Submitted By: Anu Antony
Matriculation Number: 220200551

Self-Reflection on Research Lab Project "Training a Tesseract module for GDR typewriter"

The research lab project "Training a Tesseract module for GDR typewriter" started with a mixture of anxiety, excitement, and fear. The confusion and fear were because of how this research lab project could proceeded, and what all the hindrances had to be faced. On the other hand, the excitement was working and collaborating with a team of hardworking and enthusiastic colleagues and learning something new. Even though the research lab project started with anxiety and fear, eventually, because of our whole team's support, hard work, and patience, I was able to tackle all of it and was able to make a significant contribution to this research lab project's completion.

Technical Growth

The research lab project has provided me with a vast knowledge-gaining experience. My efforts not only helped me to understand Tesseract but also helped me comprehend other OCR platforms developed by different companies or research groups. Before commencing the actual implementation of this topic, I had to understand other modules related to this research lab project. I began with the OCR concept itself, acquired maximum knowledge, and laid a foundation before delving into the core concept. Learning about the internal functioning of the Tesseract model training, such as TIFF file creation and how LSTMF and box files are treated internally, was an exciting experience. During the research lab project implementation phase, I had to perform model training and testing. Refreshing the concepts that I had learned during my machine learning and data science courses helped me to complete the implementation phase easily. Since our research lab project consists of only a few Python scripts, I had to work with the Python programming language and a bit of shell programming as well. The training of Tesseract involved the installation and use of "*make*" commands, which was also new to me. During the training process, I had to delve deeply into *make* commands as well. Until then, I had only dealt with *Ant*, *Maven*, and *Gradle* build automation tools. With this research lab project, I also had to explore make commands and the workings of "*MakeFile*", which became a new addition to my skill set. Overall, I was able to gain expertise in the internal technicalities of the OCR technology, especially the Tesseract-OCR engine, and improved my programming abilities in Python as well.

Contributions

In our research lab project, my major contributions are

1. Conducted research and responsible for the training of Tesseract-OCR model, including the ground-truth data generation (tiff, box, lstmf and gt.txt files).
2. Created evaluation of our trained Tesseract-OCR model, by extensive researching and implementing matrices such as CER and WER.
3. Responsible for implementing the matrices for readability check in referencing results from ABBYY FineReader and OCR4all models.
4. Actively done the CER and WER metrics evaluation on the referencing ABBYY FineReader and OCR4all model results
5. Conducted comparison of our trained model results with ABBYY FineReader and OCR4all model results.

Submitted By: Anu Antony
Matriculation Number: 220200551

Also, I have made a minor contribution to

6. Research Literature search and research lab project report documentation, especially in the proof reading and giving update suggestions.

Teamwork and Collaboration

Working with my colleagues was quite a happy time. We have followed almost an agile mode of working. I believe, that to make a healthy team and collaborative environment, communication is a must thing and there should not be any communication gaps. We used all the tools that we could use for better team management in this research lab. We all approached this research lab in a real professional way. We used the university's MS Teams application for all our meetings. We had our dedicated channel for conducting meetings and sharing information. We had conducted meetings at least once every week during our research lab tenure. This helped effective collaboration and knowledge sharing. Most of the time we conducted online live working sessions. This method is applied to reduce time wasting and effective problem-solving. Also, this enabled flexibility among us and helped to get a feel of collective working. Apart from that, we started a WhatsApp group for any time personal and group communications. This is to make sure that, each update is reached on time among our members irrespective of their status at that time. This was also a viable solution approach to make quick decisions when required. Throughout our research lab period, we built and maintained a healthy team atmosphere, since each one is from a different background, we had a good time exchanging skills and knowledge, which helped significantly improve our overall productivity. This personally helped to solve some of my blocks during the development phase.

Conclusion

To summarize, the research lab "Training a Tesseract module for GDR typewriter" has been quite an enthusiastic journey of both technical, interpersonal, and team development. I can say that I was able to make some remarkable contributions to this research lab project. During this research lab phase, I gained a deeper knowledge of the vast OCR topic. The knowledge that gained especially of Tesseract OCR and the internal functioning of Tesseract OCR helped me to understand the depth of this technology. The experience that I gained during this period molded me into a position to take on all the upcoming challenging tasks.

Furthermore, I would like to convey my sincere thanks to Prof. Dr. Jens Dörpinghaus for his constant direction and assistance throughout our research lab on the topic of "Tesseract Module Training for the GDR Typewriter." I really value all his assistance in making this research lab project a satisfying and learning opportunity.

Regards,
Anu Antony