

Sprint 3 Goal and User Stories

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.

The general expectations for sprint 3 deliverables are that the software team will improve what is currently implemented rather than expanding the scope of the project, as agreed upon during the Client meeting that occurred on the 18th of September 2024. Software accuracy has been identified as a higher priority than speed so as a result the speed requirements of the software have been reduced to reflect the requirement for higher accuracy of the OCR extraction process.

Goals to be Completed by End of Sprint 3

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.

Client Priority: Low

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.

Client Priority: High

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.

Client Priority: Medium

User Story 4

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

Acceptance Criteria: The OCR extraction process should operate at a minimum of 1 page processed per minute on the specified user hardware agreed upon.

Client Priority: Medium

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.

Client Priority: High

User Story 6

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.

Client Priority: Medium

User Story 7

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.

Client Priority: Medium

User Story 8

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.

Client Priority: Low

User Story 9

As a user I would like to be able to crop the PDF documents that I upload so that only the relevant information is captured by the OCR extraction process.

Acceptance Criteria: the software should have a crop tool to allow the client to select the parts of the PDF which will be processed by the OCR.

Client Priority: Low

User Story 10

As a user I would like to be able to upload a PDF document with many pages all at once rather than having to upload every page individually.

Acceptance Criteria: the software should handle the uploading of a PDF with many pages so that a user can upload a PDF with more than one page.

Client Priority: High

User Story 11

As a user I would like the OCR tools to extract text at an extremely high accuracy to reduce the need for user intervention.

Acceptance Criteria: the software should achieve 99% extraction accuracy on documents deemed as having a 'standard level of quality' – while there is no measurable value for quality here, OCR extraction should be very accurate.

Client Priority: High

User Story/Goal Completion Table

Story	Priority	Status
1	Low	
2	High	
3	Medium	
4	Medium	
5	High	
6	Medium	
7	Medium	
8	Low	
9	Low	
10	High	
11	High	