**Project title:** Querying semi-structured data

**Student:** Ben Ramsay Foster 1705054

**Supervisor:** Michael Fairbank

**Abstract**

A program to automate the process of merging multiple Excel tables that contain overlapping sets of data, this ranges from tables on different spreadsheet pages to multiple files being merged into one. The main goal of the project was on automating the process so that a user of any skill could simple import in the correct files and have a result returned to them. Of course, errors can occur, so the user is able to override the automated decisions made while still being able to merge tables more efficiently. To help with the merging process the program can automatically check for inconsistencies in the data to help with large tables where it would be difficult to spot a single mistake or inconsistency.

Using Java and its object orientated nature helped in making a clear structure with room for modification and for new file formats to be included and combined with the Apache POI module allowed for Excel spreadsheets (.xlsx) files to be opened and read quickly while also gaining access to important information to help understand the sheets such as formulas used to calculate a value and how each cell is styles to help identify if it belongs to a table or is just a short description.

The aim of the project was to help create a solution to help make the process of getting data from one table to another faster while also saving the user the time it would take to go through and check the data as the current solutions either require an in depth understanding of Microsoft Excel or for the user to specify both where the tables are located and how they correlate to each other both without any error checking. While this solution can be used by anyone its simplicity does hold it back for high skilled users meaning that in the future I intent on creating a cli version to allow higher skilled users to use the program with more depth and control than just a full join on all selected tables.