

PDF TO CSV CONVERTOR

BY ADITHIYAN M

INTRODUCTION

- This project aims to simplify the process of converting PDF files containing student information into CSV format.
- The application is designed to extract specific data from PDFs and format it into a structured CSV file, making it easier to manage and analyze.

PROBLEM DEFINITION

Handling student data stored in PDF files manually is inefficient and error-prone. The varying formats of PDFs make data extraction challenging and time-consuming. This project aims to automate the conversion of PDF data into a structured CSV format, improving accuracy and efficiency in data management.

OBJECTIVE

- The objective of this project is to automate the extraction of student information from PDF files and convert it into a structured CSV format.
- This will enhance data management efficiency, reduce the potential for human error, and streamline the process of handling large volumes of student data.

SURVEY/IDEAS

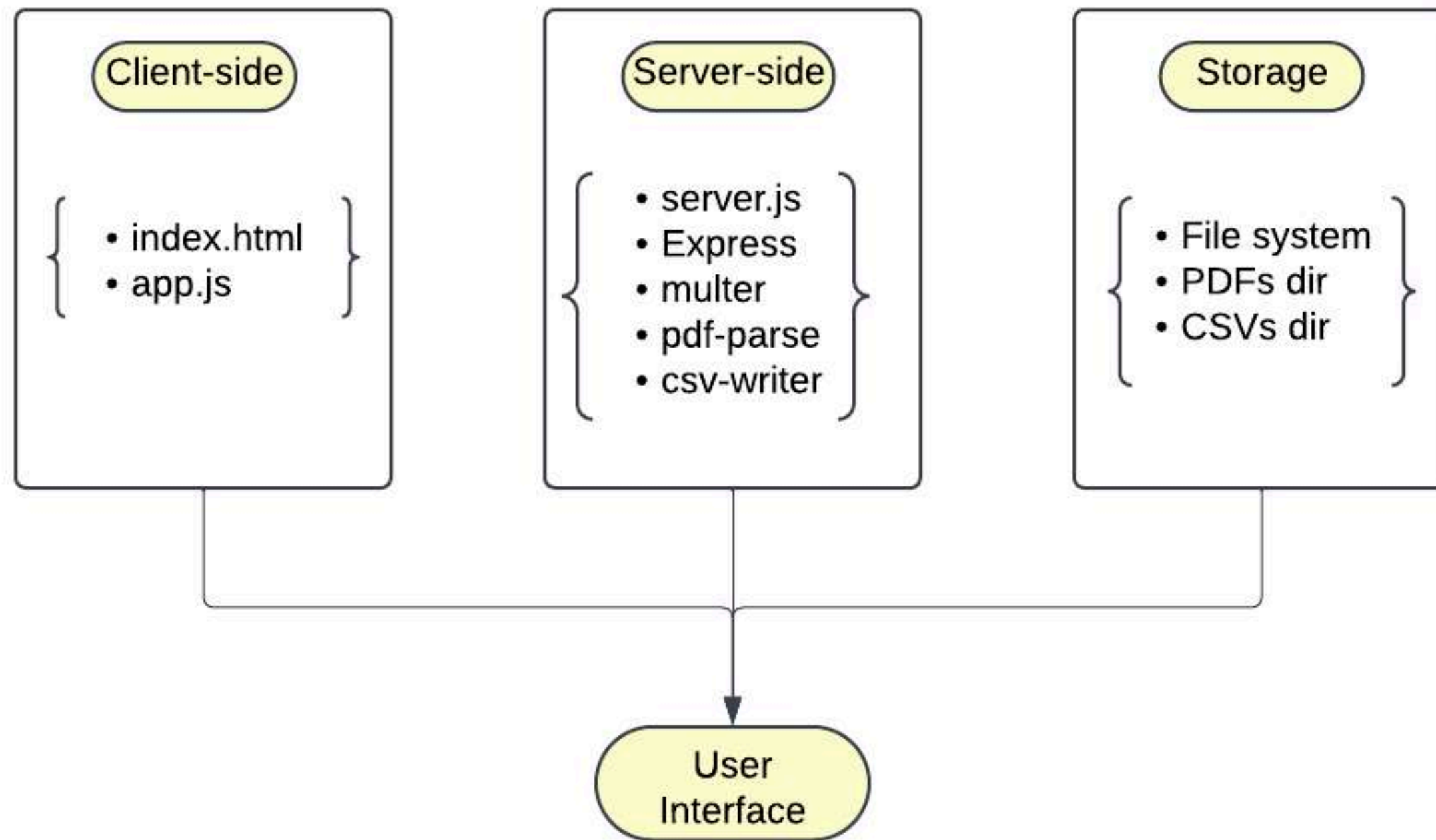
Survey

Conducted a survey among educational institutions to identify common challenges in managing student data from PDFs.

Ideas

Based on the survey, the idea emerged to develop a web-based tool that automates the extraction and conversion of PDF data into CSV format, ensuring accuracy and efficiency in data handling.

SYSTEM ARCHITECTURE



Client-side

Provides the user interface for uploading PDF files, entering the username length, and downloading the converted CSV files. Handles user interactions through event listeners.

Server-side

Manages the server, handles file uploads using *Multer*, parses PDF files using *pdf-parse*, and converts the extracted data into CSV format using *csv-writer*.

Storage

Stores uploaded PDF files and generated CSV files in specific directories to ensure organized file management.

DETAILED DESIGN

Client-side

index.html

- **Structure:** Provides the basic structure of the user interface including file input, buttons for conversion and clearing, and a download link.
- **Dark Mode Toggle:** Includes a button to toggle between light and dark modes for better user experience.

Client-side

app.js

- Event Listeners: Manages file selection, form submission and dark mode toggle.
- File Upload: Sends PDF file and username length to the server via POST.
- Response Handling: Displays download link for CSV on success, handles errors.
- Clear Functionality: Resets input fields and UI.

Server-side

server.js

- Dependencies: Uses `express`, `multer`, `path`, `pdf-parse`, `fs`, `csv-writer`.
- Directory Setup: Ensures upload and CSV directories exist.
- File Upload Handling: Manages file uploads with `multer`.
- PDF Parsing: Extracts data from PDFs using `pdf-parse`.
- CSV Conversion: Converts data to CSV format with `csv-writer`.
- Unique Filenames: Ensures unique filenames to avoid overwriting.
- API Endpoint: `/convert` endpoint handles PDF uploads and CSV conversion.
- Server Setup: Listens on a specified port for requests.

TECHNOLOGIES USED

- Frontend:
HTML, CSS, JavaScript
- Backend:
Node.js, Express.js
- Libraries:
Multer (for file uploads), pdf-parse (for PDF parsing),
csv-writer (for CSV generation)
- Other:
File System (for storing files), Path (for handling file paths)

SCREENSHOTS

User Interface

Welcome to Our Webpage

PDF to CSV Converter

Choose File No file chosen

Username length

Convert

Thank you for using our Service. See you Again 😊

Dark Mode

Welcome to Our Webpage

PDF to CSV Converter

Choose File No file chosen

Username length

Convert

Thank you for using our Service. See you Again 😊

Download Page

Welcome to Our Webpage

PDF to CSV Converter

Choose File namelist.pdf

5

Convert

Download CSV

Clear

Conversion successful! Click to download.

Thank you for using our Service. See you Again

RESULT

- Successfully converted multiple PDF files to CSV format with accurate data extraction.
- The project significantly reduces the time and effort required to manage student data. Future improvements could include support for more complex PDF formats and additional data validation features.

REFERENCES

- Node.js Documentation: Comprehensive guide and API reference for Node.js
- Express.js Documentation: The web framework used to handle routing and middleware
- Multer Documentation: The middleware used for handling file uploads
- pdf-parse and csv-writer Documentation: For manipulation of pdf and csv files



THANK YOU