**Sprint 2 Goal and User Story Evaluation**

**CITS3200 Group 38**

**Purpose**

The purpose of this document is to define the final set of user stories and goals to see the project to completion. It includes stories that were moved from delivery in sprint 2 to delivery in sprint 3 and a new set of stories which will be delivered as the final software product.

**Goals to be Completed by End of Sprint 2**

**User Story 1**

As a user I want to be able to install the OCR software on my windows device easily rather than having to clone a GitHub repository and manage the setup of the software myself.

**Acceptance Criteria**: The software should be installed onto the client’s machine using an installer and should require minimal setup procedures rather than cloning the GitHub repository and requiring manual setup.

**$ Value Test:** $--

**User Story 2**

As a user I want the errors in OCR extraction to be visualised so that it is easy to see where I need to manually review in the CSV or PDF document that I supplied to the software.

**Acceptance Criteria**: The software should have the functionality in the UI to flag/highlight cells in the output CSV that have a low confidence score or detected errors in OCR extraction.

**$ Value Test:** $--

**User Story 3**

As a user I want an easy to navigate user interface for the software so that non-technical users can run the software and process PDF documents without requiring knowledge of command line tools.

**Acceptance Criteria**: The software should have a minimal and easy to navigate user interface so that a user does not need to run the software from the command line interface.

**$ Value Test:** $--

**User Story 4**

As a user I want the OCR extraction software to be able to process the PDF documents I provide it at an appropriate speed so that I can maintain efficient business processes.

**Acceptance Criteria**: The OCR extraction process should operate at a minimum of 6 pages processed per minute.

**$ Value Test:** $--

**User Story 5**

As a user I want robust software that is not error prone so that when I use it to process PDF documents it will not run into issues that require debugging.

**Acceptance Criteria**: The software requires extensive testing in the form of unit tests and thorough user testing to ensure it is not error prone and can handle errors appropriately.

**$ Value Test:** $--

**User Story 6**

As a user I want unreadable cells such as headers to be output as an image so that I can still read the header of the table despite the error in the OCR extraction process.

**Acceptance Criteria**: The software should output the extracted header row in an image format if it cannot be accurately extracted due to issues such as language, etc.

**$ Value Test:** $--

**User Story 7**

As a user I want to know what the minimum quality of the documents I can provide to the software before the OCR extraction process becomes inaccurate so that I don’t waste my time trying to extract text from documents that are too low quality.

**Acceptance Criteria**: develop a standard for document quality which specifies the minimum quality of a PDF document to be reliably scanned and have the text extracted to a high degree of accuracy.

**$ Value Test:** $--

**User Story 8**

As a user I want to be able to quickly see a snapshot of the output after the OCR has processed the PDF document so I can efficiently evaluate the document.

**Acceptance Criteria**: the software should export a snapshot of the CSV output to the UI so that it can be viewed within the UI rather than forcing the user to exit the application and look for the file.

**$ Value Test:** $--

**User Story 9**

As a user I want to be able to ensure my documents remain organised and would therefore like to specify the location where the CSV outputs are saved after my documents are processed.

**Acceptance Criteria**: the software should allow the user to select a save location for the documents when they upload them.

**$ Value Test:** $--

**User Story 10**

As a user it would be great if I could email the CSV output to myself or colleagues straight from the software after the documents have been processed.

**Acceptance Criteria**: the software should allow the user to email the CSV output after the PDF documents have been processed.

**$ Value Test:** $--