DECLARATION

We hereby declare that this project work entitled "**DIGITAL FILE MANIPULATION**" has been prepared by us during the year 2021-22 under the guidance of **Mrs. Roopa K** ,Department of Computer Science, Bhandarkars' Arts and Science College, Kundapur in the partial fulfilment of BCA degree prescribed by the Mangalore University.

We also declare that this project is the outcome of our own effort , that it has not submitted to any other university for the award of any degree.

Date:

Student Name: Sheethal Kumar Shetty

Reg no:191122730

Student Name: Rakshith Kumar Shetty

Reg no:191122854

Student Name: Vivek

Reg no:191122880

ACKNOWLEDGEMENT

We consider it privilege to express our profound gratitude and respect to those who guided us in the successful completion of this project.

We would also like to express our gratitude towards our projects guide Mrs. **Roopa K**, Lecturer in Department of Computer Science for his valuable guidance throughout the period of project.

We would like to express our gratitude to **Mrs. Vijayalakshmi N Shetty,** Head of the Department of Computer Science for her kind concern and encouragement during the completion of our project.

We are sincerely thankful to **Dr. N. P. Narayana Shetty,** Principal of our college. For providing the opportunity and facility to develop this project.

We are also thankful to the entire faculty member of our department, our parents and friends who helped directly or indirectly to do this project through their support, encouragement, and guidance.

Thanking You,

Project Team:

Sheethalkumarshetty

Rakshithkumarshetty

Vivek

Table of Contents

List of Figures

List of Tables

Conclusion

In conclusion, this project was successfully implemented using Python 3.10.4. It is capable of carrying out various file manipulations like converting pdf to text, pdf to docx, text to docx, docx to text, excel to text, csv to text, excel to csv, search columns in csv file. Image processing like extracting text from image, extracting text from region of interest and highlighting text in image. Audio-Video manipulation like extracting text from video, extracting text from audio and extracting audio from video.

During the implantation we have faced many challenges in reading text from image and its processing, converting audio to text and handling pdf files. We handled all the challenges successfully. And also handles the various exception during the development of project work.

We have learned about GUI of python, handling image for OCR and handling various files for conversions.

Moreover, this project helped us to understand Software Development life Cycle(SDLC.),Time bound work, team spirit and preparing project document, project testing, GUI designing and presentation.

In addition to that, we learned Python, coding and image processing tools like Pytesseract.

Finally concluded that we tried to fulfill the objectives of project work and goal of our project Digital File Manipulation.

Rakshith Kumar Shetty	191122854
Sheethal Kumar Shetty	191122730
Vivek	191122880

Limitations

- Image must be clear and font must be identifiable and not wavy.
- Text in the image must be in English.
- Audio size must be under 10MB.
- Video size must be under 20MB.
- Text must be English for text to audio conversion
- Internet connection must be good for Audio-Video manipulation.
- Pdf file must not be password protected.
- If a word or pdf contains image, then text from that image will not be retrived while pdf to text and word to text conversion.
- Large files takes more time (more than 5 Min) to convert.
- Only English text will be extracted from the image.

Scope for enhancement (future scope)

We can enhance the image processing system and add more functionalities to the system using image processing. Enhance the text recognition engine to get better result by using machine learning , to extract text from video and extract text from Pdf image

Abbreviations and Acronyms (list)

SRS- software Requirement Specification.

DFD- Data Flow Diagram.

CFD- Control Flow Diagram.

Bibliography/References (list in specified format)

Website referred:

1. https://tkinter-docs.readthedocs.io/en/latest/index.html
2. https://pypi.org/project/pytesseract/
3. https://pypi.org/account/register/
4. https://numpy.org/
5. https://www.geeksforgeeks.org/
6. https://ffmpeg.org/
7. https://opencv.org/
8. https://pypi.org/project/moviepy/
9. https://pypi.org/project/gTTS/
10. https://www.google.com/

Book referred:

- An integrated approach to software engineering -Pankaj jalote.