

Universidade Cruzeiro do Sul

Henrik Beck

Peterson Garcia

Victor Henrique Ranalli Barbosa

Construção de uma assistente virtual para controlar o gerenciador de janelas i3WM no
Linux através de comandos por voz

São Paulo

2021

Henrik Beck
Victor Henrique Ranalli Barbosa
Peterson Garcia

Construção de uma assistente virtual para controlar o gerenciador de janelas i3WM no
Linux através de comandos por voz

Trabalho de graduação
interdisciplinar (parte 1) destinado a
Universidade Cruzeiro do Sul como parte
das exigências para a obtenção do título de
bacharel em Ciências da Computação.

Orientador Profº. Manuel F.
Paradela Ledón

São Paulo
2021

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 multimídia 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.

SUMÁRIO

1 INTRODUÇÃO.....	5
1.1 CAPTURA DE ÁUDIO.....	5
1.2 ACESSIBILIDADE.....	5
1.3 LICENÇA DE USO.....	5
1.4 LINUX.....	6
1.5 INTERFACE GRÁFICA.....	6
1.5.1 Conceito.....	6
1.5.2 Gerenciador de janelas i3wm.....	6
1.6 BIBLIOTECAS DE RECONHECIMENTO DE ÁUDIO.....	7
1.6.1 SpeechRecognition.....	7
1.6.2 Sphinx4.....	7
1.6.3 Watson Speech to Text.....	7
1.6.4 Critérios de escolha da biblioteca de reconhecimento de áudio.....	7
2 DESENVOLVIMENTO DO PROJETO.....	8
2.1 FUNCIONALIDADES DO PROJETO.....	8
2.2 ENGENHARIA DE SOFTWARE.....	8
2.2.1 Histórias de usuários.....	8
2.2.2 Regras de negócio.....	8
2.2.3 Requisitos funcionais.....	8
2.2.4 Requisitos não funcionais.....	8
2.2.5 Diagrama de caso de uso.....	9
3 APÊNDICE.....	10
3.1 PRIORIDADES.....	10
3.2 CATEGORIAS.....	10
3.3 HISTÓRIAS DE USUÁRIOS.....	11
3.4 REGRAS DE NEGÓCIO.....	12
3.5 REQUISITOS FUNCIONAIS.....	13
3.6 REQUISITOS NÃO FUNCIONAIS.....	14
4 REFERÊNCIAS BIBLIOGRÁFICAS.....	15

1 INTRODUÇÃO

1.1 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).

1.2 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.

1.3 LICENÇA DE USO

São muitas as maneiras pela qual um software pode ser licenciado, podendo ser essencialmente dividido 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).

Para 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
2. Inclusão do código fonte.
3. Permitir modificações e trabalhos derivados.
4. Integridade do código fonte do autor.
5. Não haver discriminação às pessoas ou grupos.
6. Não haver discriminação a respeito das áreas comercialmente exploradas.

7. 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.
8. A licença de uso não deve ser específica a um produto, desta forma aplicado a mesma licença para suas partes e redistribuições.
9. A licença não deve restringir outro software.
10. A licença deve ser neutra as tecnologias ou estilo de interface.

A licença GNU (acrônimo de *GNU is Not Unix* - GNU não é Unix, 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 serem capturados por grandes empresas da informática.

1.4 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. Este projeto está sob licença GNU, sendo assim muitas vezes chamado de GNU/Linux. Há diversos benefícios em utilizar o Linux, tais como o ótimo custo benefício, pois pode ter pouco ou nenhum custo por máquina e ser isento de taxas de royalties ou licenciamento. É um excelente uso tanto ambientes *desktop* quando servidores.

Alguns dos exemplos de distribuições Linux existentes segundo Campos (2006), são o *Debian*, o *Fedora*, o *Gentoo*, o *Mandriva*, o *Red Hat Enterprise Linux*, o *Slackware* o *Ubuntu*, dentre outros. Contudo a distribuição *ArchLinux*, sob a óptica de Castro (2016), tem por objetivo alcançar os usuários avançado e tem um sistema de atualização nos molde de *rolling release* e seu sucesso está atrelado a sua simplicidade, pois trás por padrão um alicerce básico para que o usuário construa o sistema do jeito que preferir.

1.5 INTERFACE GRÁFICA

1.5.1 Conceito

“O computador e sua interface representam uma ferramenta cognitiva, uma extensão da memória, uma prótese cognitiva que permite tratar melhor a informação. É importante que se conheça como os processos cognitivos humanos se desenvolvem para a concepção de próteses cognitivas compatíveis com eles.”
(Cybis, 2003 P.3).

1.5.2 Gerenciador de janelas i3wm

O i3WM é um gerenciador de janelas que estende as funcionalidades do WMII e automaticamente posiciona as janelas organizando-as em um layout de árvore, na qual criam-se dois

frames e divide-os horizontalmente e verticalmente. Cada *frame* contém uma janela ou um outro para de *frames*. O layout de árvore pode ser redimensionado de acordo com as preferências do usuário, alterando assim o tamanho das janelas e cada área de trabalho virtual contém seus próprios *frames* e janelas (Bradley, 2018).

O mecanismo do i3 (o i3WM) pode ser controlado através de um *socket* UNIX e realiza a comunicação pelo utilitário *i3-msg*. Esta interface pode enviar qualquer comando disponível no arquivo de configuração do i3 para utilizar teclas de atalho (STAPELBERG, 2012).

1.6 BIBLIOTECAS DE RECONHECIMENTO DE ÁUDIO

1.6.1 SpeechRecognition

Uma biblioteca capaz de reconhecer discurso, com suporte a *APIs* online e offline para a linguagem *Python*. É uma biblioteca que está sob licença BSD e tem seu código fonte aberto (SpeechRecognition, 2019).

1.6.2 Sphinx4

Escrita completamente utilizando a linguagem de programação Java, o Sphinx-4 pode ser executado em uma variedade de plataformas sem requerir nenhuma compilação especial. É uma biblioteca que está sob licença BSD e tem seu código fonte aberto (SPHINX4, 2014).

1.6.3 Watson Speech to Text

Com o Watson Speech to Text (STT) é possível converter áudio em texto em mais de 15 idiomas incluindo Português, Inglês, Espanhol e Mandarim (IBM, 2021).

1.6.4 Critérios de escolha da biblioteca de reconhecimento de áudio

A Tabela 01 compara as biblioteca de reconhecimento de áudio.

Biblioteca	Suporte a idiomas inglês e português	Reconhecimento <i>online</i>	Reconhecimento <i>offline</i>	API gratuita
SpeechRecognition	Sim	Sim	Sim	Sim
Sphinx-4	???	???	Sim	Sim
Watson Speech to Text	Sim	Sim	Não	Até 500 requisições por dia

Tabela 01

Diante de tal comparação, optamos em utilizar a biblioteca SpeechRecognition para construção deste projeto.

2 DESENVOLVIMENTO DO PROJETO

2.1 FUNCIONALIDADES DO PROJETO

O projeto desenvolvido por este trabalho tem como principal funcionalidade o comando de voz para controlar o gerenciador de janelas I3WM da interface gráfica I3. Sendo o sistema será capaz de:

1. Registrar arquivos de *logs*.
2. Gravar a voz do usuário.
3. Pronunciar palavras/frases em inglês.
4. Pronunciar palavras/frases em português.
5. Reconhecer a voz do usuário em ambiente de conexão offline.
6. Reconhecer a voz do usuário no idioma inglês.
7. Reconhecer a voz do usuário no idioma português.
8. Controlar as janelas gráficas do gerenciador de janelas I3WM.
9. Abrir softwares instalados no sistema operacional.
10. Fechar softwares instalados no sistema operacional.

2.2 ENGENHARIA DE SOFTWARE

2.2.1 Histórias de usuários

Ver apêndices 3.3

2.2.2 Regras de negócio

Ver apêndice 3.4

2.2.3 Requisitos funcionais

Ver apêndice 3.5

2.2.4 Requisitos não funcionais

Ver apêndice 3.6

2.2.5 Diagrama de caso de uso

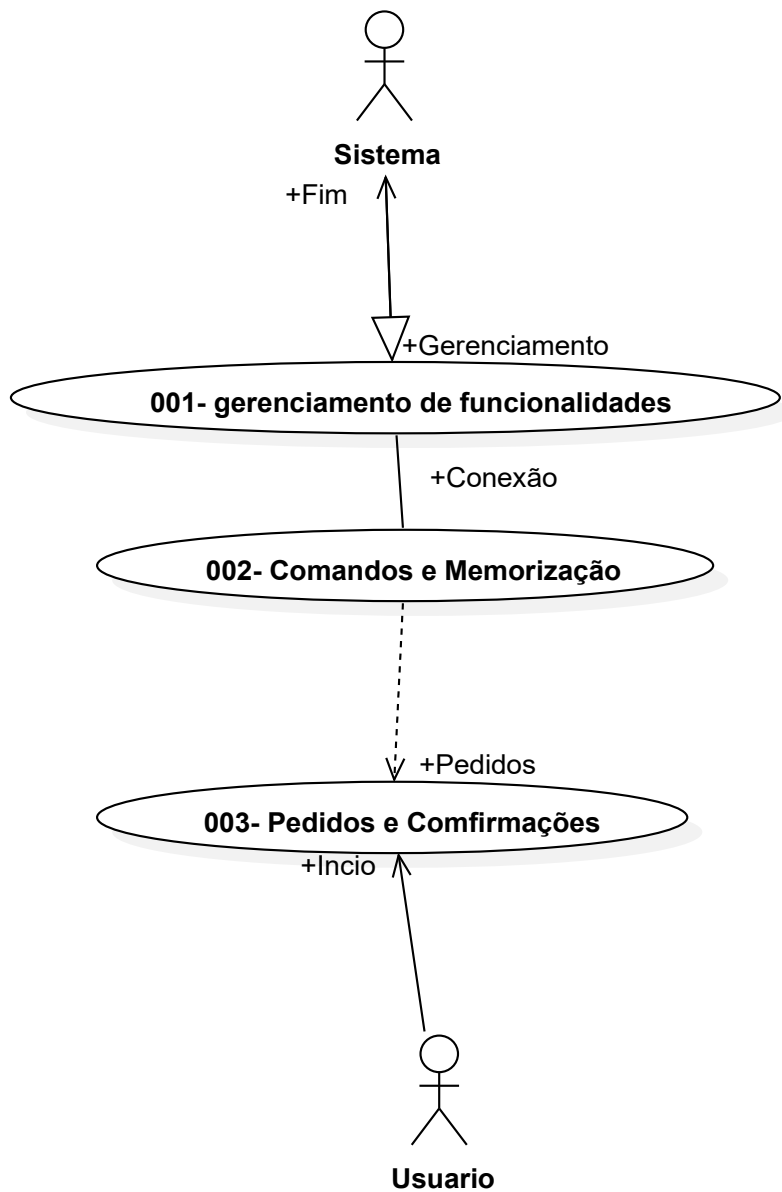


Figura 1

- **001 – Gerenciamento de funcionalidades**
Trata-se do caso de uso que verifica o pedido de seu cliente e aciona o sistema assim que verificado a sua integridade. Este caso de uso analisa o comando executado pelo usuário e em virtude da não existência do comando solicitado na memória, ele retorna falso.
- **002 – Comandos e memorização**
Etapa na qual prioriza-se a agilidade dentro do sistema onde é verificado se o usuário já solicitou ou não esta 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 este caso é refere-se a realização caso o usuário solicitou um comando que a assistente virtual possa atender ou não. Desta forma retornam-se as respostas como dativas a ele, como por exemplo: “Este comando é inválido. Tente novamente”. Ou, em caso afirmativa, retorna-se a execução do comando solicitado.

3 APÊNDICE

3.1 PRIORIDADES

Valor	Prioridade Descrição
1	Nem um pouco importante
2	Ligeiramente importante
3	Importante
4	Muito importante
5	Extremamente importante

3.2 CATEGORIAS

Valor	Categorias Nome
1	Interação com o sistema
2	Funcionalidades
3	Reconhecimento de voz
4	Fala do sistema
5	Registro do sistema
6	Configurações do sistema

3.3 HISTÓRIAS DE USUÁRIOS

Histórias de usuários					
ID	Persona	Planning poker	Prioridad e	Categoria s	Descrição
		Extremamente			
HU-01	#UB	difícil	5	1	Gostaria que a assistente virtual reconhecesse meus comandos de voz
HU-02	#S	Difícil	1	3	Gostaria que a assistente virtual compreendesse o idioma inglês
HU-03	#S	Difícil	4	3	Gostaria que a assistente virtual compreendesse o idioma português
HU-04	#S	Difícil	2	2	Gostaria que a assistente virtual fosse capaz de abrir programas do computador
HU-05	#S	Fácil	3	2	Gostaria que a assistente virtual fosse capaz de fechar programas do computador
HU-06	#S	Normal	3	5	Gostaria que a assistente virtual registrasse suas ações em um arquivo de log
HU-07	#S	Normal	4	2	Gostaria que a assistente virtual fosse capaz de controlar as janelas da interface gráfica
HU-08	#S	Difícil	3	2	Gostaria que a assistente virtual fosse capaz de controlar o nível de luminosidade da tela do computador
HU-09	#S	Difícil	3	2	Gostaria que a assistente virtual fosse capaz de controlar o volume da saída de áudio do sistema operacional
HU-10	#S	Normal	5	1	Gostaria que a assistente virtual interpretasse os comandos de voz apenas quando ela for chamada
HU-11	#S	Fácil	2	6	Gostaria que a assistente virtual pudesse carregar as informações de sua configuração a partir de um arquivo externo
HU-12	#S	Normal	3	4	Gostaria que a assistente virtual me informasse caso o comando seja inválido
HU-13	#UB	Difícil	2	1	Gostaria de personalizar os comandos da assistente virtual
HU-14	#S	Normal	3	5	Gostaria que a assistente virtual gravasse a solicitação do último comando de voz
HU-15	#S	Difícil	4	3	Gostaria que a assistente virtual compreendesse os comandos mesmo quando offline
HU-16	#S	Normal	2	6	Gostaria que a assistente virtual não solicitasse permissões de privacidade do usuário

3.4 REGRAS DE NEGÓCIO

ID	Regras de negócio Descrição
RN- 01	Usuário deverá ter acesso ao sistema sem a necessidade de privilégios de administrador
RN- 02	O usuário não precisará ter permissão de administrador do sistema para executar a aplicação O usuário deverá editar o arquivo de texto correspondente as configurações do sistema caso deseje realizar
RN- 03	customizá-lo
RN- 04	O sistema não disponibilizará interface gráfica ao usuário
RN- 05	O usuário deverá ser informado caso um comando seja invalidado pelo sistema
RN- 06	O usuário deverá estar ciente que a utilização do software será de sua inteira responsabilidade
RN- 07	Os comandos validados pela assistente virtual deverão ser executadas através do terminal Linux
RN- 08	Deverá ser reconhecido pelo sistema comandos no idioma inglês
RN- 09	Deverá ser reconhecido pelo sistema comandos no idioma português
RN- 10	A assistente virtual deverá ser possibilitada de customização
RN- 11	A assistente virtual deverá reconhecer comandos de voz emitidos pelo usuário
RN- 12	A assistente virtual só estará em execução de segundo plano
RN- 13	A assistente virtual deverá ser ativada apenas quando chamada através de ser codinome (padrão: VAL)
RN- 14	Deverá ser mantido pelo usuário a gravação apenas do seu último comando
RN- 15	Deverá ser criado pela assistente virtual um arquivo de log de funcionamento do sistema
RN- 16	Deverá ser carregado as configurações do sistema através de um arquivo externo
RN- 17	Deverá ser reconhecido os comandos de voz mesmo em ambiente de conexão offline

3.5 REQUISITOS FUNCIONAIS

Requisitos Funcionais		Histórias de	Regras de
ID	Descrição	usuários	negócio
RF-01	O usuário requisitará uma funcionalidade ao sistema	HU-02	RN- 13
RF-02	O sistema deverá registrar arquivos de logs	HU-07	RN- 15
RF-03	O sistema deverá gravar a voz do usuário	HU-15	RN- 14
RF-04	O sistema deverá permitir apenas o uso de funcionalidades disponíveis	HU-13	RN- 05
RF-05	O sistema controlará as janelas da interface gráfica	HU-08	RN- 07
RF-06	O sistema deverá reconhecer a voz do usuário offline	HU-16	RN- 17
RF-07	As bibliotecas requeridas para o funcionamento do sistema deverão armazenar os idioma inglês	HU-03	RN- 08
RF-08	As bibliotecas requeridas para o funcionamento do sistema deverão armazenar os idioma português	HU-04	RN- 09
RF-09	O sistema deverá carregar suas configurações através de um arquivo externo de notações de objetos	HU-12	RN- 16
RF-10	O sistema deverá permitir que o usuário crie e carregue configurações personalizadas	HU-14	RN- 03

3.6 REQUISITOS NÃO FUNCIONAIS

Requisitos não funcionais				
ID	Descrição	Regras de negócio	Histórias de usuários	
RNF – 01	Apenas a última gravação da voz do usuário deverá ser mantida no sistema	RN- 14	HU-15	
RNF – 02	A biblioteca do idioma inglês será disponibilizada	RN- 08	HU-03	
RNF – 03	A biblioteca do idioma português será disponibilizada	RN- 09	HU-04	
RNF – 04	Não haverá permissões de privacidade deverão a serem solicitadas	RN- 01	HU-16	
RNF – 05	A assistente de voz deverá informar o usuário caso o comando solicitado seja inválido através de emissão sonora por meio de sua própria fala	RN- 05	HU-13	
RNF – 06	A assistente de voz deverá informar o usuário caso o comando solicitado seja inválido através de texto por meio da notificação do sistema	RN- 05	HU-13	

4 REFERÊNCIAS BIBLIOGRÁFICAS

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.

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

BRADLEY, Evan. An Infinite-Pane, Zooming User Interface Window Manager and Survey of X Window Managers. 2018. Tese de Doutorado.

CAMPOS, Augusto. O que é uma distribuição Linux. **BR-Linux. Florianópolis**, 2006.

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.

CASTRO, Jose Dieguez. Arch linux. In: **Introducing Linux Distros**. Apress, Berkeley, CA, 2016. p. 235-252.

CYBIS, Walter de Abreu. A identificação dos objetos de interfaces home-computador em seus atributos ergonômicos. Tese de Doutorado em Engenharia de Produção. Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis, 1994.

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.

IBM; 2021. Disponível em: <https://www.ibm.com/br-pt/cloud/watson-speech-to-text?utm_content=SRCWW&p1=Search&p4=43700052743482493&p5=e&gclid=CjwKCAiA-9uNBhBTEiwAN3iINFIcmtI-B9kCiVIYsedE4ngg7iTsWoLIuEEerpiYT1S41DIzSljJ-hoCnIkQAvD_BwE&gclsrc=aw.ds>. Acesso em: 13 dezembro 2021.

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.

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

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.

SPEECHRECOGNITION; 2019. Disponível em:
<https://github.com/Uberi/speech_recognition>. Acesso em: 13 dezembro 2021.

SPHINX4; 2014. Disponível em: <<https://github.com/cmusphinx/sphinx4>>. Acesso em: 13 dezembro 2021.

STAPELBERG, Michael. i3-msg(1). <https://build.i3wm.org/docs/i3-msg.html>. 2012.