

Submitted by: Sai venkata sahith reddy,
Kadapaiahgari
Matriculation Number: 222201978.

Self reflection on "Training a tesseract module for GDR typewriter"

Working on this project with my team has been a great learning experience that helped me to enhance my technical skills and interpersonal skills.

Technical Growth

This research helped me to grasp more knowledge about Optical Character Recognition (OCR) using Tesseract, technology I had previously not worked with. I learned how to extract the text from handwritten documents, typewritten documents, etc., using tesseract. Through this experience, I was able to learn various things like installing tesseract, configuring make file and training the model using various commands. I have improved my ability to spot performance improvements in the text extraction. I developed a competence during this course of the project that allowed me to test the accuracy of the extraction and use the tesseract's ultimate features.

My Contribution

In this research,

- I am responsible for transforming the given input documents into required formats for training the model.
- Initially I used a custom python library '**pdf2images**' to convert the PDFs to TIFF.
- Later using '**pytesseract**' extracted text into a single line to create ground-truth.
- After that, actively collaborated with other teammates to train the model.
- Actively collaborated with my teammates in evaluating the model outcome.
- With my teammates help, Performed training again in my machine.
- In documentation, I took responsibility to write,
 - ◆ Methodology which consists of,
 - system setup
 - data collection
 - data pre processing
 - training the model
 - testing the model.
 - ◆ Conclusion

Teamwork and Collaboration

The chance to collaborate closely with my colleagues was very helpful for me. We succeeded in this project because of clear communication and teamwork. We followed with a plan by organising weekly stand-up meetings, which allowed for easy workflow and made sure that everyone was on the same page with the project's goals. I understood the value of knowledge and skill exchange within the team. In order to overcome obstacles, we used both our group's collective problem solving abilities and the interaction of our different skills

Submitted by: Sai venkata sahith reddy,
Kadapaiahgari
Matriculation Number: 222201978.

and backgrounds. It was encouraging to see how the many contributions made by the team members improved the project as a whole.

Conclusion

To summarise, this **Training a tesseract module for GDR typewriter** project has been a journey of technical and interpersonal development. I gained a deeper understanding of large data processing and data integrity, data transformation, Model training, and Model evaluation. I was also able to comprehend the value of common knowledge and develop good teamwork abilities thanks to the collaborative environment. As I think back on this project, I'm not only happy with how well it turned out, but I'm also appreciative of the teamwork and lessons we all gained along the way. With the knowledge and insights I gained from this experience, I will be better prepared to take on new challenges and support the success of any team I work on in the future.

I would like to express my heartfelt gratitude to Prof. Dr. Jens Dörpinghaus, for his unwavering guidance and support throughout our Training a Tesseract Module for GDR Typewriter project. His expertise and mentorship played an instrumental role in shaping our project's success. His contributions were invaluable, and I am truly appreciative of their assistance in making this project a rewarding and educational experience.

Best regards,
Sai venkata sahith reddy Kadapaiahgari