Remote desktop access from cell phones “Desktop Anywhere”

By: Bassant Hossam, Hossam Hatem, Andrew Samir, Sara Mazhar, Rehab Youssef

Supervised by: Prof. Dr. Ahmed Salah , TA. Zeina Rayan

Faculty of Computer and Information Sciences - Ain Shams University

Abstract

The rapid advancement of technology has significantly impacted our lives, work, and interactions. It has significantly shaped human civilization, making repetitive tasks easier in terms of time, effort, or both.

Desktop device users face a variety of obstacles while using desktop computers, including Limited Physical Access as Users face the challenge of being physically present at their desktop computers to access or control them, inefficient File and Folder Retrieval in terms of time especially where huge data exists, time-consuming data management across many PCs, also users with disabilities may have difficulty typing, while others may struggle with manual activities such as application launches or shutdowns. Furthermore, individuals who do not properly schedule their work risk becoming overwhelmed, lowering pleasure and productivity.

These obstacles highlight the need for innovative solutions to improve user experience and efficiency. Also, as a response and a contribution to the global orientation, Desktop Anywhere is here. It is a software which developed as a comprehensive and integrated system consisting of two interconnected components: a mobile application and a desktop application communicating through a server. Desktop Anywhere is developed to enable users to access and control multiple personal desktop computers remotely. It aims to provide a seamless experience for users to interact with their desktop screens, manage files, execute tasks through virtual touchpad and keyboard, and utilize voice commands in Arabic, which is done by more than 95% accuracy, all through their mobile devices which will have the ability to be automatically connected to all paired devices without authentication. The primary goal is to bridge the gap between devices and redefine remote interactions, enhancing convenience, efficiency, and accessibility in the realm of digital connectivity.

Introduction

"Desktop Anywhere“ system is a necessary and timely response to the challenges faced in the field of desktop computing. The need for improved accessibility, streamlined data management, inclusivity, and remote connectivity has motivated the idea behind this system. By addressing these needs, the project enhances user experience, improves productivity, and aligns with the ongoing advancements in remote work and digital collaboration.

Methods

In our software project, we have implemented several key modules. The project consists of the following modules: Server, Mobile App, Desktop App, Model Finetuning, and Dataset Creation.

Server

The server is responsible for handling communication between the mobile and desktop applications. It was developed using NodeJS and ExpressJS to manage real-time data transmission and ensure seamless interaction between client applications. The server acts as a central hub, facilitating the exchange of data and commands between the mobile and desktop environments.

Mobile App

The mobile app provides a user-friendly interface for voice command entry and real-time interactions. It was developed using the Dart programming language and the Flutter framework. For real-time view features, we integrated WebRTC, an open-source technology that enables the transmission of sound, video, and data over the internet. This integration allows for real-time communication across different devices, enhancing the user experience.

Desktop App

The desktop app allows users to interact with the system via voice commands on a desktop environment. It was developed using Python and communicates with the server to execute commands and retrieve data. The desktop app provides a seamless and efficient way for users to manage their tasks through voice commands. Additionally, users can remotely access and control the desktop or several desktops using the mobile application, offering comprehensive remote management capabilities.

Model Finetuning

We utilized two AI models for voice command understanding: Whisper and AraBERT. Whisper is a transformer-based model designed for speech-to-text conversion, while AraBERT is used for classification and Named Entity Recognition (NER) tasks. The process begins when a user initiates a voice command using the mobile app. The audio input is processed by the Whisper model, which generates a textual transcription of the user's voice command. The transcribed text is then classified using the AraBERT model to determine the type of command issued by the user. Possible command categories include search, set, add note, close, restart, and sleep.

Once the command type is identified, AraBERT performs NER to extract useful information from the command, such as file names, folder names, time, and other relevant entities. After extracting these entities and understanding the command context, the appropriate action is executed on the desktop app.

Dataset Creation

To finetune the Whisper and AraBERT models, we created a customized dataset in two forms: text and voice records. The dataset covers the six main commands allowed in our system for users to ask verbally. The steps in the dataset creation were as follows:

1. **Text Data Creation**: We created 1872 rows of text data.
2. **Voice Record Collection**: We found volunteers who each recorded an average of 20 sentences, resulting in 1932 voice records. The difference in the number of records is due to the need for balance in the text dataset, which affects classification and NER tasks.

This comprehensive dataset was then used for finetuning the models.

Results

Expected Results:

1. Collecting Voice Dataset for voice command to be executed on desktop that consists of Arabic and English command.
2. Collecting Dataset for Multi class classification in Arabic.
3. Collecting Dataset for NER Arabic.
4. Chose suitable models and Finetune them for the following tasks speech to text, classification and Name Entity Recognition to reach our goal which is getting accuracy 95% or higher.
5. Building Mobile app, Desktop app and server then deploy our server on free online server.
6. Building Command execution system.

Actual Results:

1. Voice Dataset collected.
2. Dataset for Multi class classification collected.
3. Dataset for NER Arabic collected.
4. Models for speech to text, classification and Name Entity Recognition has been chosen and finetuned depends on the following trials and previously mentioned features of each chosen model.
5. Mobile app, Desktop app and server built and the Server deployed.
6. Command execution system built.

Final results after a lot of trials for ai based part:

1. Speech to Text Task: we chose Whisper model for this task we finetuned it with 1932 record which met our targeted accuracy, and the results we got from final trail is displayed in table 1.
2. Results of finetuning Ara-BERT for Classification task mentioned in table 2.
3. Results of finetuning Ara-BERT for NER task mentioned in table 3.

	Customized Dataset
SIZE	1932
SPLITS	Train: 1544 Test: 387
WER Before Fine-Tuning	28 %
WER After Fine-Tuning	4.6%

Table 1

	AraBERT
Dataset	1872 row: Train = 1497 Test = 375
Accuracy Before Fine-Tuning	25 %
Accuracy After Fine-Tuning	100 %

Table 2

	AraBERT
Dataset	1872 row: Train = 1497 Test = 375
Accuracy Before Fine-Tuning	43 %
Accuracy After Fine-Tuning	98 %

Table 3

Conclusion

Eventually, in this project we tried to bridge the gap between physical presence and desktop accessibility. Through Desktop Anywhere, a comprehensive system designed for users to be user-friendly/ the mobile application empowers users to access and control their personal desktops from anywhere through the server. This remote access capability simplifies file management by allowing users to effortlessly manage and navigate through files and resources across multiple desktops concurrently.

Desktop Anywhere embraces inclusivity by providing accessibility features like voice commands (currently in Arabic) that cater to a wider range of users. For individuals who prefer hands-free interaction or those engaged in multitasking activities, Desktop Anywhere provides voice control and virtual input methods for executing tasks and controlling applications. Repetitive tasks like shutdowns or application launches can be automated, streamlining daily routines and boosting efficiency.

The ability to manage tasks effectively is crucial in today's fast-paced world. Desktop Anywhere addresses this need by offering task scheduling, allowing users to organize activities seamlessly across both desktop and mobile devices.

References

1. W. Antoun, F. Baly, and H. Hajj, “AraBERT: Transformer-based Model for Arabic Language Understanding,” arXiv:2003.00104 [cs], Mar. 2020, Available: https://arxiv.org/abs/2003.00104
2. WebRTC, “WebRTC Home | WebRTC,” Webrtc.org, 2017. https://webrtc.org/
3. Abdelhalim Hafedh Dahou, Mohamed Amine Cheragui, and A. Abdelali, “Performance Analysis of Arabic Pre-Trained Models on Named Entity Recognition Task,” Jan. 2023, doi: https://doi.org/10.26615/978-954-452-092-2_051.
4. A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision,” *arXiv:2212.04356 [cs, eess]*, Dec. 2022, Available: https://arxiv.org/abs/2212.04356