

User Stories

Story 1

Story: 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

Story 2

Story: 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

Story 3

Story: 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

Story 4

Story: 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

Story 4

Story: 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: -

Story 5

Story: 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

Story 6

Story: 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: -

Story 7

Story: 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

Story 8

Story: 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

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