FileFlip: PDF to CSV/XLSX Converter

Architecture Overview

FileFlip is a web application that converts PDF documents to CSV, XLSX, and other worksheet formats, with a focus on accounting data for integration with Sage accounting software.

Tech Stack

Backend:

- Python 3.10+
- FastAPI for API endpoints
- pdfplumber/tabula-py for PDF data extraction
- pandas for data manipulation
- openpyxl for Excel file generation

Frontend:

- TypeScript
- React 18
- Tailwind CSS
- IBM Carbon Design System
- Vite for build tooling

Deployment:

- Docker containers
- GitHub Actions for CI/CD

System Components

1. PDF Processing Engine

The core of FileFlip is the PDF processing engine that accurately extracts tabular data from various PDF formats:

- Table detection and extraction
- Text recognition and structuring
- Headers and columns identification
- Data type inference

2. Format Conversion Module

Once the data is extracted, the conversion module transforms it into the desired format:

- CSV generation with proper delimiter handling
- XLSX creation with formatting

- Format validation for Sage compatibility

3. User Interface

The frontend provides an intuitive interface for:

- PDF upload and preview
- Table selection and editing
- Format configuration
- Download options
- Error handling and reporting

Data Flow

1. User uploads PDF document
2. Backend processes and extracts table data
3. Preview is shown to user for verification/editing
4. User selects output format (CSV/XLSX)
5. Conversion is performed
6. User downloads the converted file

Folder Structure

```
fileflip/
├── backend/              # Python FastAPI application
│   ├── app/
│   │   ├── api/          # API endpoints
│   │   ├── core/         # Core application code
│   │   ├── models/       # Data models
│   │   ├── services/     # Business logic
│   │   └── utils/        # Utility functions
│   ├── tests/            # Backend tests
│   ├── Dockerfile        # Backend Docker configuration
│   └── requirements.txt  # Python dependencies
│
├── frontend/             # TypeScript React application
│   ├── public/           # Static files
│   ├── src/
│   │   ├── assets/       # Images, styles, etc.
│   │   ├── components/   # React components
│   │   ├── hooks/        # Custom React hooks
│   │   ├── pages/        # Application pages
│   │   ├── services/     # API services
│   │   ├── types/        # TypeScript type definitions
│   │   ├── utils/        # Utility functions
│   │   ├── App.tsx       # Main application component
│   │   └── main.tsx      # Application entry point
│   ├── Dockerfile        # Frontend Docker configuration
│   ├── package.json      # Frontend dependencies
│   └── tsconfig.json     # TypeScript configuration
│
├── docker-compose.yml    # Docker Compose configuration
├── README.md             # Project documentation
└── .github/              # GitHub Actions workflows
```