

Paul Trust

PhD Student(Cohort 2)

**Supervisors: Dr.Rosane Minghim and Dr.Ahmed
Zahran**

Project Title

Novel Methods for User Centered Analytics of Documents

Project Summary

The world is being faced with massive explosion of information like never due to the advent of internet and technology, which has made it possible to have digital libraries, scanned documents, social media platforms where millions of messages are shared every day and several websites with all sorts of information from education to advertising. Most of this information comes in unstructured form as text and images, which further results in bottlenecks to their analysis. Recent adoption of neural models for documents has made significant progress in automated document analysis. However, computational methods alone cannot uncover the insights that lie in these documents fully. There is thus a need to augment computational methods with the user's knowledge. The project involves developing algorithms and strategies that leverages on both computational methods and expert's knowledge to analyze these document collections.

Potential Output and Impact

We intend to develop strategies and algorithms where the user has influence over the modelling process through interactive visualizations. This will increase trust in model results and further the adoption of such models. Potential applications include feedback analysis, analysis of historical documents, document summarization, support to job selection processes, among others.