# TGI

## Resumo

O desenvolvimento e a configuração de uma assistente virtual é um meio de proporcionar uma ferramenta de integração entre o sistema operacional e humanos, um cenário na qual pessoas que possuem mobilidade reduzida possam se beneficiar. Disponibilizar um leque maior de possibilidades a respeito de ferramentas de acessibilidade de software pode fazer com que o usuário decida qual delas seja melhor para sua individualidade. A existência dos recursos de controle do comportamento de janelas, abertura e fechamento de programas, além do gerenciamento de controle multimidia por meio da interação por voz são alguns dos benefícios a serem desfrutados deste software.

O objetivo deste projeto consiste na construção de uma assistente virtual para controlar o gerenciador de janelas i3WM no Linux através de comandos por voz. Para esta tarefa está sendo utilizado a linguagem de programação Python em conjunto com bibliotecas de código fonte aberto que sejam capazes de usufruir do sistema de controle de áudio do computador; reconhecer e transcrever a voz humana em texto; e, fazer o computador pronuncie palavras a partir de textos.

## Relatório - parte 2

Após definirmos que o objetivo do trabalho de graduação interdisciplinar consistiria na construção de uma assistente virtual para controlar o gerenciador de janelas i3WM no Linux através de comandos por voz, iniciamos uma pesquisa sobre as tecnologias existentes com o intuito de assimilar e distinguir o cenário em que pretenderíamos atuar.

A escolha de um nome apropriado e condizente com a descrição do projeto (A VAL - Voice Assistant for Linux -, ou em tradução livre, Assistente de Voz para Linux), tem nos direcionado a delimitar o escopo a ser trabalhado. Tal definição nos propícia flexibilidade para implementação de funcionalidades.

Ao decorrer do processo de pesquisa sobre as tecnologias necessárias para criação da assistente de voz, pôde-se identificar a existência de algumas bibliotecas de código que permitissem o reconhecimento e a transcrição da voz humana no computador. Escolher quais destas bibliotecas utilizar emplacaria tanto no licenciamento do projeto; no processo de desenvolvimento do código fonte, e sobretudo, no processo de criação da engenheira de software. Ao avaliarmos tais bibliotecas, escolhemos a Speech Recognition por conta da sua licença de uso e estar recebendo atualização dos desenvolvedores do projeto. Estes critérios foram fundamentais para exclusão das demais alternativas, tais como a Watson Speech to Text (IBM - Python) e o Sphinx4 (Java), além de direcionar a codificação do projeto para a linguagem Python.

De imediato, demos início ao desenvolvimento do projeto apenas em caráter experimental visando conhecer e explorar as possibilidades que poderíamos utilizar. Após um longo período de sofrimento conseguimos identificar as dependências necessárias para o funcionamento da VAL. A dificuldade mencionada deve-se a incompatibilidade de versionamento entre o PyAudio e o Speech Recognition em versões superiores a 3.6 do Python.

Utilizamos o sistema operacional ArchLinux como ambiente de testes e isto deve-se aos seguintes motivos: familiaridade de um dos integrantes do grupo ter com este sistema operacional; licença de uso do sistema operacional em questão; o sistema operacional possuir uma vasta documentação; e por fim, a organização do gerenciamento de pacotes de programas, sendo claramente dividido entre repositórios disponibilizados pelos mantenedores oficiais do projeto e os softwares disponibilizados pelos demais usuários, a Arch User Repository (AUR), a qual poderia facilitar nos processos de busca e instalação de dependências dos softwares.

Superar estes obstáculos e entender os motivos pelos quais diversas mensagens de erros surgiram durante todo este processo, nos condicionou a ter uma visão do comportamento do software. Identificamos possíveis limitações das bibliotecas quanto a detecção do reconhecimento de voz e também da pronúncia provinda a partir de texto no ambiente Linux ser um pouco robotizada. Todo este conhecimento adquirido nos ajudou a iniciarmos o processo de engenharia de software, onde estamos nos empenhando atualmente. Simplificar e organizar através de documentação é de fundamental importância para que possamos reiniciar o processo de dedicação de forma muito mais objetiva, uma vez que a VAL se mostrou capaz de abrir e fechar programas; focar em janelas na tela; mover o posicionamento das janelas; mover as janelas para diferentes áreas de trabalho; e, acessar diferentes áreas de trabalho.

## Short paper

### Lista de tarefas

- Falar sobre

- [x] Captura de áudio.

- [x] Inclusão digital.

-[X] Engenharia de Software

- [ ] Acessibilidade.

- [ ] Linux.

- [ ] Conceituar interface gráfica.

- [ ] Conceituar gerenciador de janela.

- [ ] Diferenciar interface gráfica vs gerenciador de janela.

- [ ] Interface gráfica i3 e gerenciador de janelas i3wm.

uso offline

Bibliotecas

Dificuldades enfrentadas

ASDF

<!--

Speech recognition module for Python, supporting several engines and APIs, online and offline.

https://github.com/Uberi/speech\_recognition

3 years ago

Watson Speech to Text

https://www.ibm.com/br-pt/cloud/watson-speech-to-text

Closed source

Sphinx4 - Java

https://github.com/cmusphinx/sphinx4

2 years ago

Offline Text To Speech synthesis for python

https://github.com/nateshmbhat/pyttsx3

11 months ago

-->

<!--

As tecnologias de reconhecimento de voz não são sistemas exclusivamente destinados à acessibilidade, mas tem servido com sucesso nessa área

(LIMA FILHO e WAECHTER, 2013)

LIMA FILHO, Marcos Antônio de; WAECHTER, Hans da Nóbrega. Tecnologias assistivas presentes no tablet e seu potencial para uma educação inclusiva para pessoas com deficiência visual. Revista Brasileira de Tradução Visual, Recife, n. 15, 2013.

-->

### Introdução

#### Captura de áudio

De acordo com Cardoso (2001), O reconhecimento de voz pode ser definido como a capacidade de um computador converter a sonoridade das palavras em código binário, o qual é compreensível pelo computador.

Partindo deste conceito, Alexandra et al. (2003) dizem que o som pode ser capturado a partir de um microfone simples no caso de um solução que não exige muita qualidade, ou através de uma mesa de som e amplificador para ambientes mais complexos possibilitando o balanceamento de sinal e ajuste de volume. Para a construção deste projeto nos concentramos em dispositivos de capturas de áudio simples, uma vez que tais recursos sejam mais acessíveis ao usuário comum perante ao custo de aquisição e o nível de complexidade de configuração.

Ao capturar o sinal de áudio para o computador, sua recepção digital frutifica um amplo panorama de possibilidades proporcionando inclusive a excentricidade de "conversar com a máquina". O valor de controlar o computador a partir de comandos de voz é deixar as mãos livres para outras funções, além de ser muito útil para propiciar maior acessibilidade às pessoas com necessidades especiais, pois oferece a oportunidade de usar o computador de uma forma que antes não podiam (SCOPE, s.d, apud DOS SANTOS, BERNAL, FERNANDEZ, 2018).

#### Acessibilidade

Em um levantamento realizado pelo Censo Demográfico de 2010 (IBGE, 2016), aproximadamente 23,9% da população se declaram com algum tipo de deficiência, seja visual, mental, motora ou auditiva, desta forma representando cerca de 46 milhões de brasileiros. Esta parcela populacional é diariamente afetada pelas barreiras de acessibilidade que permeiam o ambiente físico urbanista que vivemos e muitas vezes este cenário é estendido ao ambiente virtual por conta das limitações e, em particular a deficiência motora, cuja representatividade é de 6,95%, ou seja, aproximadamente 13 milhões de pessoas.

O Instituto de Tecnologia Social ITS (2008) apresentou a importância do desenvolvimento de ferramentas computacionais voltadas as necessidades das pessoas com deficiência. Uma vez que estas pessoas adquirem a independência da realização de tarefas computacionais básicas, como por exemplo, redigir textos e utilizar as redes sociais, elas sentem-se iguais aos demais usuários, aumentando assim sua autoestima através desta integração tecnológica.

#### Licença de uso

São muitas as maneiras pela qual um software pode ser licenciado, podendo ser essencialmente divido entre os grupos de softwares proprietários e os de softwares livres. E esta questão das licenças de uso é muito importante para o desenvolvido e a adoção de programa em caráter especial o software livre. Tais programas em geral são de simples obtenção, porém isto não significa que uma pessoa possa manipulá-lo da forma como queira (SABINO e KON, 2009).

A definição de código aberto não significa apenas o acesso ao código fonte havendo assim um conjunto de critérios conforme sintetizado a seguir:

1. Redistribuição livre

1. Inclusão do código conte.

1. Permitir modificações e trabalhos derivados.

1. Integridade do código fonte do autor.

1. Não haver discriminação às pessoas ou grupos.

1. Não haver discriminação a respeito das áreas comercialmente exploradas.

1. Os direitos associados ao programa devem ser aplicáveis a todos para quem o programa é redistribuído, sem a necessidade de execução de licenças adicionais para essas partes.

1. A licença de uso não deve ser específica a um produto, desta forma aplicado a mesma licença para suas partes e redistribuições.

1. A licença não deve restringir outro software.

1. A licença deve ser neutra as tecnologias ou estilo de interface.

<!--

(Trecho adaptado de SABINO, Vanessa; KON, Fabio. Licenças de software livre história e características. 2009).

-->

A licença GNU (acrônimo de GNU is Not Unix - GNU não é Unux, em tradução livre) foi criada por Richard Stallman em 1983 e mantido atualmente pela Free Software Foundation (FSF) e concede quatro liberdades aos seus usuários, sendo elas: liberdade para executar o programa como quiser; liberdade para realizar quantas cópias julgar necessário; liberdade de realizar modificações no programa a próprio gosto; e a liberdade para distribuir versões melhoradas (SABINO, 2011). Para Gouvêa e Jamil, 2002, Richard Stallman tomou esta iniciativa com o intuito de que programas desenvolvidos por comunidades de programadores ou universidades fossem impedidos de seres capturados por grandes empresas da informática.

#### Linux

Conforme os autores Ball e Buff (2004), Linux é o núcleo (ou kernel em inglês) de um sistema operacional gratuito, inspirado no Mini e que foi desenvolvido por Linus Benedict Torvalds no ano de 1991. Este é um programa executado no momento de inicialização do computador fornecendo uma interface entre o que é exibido ao usuário e o hardware.

O Linux está a sob a licença GNU

<!--

Alguns exemplos de distribuições não derivadas existentes são Red Hat3, Debian4, Slackware5. E alguns exemplos de distribuição derivada (que é baseada em um sistema operacional que já existe) são elas Fedora6, Ubuntu7.

3 Disponível em: <https://www.redhat.com/pt-br>. Acesso em: 12 set. 2015 4 Disponível em: <https://www.debian.org/index.pt.html>. Acesso em: 12 set. 2015 5 Disponível em: <http://www.slackware.com>. Acesso em: 12 set. 2015 6 Disponível em: <https://getfedora.org/pt\_BR>. Acesso em: 12 set. 2015 7 Disponível em: <http://www.ubuntu.com>. Acesso em: 12 set. 2015

-->

<!--

Segundo Ball e Duff (2004) existem grandes vantagens em se utilizar Linux.

Várias delas podem ser mencionadas. O Linux é um ótimo custo benefício, devido

ao fato de o Linux fornece um excelente retorno sobre o investimento, isto tudo

deve-se ao fato de o Linux ter pouco ou nenhum custo por máquina, além de não ter

taxas de royalties ou licenciamento, sendo assim uma única cópia de uma

distribuição Linux pode ser utilizada infinitamente. O Linux é um excelente Desktop,

também é uma excelente plataforma de Servidor.

BALL, B.; DUFF, H. Dominando Linux Red Hat e Fedora: Conhecimento, Soluções, Especialização. São Paulo: Makron Books. 2004.

-->

<!--

nitidez

precisão

perceptibilidade

panorama

captura do sinal

recepção digital

o processo de disponizar o áudio e até mesmo o vídeo seguem as etapas de captura do sinal (câmeras e microfones); recepção e compactação digital do sinal d

O processo de disponibilizar áudio e vídeo na internet pode ser dividido nas seguintes etapas: captura do sinal (câmeras, microfones); recepção e compactação digital do sinal de vídeo (codificador); transmissão do vídeo digital (servidor); digitalização de vídeos e videoconferência (ALEXANDRA et al., 2003).

O som pode ser capturado de um microfone simples, para uma solução que não exige muita qualidade, ou através de mesa de som e amplificador, que possibilite balanceamento de sinal e ajuste de volume (ALEXANDRA et al., 2003).

Partindo deste conceito, uma vez que o computador é capaz de obter ??? através de dispositivos de captura de áudio, tal como o microfone, e também de reproduzir sons através de dispositivos de saída de áudio, tal como o autofalante ou o fone de ouvido um imenso leque de possibilidades emerge-se, dentre eles, o de

O que fazemos é converter a voz humana em texto, realizar o tratamento da frase obtida para identificar e validar comandos. Uma vez validado, este comando será executado.

-->

Diante do exposto, este trabalho propõe uma assistente para controlar o gerenciador de janela i3wm no Linux através de comandos por voz. A base da pesquisa bibliográfica dá-se pela investigação dos seguinte termos: ???

<!--

comando de voz acessibilidade

deficiencia motora computador

linux

linux acessibilidade

mobilidade reduzida computação

virtual assistant

voice assistant

speech recognition

interface gráfica

i3wm

-->

### Projetos similares

Alexa

Corta

Jarvis

https://github.com/MichalKacprzak99/jarvis

introduce yourself

convert text to speech

handle microphone input

welcome you, different greeting depending on the time of day

tell a joke

say goodbye(stop working)

set timer

take note and read the last note

search in google

search in Wikipedia

tell a weather in given city

### Funcionalidades do projeto

O Projeto irá ter como principal funcionalidade o comando de voz que será emitido pelo usuário cliente, assim o sistema mudara a pedido dos padrões do usuário isso caso o sistema possa realizar tal ação. Assim ajudando varias pessoas que tem dificuldades de mexer no computado seja ela tenha algum tipo de deficiência mórbida ou não, é importante entender que o sistema será totalmente endereçado a essas pessoas.

### Engenharia de Software: Uma Amostra Bruto de Algo Realista

### Regras de Negócio: valos falar das mais importantes que o nosso sistema já possui

### História de Usuários: Algumas Histórias de Usuário Importantes

1. Gostaria que a assistente virtual reconhecesse meus comandos de voz:

Basicamente a essência do nosso sistema onde teremos o core, do onde o usuário usara maior parte dela para a execução de comando por voz, isso é o que faz o sistema ser o que ele é... Sua implementação é uma das mais complicadas e exige tempo e esforço para refinar ela e chegar ao estado perfeito e aceitável no final do projeto

1. Gostaria que a assistente virtual fosse capaz de controlar as janelas da interface gráfica:

Controle das janelas aos modos do usuário, onde através de seus comando por foz poderá personalizar as janelas sem precisar mover o mouse.

1. Gostaria que a assistente virtual interpretasse os comandos de voz apenas quando ela for chamada:

Famoso reconhecimento de voz que nossa interface terá assim para obedecer a um mestre especifico quando for chamada e assim executar as necessidades dele como personalização de janelas, funções como abrir o spotifi ou misics ou simplesmente abrir seu firefox

1. Gostaria que a assistente virtual me informasse caso o comando seja inválido:

Informação que daremos caso o usuário peça coisas impossíveis que o sistema não pode atender como instalar o Windows... deixando a brincadeira de lado esse requisito basicamente vai informá-lo caso ele execute algo que não configuramos na interface

### Requisitos Funcionais

1. O usuário requisitará uma funcionalidade ao sistema: como explicado acima o importante e estabelecer comunicação entre humano e maquina, então esse requisito é extremamente essencial já que ele tem sua descrição de estabelecer a conexão de máquina e usuário fazendo assim a maquina ser submetida aos pedidos de seu usuário.
2. O sistema deverá gravar a voz do usuário: é super importante que a maquina se lembre da voz de seu mestre para assim ela poder dar o GEP de tipo, esse é o mesmo usuário e eu o reconheço então eu irei atender todas as suas requisições (essa metáfora pode ser meio besta mas faz total sentido para o entendimento de algo tão complexo, Sommerville já dizia que é muito mais preferível uma metáfora bobinha do que um processo esdrúxulo sem sentido).

### Diagrama de Caso de Uso



001 – Gerenciamento de Funcionalidades

Aqui temos o caso de uso que verifica o pedido de seu cliente e o manda para o sistema depois de checado, esse caso irá analisar o comando executado pelo usuário e caso ele não esteja memorizado é por que ele precisa ser validado e é aqui que iremos validar ele ou não...

002 – Comandos e Memorização

Uma etapa que prioriza a agilidade dentro do sistema onde iremos checar se o usuário já pediu ou não essa função, se ele já pediu não precisaremos pedir ao sistema sua checagem esse comando já está dentro das normas. Ou seja ele será executado de uma forma ou de outra.

003 – Pedidos e Confirmações

Aqui temos o clássico caso de uso da confirmação ou da ***NEGAÇÃO,*** assim esse caso é super importante para realizarmos se o usuário pediu um comando que nossa agente possa atender ou não, assim retornando respostas como dativas a ele como tipo: não posso fazer isso aminho tente outra coisa. Ou simplesmente retornando o que ele pediu fazendo as alterações caso seja positiva a resposta pelo lado do sistema.

### Bibliografia

CARDOSO, S. H. Tecnologia de reconhecimento de voz. 2001. Disponível em: <http://www.edumed.org.br/cursos/biblioteca/recognition1.html>. Acesso em: 2 maio 2016.

ALEXANDRA Ap. Marcelino Toscaro, CLÁUDIO Martinez, LUCIANA Meneghel, MARCELO Araújo, FRANCO (Coordenador do projeto e organizador da publicação), RENATA A. Fonseca del Castillo, ROANDER Scherrer, RUBENS Queiroz de Almeida (Coordenador da DSC-CCUEC). Preparação de Conteúdo para WEB. 2003. EAD- Unicamp. 62f.

SCOPE

IBGE. Censo Demográfico 2010 - Características gerais da população, religião e pessoas com deficiência - Características urbanísticas do entorno dos domicílios. Rio de Janeiro: cap. 1.3. Instituto Brasileiro de Geografia e Estatística - IBGE, 2012.

ITS. Tecnologia Assistiva nas Escolas - Recursos básicos de acessibilidades ́sociodigital para pessoas com deficiência. São Paulo: cap. 1 e 2. Instituto de Tecnologia Social (ITS Brasil), 2008.

SABINO, V. C. Um estudo sistemático de licenças de software livre. Dissertação apresentada ao instituto de matemática e estatística da Universidade de São Paulo (USP). São Paulo, 2011.

SABINO, Vanessa; KON, Fabio. Licenças de software livre história e características. 2009.

GOUVÊA, Bernardo Andrade; JAMIL, George Leal. Linux: Uma análise histórica mercadológica e funcional de um sistema operacional. Revista Pretexto, 2002.

DOS SANTOS, Carlos Roberto; BERNAL, Janaína Cristina Galvez; FERNANDEZ, Breno Ortega. AUTOMAÇÃO RESIDENCIAL POR COMANDO DE VOZ PARA VIABILIZAR A ACESSIBILIDADE DE PESSOAS COM DEFICIÊNCIA. Revista Estudos & Pesquisas Unilins, v. 1, n. 1, p. 49-55, 2018.