

Sprint 2 Goal and User Story Evaluation

CITS3200 Group 38

Purpose

The purpose of this document is to evaluate the user stories and goals proposed to be delivered by the end of sprint 2. It will evaluate the whether the acceptance criteria of each goal has been met and will assess discrepancies.

The document will describe the user stories proposed to be completed in sprint 2, including the acceptance criteria and the result of the client dollar value test. For each user story the Status/Progress will be described, the acceptance criteria will be evaluated based on the Status/Progress compared to the requirement. If the acceptance criteria has not been 'Fully met' then issues, discrepancies and resolutions will be explained.

Goals to be Completed by End of Sprint 2

User Story 1

As a user I want to be able to upload PDFs containing tables and extract the data from it, so that I don't have to manually enter data into my computer.

Acceptance Criteria: users can drag and drop or select PDFs from their file system, upload them and receive the same data in a digital format.

\$ Value Test: \$10

Evaluation

Status/Progress: A user can select the PDF document they want to have processed by the software, the software can process the document and extract the data contained within the tables and output the data in a digital format for the user. Therefore, it removes the requirement for a user to manually enter data into their computer.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Story 2

As a researcher I want data to be rearranged in a usable, malleable format, so that I can perform data analysis on the information.

Acceptance Criteria: extracted data output in a csv file format, restructured into the original table structure from the PDF document.

\$ Value Test: \$20

Evaluation

Status/Progress: The current OCR process will extract the data from the given PDF document and output the extracted text into a CSV file format, maintaining the original structure of tables given in the PDF.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Story 3

As a user I want to make sure that our information is 99% accurate, so that I don't have to spend so much time manually checking data.

Acceptance Criteria: OCR library has a requirement to achieve a 99% of accuracy. Employ multiple OCR models.

\$ Value Test: \$30

Evaluation

Status/Progress: OCR process can extract text to a high degree of accuracy, with multiple OCR models being run depending on confidence scores resulting in maximum accuracy and efficiency.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Story 4

As a user I want to make sure errors in extraction are visualised, so that I don't have to manually review extracted data.

Acceptance Criteria: detected errors in extraction of the data must be flagged for the user to manually review.

\$ Value Test: \$10

Evaluation

Status/Progress: After processing the PDF document provided to the system, the system will output the results of the OCR process including low confidence information to the user to review, i.e. the detected errors/low confidence scores are flagged to the user for which they can review further manually.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Story 5

As a user I want to extract data from many PDFs at once, so that I don't have to scan 100s of documents individually.

Acceptance Criteria: allow users to upload many PDF documents at once for batch processing.

\$ Value Test: -

Evaluation

Status/Progress: Goal dropped

Acceptance: N/A

Issues/Discrepancies: N/A

Resolution: Goal has been dropped in sprint 2 as client has clarified the expected process of extracting data from PDFs with multiple users observing process at several stages, therefore it would not be suitable to upload many PDFs at once.

Discussions occurred during client meeting 4:

https://github.com/schmilly/ProfComp3200_38Project/tree/main/Meetings/Client%20Meetings/Client%20Meeting%204

Story 6

As a user I want to view relevant tables about census data, so that I don't have to manually search for the information I'm looking for.

Acceptance Criteria: arrange extracted data into logical structure (original tables), allow it to be searchable (include table title in csv)

\$ Value Test: \$20

Evaluation

Status/Progress: the extracted data from the OCR process is arranged into the original table structure which allows for the user to analyse the data and search for data. Although, the search functionality is the user's responsibility, this goal focuses on the software's ability to produce an output that allows for the functionality of search.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Story 7

As a future developer I want appropriate documentation, so that I can improve and maintain the system.

Acceptance Criteria: technical documentation for the software should be at least minimal.

\$ Value Test: -

Evaluation

Status/Progress: Doxygen documentation has been added to current system/code for maintainability, meaning the acceptance criteria for documentation at this stage has been met.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Story 8

As a user I want an easy way to install the software onto my computer, so that I don't have to learn how to use a command line interface.

Acceptance Criteria: create installation software so that a user can directly download it rather than manually from GitHub.

\$ Value Test: \$10

Evaluation

Status/Progress: N/A

Acceptance: Criteria not yet met

Issues/Discrepancies: Goal out of scope for sprint 2

Resolution: Goal has been moved to sprint 3

Story 9

As a user I want the tables I upload to be automatically detected, so that I don't have to manually specify every row and column for the 1000s of pages I upload.

Acceptance Criteria: implement automatic table detection so that users don't have to specify table rows and columns.

\$ Value Test: \$20

Evaluation

Status/Progress: Automatic table detection has been implemented in the code so that when a user uploads a PDF the columns and rows of a table will be automatically detected based on luminosity values detected in the PDF document. This results in users not having to manually specify all table rows and columns.

Acceptance: Criteria fully met

Issues/Discrepancies: N/A

Resolution: N/A

Evaluation Summary

Story	Client \$ Value test	Goal Met?
1	10	Yes
2	20	Yes
3	30	Yes
4	10	Yes
5	0	Dropped
6	20	Yes
7	0	Yes
8	10	Out of scope
9	20	Yes
Extent to which goals have been met:	110/120	7/8

\$ values may add to over \$100 as acceptance criteria can be combined.