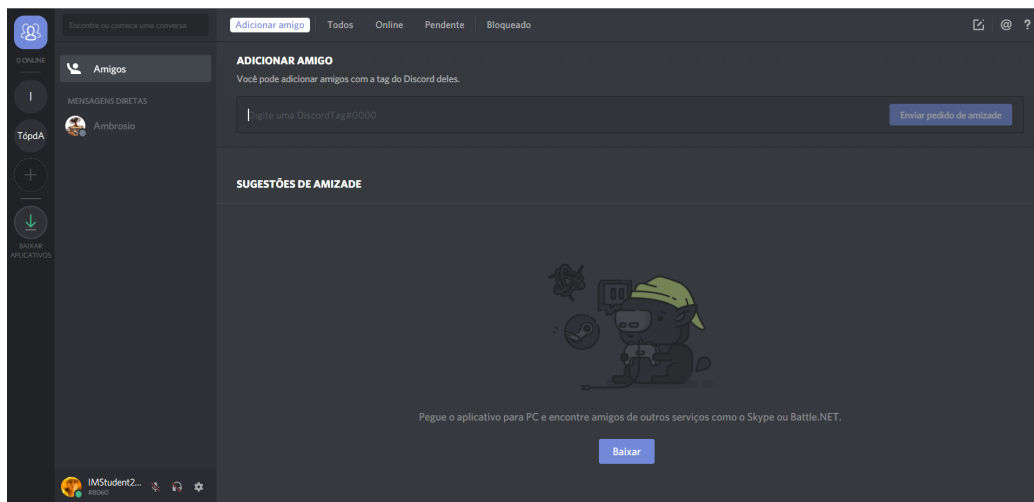


Figura 1: Plataforma *Discord* depois da autenticação

Escolhas de implementação

Nesta secção serão explicadas todas as escolhas de implementação, desde o reconhecimento dos comandos até ao feedback dado ao utilizador através do *text to speech* (TTS).

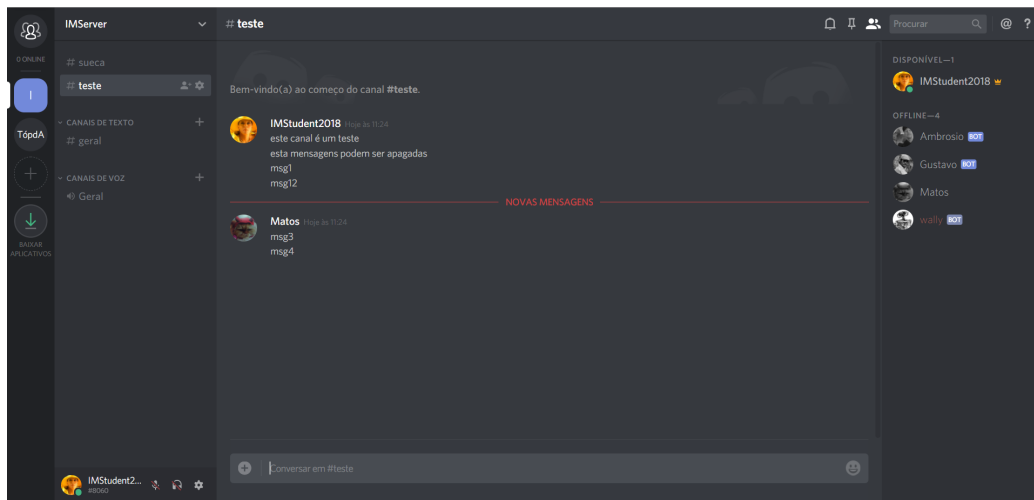


Figura 2: Vista do servidor *IMServer* no canal *teste*

Comandos implementados

Todos os comandos de voz implementados, à exceção da confirmação explícita, necessitam que comecem por *Wally*, para que o sistema saiba que o utilizador está a falar para o mesmo e não para outras pessoas, estejam estas ao lado do utilizador ou a falar pela plataforma *Discord*.

Os comandos implementados têm todos pelo menos 2 alternativas possíveis, para que o utilizador tenha alguma liberdade no que pode dizer.

Gramática

A gramática desenvolvida é uma gramática estática, no qual o utilizador tem de dizer certas palavras chaves para que o sistema entenda o comando que o utilizador quer executar. Por exemplo, para eliminar um canal o utilizador tem de dizer as seguintes palavras:

- *Wally*
- elimina, remove ou apaga
- *channel* ou canal
- teste, geral ou sueca

- *server*, servidor, grupo, comunidade ou *guild* (opcional)
- IMServer ou Tópicos de Apicultura (opcional)

Esta escolha foi feita para que o utilizador tenha a máxima liberdade possível para dizer os comandos que quer que sejam executados.

Para grande parte dos comandos desenvolvidos, o nome do servidor encontra-se opcional, ou seja, o utilizador pode dizer ou não o nome do servidor no qual quer executar o comando desejado. Por omissão, o servidor considerado é o *IMServer*, já que este é o servidor principal do utilizador. Esta escolha deve-se ao facto de o utilizador não ter a necessidade de em todos os comandos dizer o nome do servidor que deseja ver os comandos serem executados. Apenas quando o servidor não é *IMServer*, o utilizador tem de dizer o nome do servidor.

Para banir ou remover utilizadores de um servidor, o *Discord.Net*, permite que as razões da expulsão ou da remoção não sejam referidas, por isso, o utilizador tem a opção de dizer ou não a razão pela qual decidiu banir ou remover o utilizador.

Text to Speech

O *Text to Speech* (TTS) desenvolvido tem pelo menos 2 *templates* para cada feedback que o sistema dá ao utilizador, para que o utilizador nunca oiça a mesma frase várias vezes e enriquecer de alguma maneira o *flow* entre o sistema e o utilizador.

O utilizador tem sempre feedback sobre os comandos executados, visto que existe templates para qualquer comando executado com sucesso ou não.

Pipe entre Speech Recognizer e TTS

Quando o TTS fala existe a possibilidade de o *Speech Recognizer* reconhecer o que é dito pelo TTS, por isso foi necessário criar um *pipe* entre o *Speech Recognizer* e o TTS, para que o *Speech Recognizer* pare de reconhecer quando o TTS fala.

Para isto ser possível, é criado no Speech Recognizer um *pipe server* e no TTS um *pipe client*, que se irá conectar ao *pipe server* do *Speech Recognizer*. Quando o TTS começa a falar, é enviado por esse *pipe* uma mensagem “ttsSpeaking”. De seguida, a mensagem é recebida no lado do *Speech Recognizer*, e consequentemente o *recognizer* para de reconhecer. Logo após o TTS terminar de falar, é enviado pelo *pipe* a mensagem “ttsNotSpeaking”, no qual quando recebida pelo *Speech Recognizer* volta a reconhecer o que é dito pelo utilizador.

Exemplo de Interação

O utilizador pode começar por pedir ao “bot” os comandos de voz que pode utilizar ou o que o “bot” consegue fazer:

- “Wally diz-me os comandos disponíveis”
- “Wally diz-me o que podes fazer”

Apagar a última mensagem do canal *teste* do servidor *IMServer*:

- “Wally apaga a última mensagem do canal teste do servidor IMServer”

Silenciar um utilizar que está a fazer muito barulho pelo microfone:

- “Wally silencia o utilizador Gustavo”

Cortar todas as comunicações com um utilizador, ou seja, não ouve o utilizador nem o utilizador o ouve a si:

- “Wally corta as comunicações com utilizador Ambrósio”

Expulsar o utilizador Gustavo do servidor Tópicos de Apicultura:

- “Wally tira o utilizador Gustavo do servidor Tópicos de Apicultura”

Banir o utilizador Ambrósio do servidor IMServer:

- “Wally bane o Ambrósio do servidor IMServer”

Feedback

O feedback dado ao utilizador pelo sistema é através do *Text to Speech*, no qual qualquer funcionalidade do sistema tem feedback, mesmo quando o utilizador quer fazer alguma tarefa que não é possível.

O sistema desenvolvido possui 4 níveis de confiança (*Four-Tiered Confidence*), que são os seguintes:

1. Quando o valor de *confidence* é menor que **30%**, é ignorado o que é reconhecido porque o utilizador poderá estar a falar com outras pessoas e o *recognizer* pode reconhecer palavras chaves que se encontra nos comandos disponíveis.
2. Quando o valor de *confidence* encontra-se entre **31% e 45%**, é pedido ao utilizador que repita o comando dito, por haver incerteza do que foi reconhecido.
3. Quando o valor de *confidence* encontra-se entre **46% e 79%**, é feita uma confirmação explícita, no qual é repetido o que o utilizador quer fazer, em forma de questão. Isto é feito para que o sistema tenha a certeza que, o que é reconhecido, é o que o utilizador deseja executar.
4. Quando o valor de *confidence* é maior que **80%**, o comando reconhecido é executado e é feita uma confirmação implícita para dar feedback ao utilizador, para que este tenha conhecimento do que foi entendido e consequentemente executado.

Quando o sistema é iniciado, é dado feedback ao utilizador, de que a aplicação foi iniciada e que esta pronta para reconhecer comandos. Este feedback inicial é essencial porque é a forma de dizer ao utilizador que o sistema esta pronto para reconhecer comandos, e se quiser pode perguntar ao mesmo quais são os comandos disponíveis e o que o sistema pode fazer, como uma forma de ajudar o utilizador a usar o sistema, caso seja a sua primeira vez a usa-lo.

Referências

- [1] Discord Brand. *Discord*. 2018. URL: <https://discordapp.com/> (acesso em 13/04/2018).

- [2] RogueException. *Discord.Net*. 2018. URL: <https://github.com/RogueException/Discord.Net> (acedido em 13/04/2018).