



universidade de aveiro

Departamento de Eletrónica, Telecomunicações e Informática

Curso: Engenharia de Computadores e Telemática

Unidade Curricular: Interação Multimodal

Ano Escolar: 2018/2019

Jarbas

Controlador de voz do sistema Windows e aplicações integradas

Trabalho 3

Trabalho elaborado por:

Cristiano Vagos nº 65169

Gabriel Patrício nº 76137

Data:

17 de Janeiro de 2019

Índice

Introdução

Arquitectura

Speech Modality.

Gesture Modality

AppGui

Gramática

Fusão

Construção do Fusion.scxml

Cenários Redundantes

Cenários Complementares

Cenários Singulares

Extra

Use Cases

Voz

Gestos

Cenário de utilização

Exemplo

Instruções para correr a aplicação

Introdução

Inspirado pelo “Jarvis” do filme Iron Man, o “Jarbas” foi criado com o intuito de tornar a interação entre o utilizador e o sistema Windows, assim como algumas das suas aplicações, mais dinâmico e prático.

Básicas interações como desligar o computador, manipular janelas, controlar volume, luminosidade e abrir uma série de aplicações como email, meteorologia, players de musica ou video são “cobertas” pela nossa aplicação com comandos de voz e gestos que tentámos humanizar o máximo possível para que a interação fosse o mais natural e simples.

Arquitectura

Speech Modality

Módulo ASR, desenvolvido a partir da base dada nas aulas. Aqui foi implementada a gramática do Jarbas (jarbasGrammar.grxml) e o processamento dos comandos de voz. Por defeito, descartamos todo o reconhecimento feito com uma confiança abaixo dos 60%, e não é sequer enviado para o servidor.

Gesture Modality

Segundo módulo implementado com base no exemplo que nos foi facultado nas aulas. Com o auxílio do KinectStudio foram gravados todos os gestos a ser utilizados nesta modalidade é criada uma base de dados à qual acedemos para que seja feito o reconhecimento dos gestos a usar no sistema Jarbas.

Dependente do gesto o grau de confiança difere pois há gestos em que é mais fácil o seu reconhecimento do que outros.

AppGui

É no AppGui é que é feita a fusão de todos os gestos e comandos de voz reconhecidos e lhes é atribuída o método/acção correspondente. São também definidos todos os comandos de feedback de voz usados pelo Sistema Jarbas.

Gramática

A gramática do Jarbas foi desenvolvida tendo por base a especificação GrammarXML da W3C, sendo desenvolvida em XML e executada no ASR.

Separou-se cada um dos comandos a efetuar em regras (rules) que os especificam, reunindo várias frases (items) em sequência, que assim que uma delas seja reconhecida (one-of) devolve o tipo de comando reconhecido através de uma “tag” que é enviada para a aplicação do Jarbas. Desta forma conseguimos modularizar as frases / palavras a serem reconhecidas pelo ASR.

Não foram colocadas todas as frases possíveis para cada tipo de comando / operação (senão a gramática ficaria gigantesca), no entanto desenvolvemos a gramática com base num diálogo normal e frequente, tentando humanizar ao máximo os comandos a serem executados (conforme recomendação dos professores).

```
<rule id="cmdDesligar">
  <item>
    <one-of>
      <item> até amanhã </item>
      <item> por hoje
        <item repeat="0-1"> já </item>
      <item> chega </item>
    </one-of>
  </item>
  <item> desliga o
    <one-of>
      <item> PC </item>
      <item> computador </item>
    </one-of>
  </item>
</one-of>
</item>
<tag>out = "desligar"</tag>
</rule>
```

Existem várias ocorrências na gramática que são opcionais. Por exemplo, podemos pedir para o Jarbas executar um comando com vários tipos de fala (pedir educadamente ou exigir, por exemplo), pelo que se criou um iniciador/terminador, regras que poderão ser ou não cumpridas de acordo com as palavras que poderão ser ditas para executar um comando (ex. “Jarbas, se possível diz-me que horas são”, ou “Jarbas, diz-me horas já”).

```

<rule id="initiator">
  <item repeat="0-1">
    <one-of>
      <item> importas-te de </item>
      <item> se faz favor </item>
      <item> por favor </item>
      <item> se possível </item>
      <item> é pá </item>
    </one-of>
  </item>
</rule>

```

```

<rule id="terminator">
  <item repeat="0-1">
    <one-of>
      <item> se faz favor </item>
      <item> por favor </item>
      <item> se possível </item>
      <item> já </item>
      <item> agora </item>
      <item> meu </item>
      <item> pá </item>
    </one-of>
  </item>
</rule>

```

Fusão

Construção do Fusion.scxml

Utilizou-se o projecto NetBeans fornecido para a geração do ficheiro de suporte à fusão, para tal encontrou-se funcionalidades que pudessem fazer sentido na utilização do Jarbas e que de alguma forma pudessem usar todos os tipos de cenários que abaixo definimos.

Cenários Redundantes

```

// REDUNDANCIAS
fg.Redundancy(Speech.S_SCREENSHOT, SecondMod.G_SCREENSHOT, Output.SCREENSHOT);
fg.Redundancy(Speech.S_LAZER, SecondMod.G_HANDS_AIR, Output.MODO_LAZER);
fg.Redundancy(Speech.S_TRABALHO, SecondMod.G_KILL, Output.MODO_TRABALHO);
fg.Redundancy(Speech.S_LOCK, SecondMod.G_JARBAS_INIT, Output.LOCK_SESSION);

```

Cenários Complementares

```
// COMPLEMENTARIDADE
fg.Complementary(Speech.S_NEXT_TRACK, SecondMod.G_SWIPE_RIGHT, Output.NEXT_TRACK);
fg.Complementary(Speech.S_PREVIOUS_TRACK, SecondMod.G_SWIPE_LEFT, Output.PREVIOUS_TRACK);
fg.Complementary(Speech.S_BRIGHTNESS_DOWN, SecondMod.G_SWIPE_RIGHT, Output.BRIGHTNESS_DOWN);
fg.Complementary(Speech.S_BRIGHTNESS_UP, SecondMod.G_SWIPE_LEFT, Output.BRIGHTNESS_UP);

fg.Complementary(Speech.S_EMAIL, SecondMod.G_SWIPE_LEFT, Output.EMAIL_LEFT);
fg.Complementary(Speech.S_EMAIL, SecondMod.G_SWIPE_RIGHT, Output.EMAIL_RIGHT);
fg.Complementary(Speech.S_CALC, SecondMod.G_SWIPE_LEFT, Output.CALC_LEFT);
fg.Complementary(Speech.S_CALC, SecondMod.G_SWIPE_RIGHT, Output.CALC_RIGHT);
fg.Complementary(Speech.S_WEBSITE, SecondMod.G_SWIPE_LEFT, Output.WEBSITE_LEFT);
fg.Complementary(Speech.S_WEBSITE, SecondMod.G_SWIPE_RIGHT, Output.WEBSITE_RIGHT);
fg.Complementary(Speech.S_METEO, SecondMod.G_SWIPE_LEFT, Output.METEO_LEFT);
fg.Complementary(Speech.S_METEO, SecondMod.G_SWIPE_RIGHT, Output.METEO_RIGHT);
fg.Complementary(Speech.S_CAMERA, SecondMod.G_SWIPE_LEFT, Output.CAMERA_LEFT);
fg.Complementary(Speech.S_CAMERA, SecondMod.G_SWIPE_RIGHT, Output.CAMERA_RIGHT);
```

Cenários Singulares

```
// SINGLE
fg.Single(SecondMod.G_SWIPE_LEFT, Output.SWIPE_LEFT);
fg.Single(SecondMod.G_SWIPE_RIGHT, Output.SWIPE_RIGHT);
```

Use Cases

Voz

Acção	Comando de voz
Saudação / dizer olá	1. “Olá Jarbas” 2. “Olá”
Insultar	1. “Cala-te” 2. “Já te calavas” 3. “Estou/tou farto/cheio/cansado de te ouvir”
Agradecer	1. “É isso tudo” 2. “Obrigado” 3. “Está/tá demais/brutal”
Perguntar o estado do Jarbas	1. “Jarbas, como é que estás hoje?” 2. “Jarbas, tudo bem contigo?” 3. “Jarbas, como te sentes?”
Perguntar quais os developers	1. “Jarbas, quem foi/é que te fez?” 2. “Jarbas, onde foste feito?”
Desligar o “Jarbas”	1. “Jarbas, xau”

	<ol style="list-style-type: none"> 2. “Jarbas, quero/podes sair” 3. “Jarbas, já não preciso de ti”
Configuração do “Jarbas”	<ol style="list-style-type: none"> 1. “Jarbas, quero configurar-te” 2. “Jarbas, mudar configurações” 3. “Jarbas, vamos/que tal mudar(mos) as tuas opções/configurações?”

Controlo de janelas

Acção	Comando de voz
Fechar janela	<ol style="list-style-type: none"> 1. “Jarbas, fecha a primeira/segunda/terceira janela”
Maximizar janela	<ol style="list-style-type: none"> 1. “Jarbas, maximiza a primeira/segunda/terceira janela”
Minimizar janela	<ol style="list-style-type: none"> 4. “Jarbas, minimiza a primeira/segunda/terceira janela”

Controlo do Sistema Operativo

Acção	Comando de voz
Tirar screenshot	<ol style="list-style-type: none"> 1. “Jarbas, tira foto ao ecrã” 2. “Jarbas, tira/tira-me um printscreen”
Desligar o sistema	<ol style="list-style-type: none"> 1. “Jarbas, até amanhã” 2. “Jarbas, por hoje chega” 3. “Jarbas, desliga o PC”
Bloquear Sessão	<ol style="list-style-type: none"> 1. “Jarbas, vou sair, bloqueia-me a sessão” 2. “Jarbas, bloqueia a sessão”
Informar percentagem de bateria	<ol style="list-style-type: none"> 1. “Jarbas, qual é o estado da bateria?” 2. “Jarbas, quanta bateria tens?”

Controlo de Aplicações

Acção	Comando de voz
Utilizar Câmara	<ol style="list-style-type: none"> 1. “Jarbas, tira-me uma foto” 2. “Jarbas, mostra-me a webcam”
Saber a data/horas	<ol style="list-style-type: none"> 1. “Jarbas, diz-me que dia é hoje” 2. “Jarbas, posso saber a que dia estamos?” 3. “Jarbas, diz-me em que ano estamos” 4. “Jarbas, mostra-me as horas” 5. “Jarbas, que horas são?”
Desligar “Jarbas”	<ol style="list-style-type: none"> 1. “Jarbas, xau” 2. “Jarbas, já não preciso de ti”
Saber meteorologia	<ol style="list-style-type: none"> 1. “Jarbas, posso saber o estado do tempo?” 2. “Jarbas, sabes se preciso de ir de guarda chuva?” 3. “Jarbas, sabes se tenho de levar chapéu?” 4. “Jarbas, diz-me se está bom para ir à praia”
Abrir Email	<ol style="list-style-type: none"> 1. “Jarbas, mostra-me o meu mail” 2. “Jarbas, mostra aí a minha caixa de correio”
Iniciar leitor de música (Spotify ou Windows Media Player) <i>Spotify abre por defeito, Windows Media Player caso não haja Spotify instalado</i>	<ol style="list-style-type: none"> 1. “Jarbas, dá-me música” 2. “Jarbas, mete aí algo/uma cena a tocar” 3. “Jarbas, vamos ouvir música/alguma coisa”
Controlo de leitor de música (Play/Pause, Stop, Mudar faixa, Aumentar/diminuir volume)	<ol style="list-style-type: none"> 1. “Jarbas, não gosto dessa música” 2. “Jarbas, desliga o som” 3. “Jarbas, não ouvi” 4. “Jarbas, mete o som mais alto” 5. “Jarbas, estava a gostar da outra música”

Gestos

Independente

Acção	Gesto
-------	-------

Entrar em Modo de Trabalho	Juntar mãos fechadas ao nível do peito
Entrar no Modo de Lazer	Levar mãos à nuca e inclinar ligeiramente para trás
Terminar sessão	Levantar-se da cadeira e sair da frente do computador
Tirar screenshot	Dar “soco” com mão esquerda para a frente ao nível do peito
Saudar Jarbas	Acenar com mão direita ao nível da cabeça

Modo Trabalho

Acção	Gesto
Executa a ação definida para este gesto, pode ser modificada pelo utilizador (Default: Abre caixa de correio)	Abrir o braço esquerdo do peito para fora com a mão fechada
Executa a ação definida para este gesto, pode ser modificada pelo utilizador (Default: Abre o browser)	Abrir o braço direito do peito para fora com a mão fechada

Modo Lazer

Acção	Gesto
Executa a ação definida para este gesto, pode ser modificada pelo utilizador (Default: Abre player de musica)	Abrir o braço esquerdo do peito para fora com a mão fechada
Executa a ação definida para este gesto, pode ser modificada pelo utilizador (Default: Abre player de video/filmes)	Abrir o braço direito do peito para fora com a mão fechada

Fusion

Independente (Redundantes)

Acção	Gesto	Comando de Voz
Tirar screenshot	Dar “soco” com mão esquerda para a frente ao nível do peito	<ol style="list-style-type: none"> 1. “Jarbas, tira foto ao ecrã” 2. “Jarbas, tira/tira-me um printscreen”

Terminar sessão	Levantar-se da cadeira e sair da frente do computador	<ol style="list-style-type: none"> 1. “Jarbas, vou sair, bloqueia-me a sessão” 2. “Jarbas, bloqueia a sessão”
-----------------	---	---

Modo Trabalho (Complementares)

Acção	Gesto	Comando de Voz
Define a ação dada pelo comando de voz para este gesto quando no modo de trabalho	Abrir o braço esquerdo do peito para fora com a mão fechada	*Ditar qualquer comando de voz que queiramos associar a este gesto
Define a ação dada pelo comando de voz para este gesto quando no modo de trabalho	Abrir o braço direito do peito para fora com a mão fechada	*Ditar qualquer comando de voz que queiramos associar a este gesto

Modo Lazer (Complementares)

Acção	Gesto	Comando de Voz
Define a ação dada pelo comando de voz para este gesto quando no modo de lazer	Abrir o braço esquerdo do peito para fora com a mão fechada	*Ditar qualquer comando de voz que queiramos associar a este gesto
Define a ação dada pelo comando de voz para este gesto quando no modo de lazer	Abrir o braço direito do peito para fora com a mão fechada	*Ditar qualquer comando de voz que queiramos associar a este gesto

Cenário de utilização

Exemplo (Usado na apresentação)

O João liga o seu computador, no qual o Jarbas já está instalado e pronto a usar.

O João inicia a sua interação com o Jarbas dizendo “Olá Jarbas” ou qualquer outro tipo de comando de voz de saudação.

O “Jarbas” apresenta-se e contextualiza João acerca de qual o seu propósito e dá directrizes de utilização.

Por defeito o João está a usar o modo de trabalho do Jarbas.

O João faz o gesto de swipe para a esquerda e é aberto o browser na pagina elearning.ua para que ele possa começar a estudar.

O João quer tirar um screenshot de um slide importante e faz o gesto de screenshot e diz “Jarbas, tira-me um print ao ecrã”, a imagem é então guardada no ambiente de trabalho.

Sendo que o João usa bastante a calculadora quando trabalha ele decide configurar o gesto swipe esquerda para lhe dar acesso imediato à calculadora.

Para isso ele faz o gesto e de seguida diz o comando de voz que lhe dá acesso à calculadora. O Jarbas confirma que ele acaba de definir um novo comando associado a esse gesto e está pronto a ser utilizado.

O João acaba de estudar e quer descontrair um pouco portanto ele entra no modo de lazer fazendo o gesto correspondente e são dadas mais uma vez directrizes por parte do Jarbas sobre a que ação corresponde a cada gesto.

O João quer ouvir música portanto ele faz swipe right para abrir o seu player de música.

Ele adora ouvir música por isso vai configurar o Jarbas para que possa ser alterada a música com gestos para ele poder relaxar por completo sem ter de tocar mais no PC.

O João faz então o gesto swipe para a direita e diz “Jarbas, não gosto dessa música” a partir deste momento sempre que o João fizer swipe para a direita ou esquerda ele consegue navegar de música para música.

Já é tarde e o João está cansado por isso ele levanta-se e o Jarbas automaticamente bloqueia a sessão.

Instruções para correr a aplicação

1. Executar Fusion Engine RunTime
2. Executar IM RunTime
3. Executar Speech Modality
4. Executar Gesture Modality
5. Executar AppGui