CMSC 25-316

Financial Receipt Capture and Analysis System

Project Proposal

Prepared for

Mehul Garnara

Capital One

By

Lindsey Marandina, Joshua Whitchurch, Josh Tarongoy, and Will Southerland

Under the supervision of

Thang Dinh

October 8th, 2024

Executive Summary

Project Overview

This project aims to develop an application that streamlines the process of capturing and managing hard copy credit card transaction receipts through capture, storage, and analysis of physical receipts. The system will allow users to:

- Search receipts by date, vendor, category, and other criteria.
- Analyze spending across different periods and categories.
- Identify spending trends to gain valuable insights.
- Utilize a categorization system to enhance financial tracking and analysis.

Project Goals/Objectives

To minimize costs and maximize scalability, the project will utilize open-source software. Key design objectives are:

- Receipt capture and storage by digitizing and storing physical receipts.
- Categorization by automatically and/or manually classifying receipts based on common spending categories.
- Search functionality that allows users to locate specific receipts efficiently.
- Spending analysis through comparing and analyzing spending over time and across spending categories.
- Trend Identification through identifying spending patterns to inform financial decisions.

Design Specifications

The system will meet the following specifications:

- Image processing: Accepting and processing receipts in various formats (e.g., JPG, PNG).
- OCR accuracy: Ensuring accurate extraction of text from receipts using OCR technology.
- Key information extraction: Identifying critical data points from receipts (e.g., date, vendor, amount).
- Categorization**:** Providing both automatic and/or manual categorization options.
- Search capabilities: Enabling efficient searching based on multiple criteria.
- Analysis tools: Offering tools for analyzing spending trends and patterns.
- Open-source integration: Incorporating at least one open-source project.

Current Progress and Future Plans

The team has initiated research into OCR technologies and is currently evaluating Tesseract for its accuracy. DynamoDB has been selected as the primary database, and efforts are underway to integrate data from Tesseract. Future plans include:

- Machine learning: Training a machine learning algorithm using the receipt database to enhance categorization and analysis.
- Data visualization: Developing visualizations to present spending data effectively.
- UI design: Creating a user-friendly web interface.

Table of Contents

Section A. Problem Statement

Many people face challenges keeping track of spending and are often unaware of how much they are actually spending versus how much they think they're spending. With many often spending more than what they can afford leading to credit card bills racking up. The manual processes of storing, searching, and analyzing physical receipts is time-consuming, inefficient, and prone to errors. Financial receipts are also often lost, disorganized, or difficult to analyze, making it challenging for individuals to track their spending and make informed financial decisions. This project will allow individuals to keep track of their spending by digitally uploading receipts so they are able to make informed financial decisions. It will also allow individuals to know just exactly how much they are spending and where they are spending it.

Capital One is an American financial services company known for their innovative approach to banking and credit card services. Capital One offers many different credit cards to fit the needs of many different customers. The company follows a data-driven approach with a "reverence for insights derived from data" (Bean, 2024) and "adopted a "You Build, Your Data" approach to give our data users greater ownership and accountability over their data" (Lebonitte, 2024).

The problem has currently been addressed through applications that allow users to track their expenses through connecting their accounts to track spending and income. One well known application is Rocket Money. Rocket Money "is a personal finance app…that helps you manage your money, lower your bills, and even cancel unwanted subscriptions" (Hamilton, 2022). Other banking companies, like Bank of America, have similar expense and spending insights. Bank of America has "Category-specific interactive charts and longer-term spending trends help you identify where you are frequently over or under budget" (*Use the spending & budgeting tool to get a clear view of your finances*). One study found that "financial literacy helps people make informed and responsible financial decisions, boosting financial stability and reducing financial worries" (Bai, 2023).

Section B. Engineering Design Requirements

The current design requirement is to use as many open source projects as we can. This requirement was decided upon since web services can become pricey depending on scale. On our project scale, cost would be minimal. However, looking at the scalability of the project is important, therefore web services would not be cost effective in scaling the project up.

B.1 Project Goals (i.e. Client Needs)

The goals of this project is to create an application that digitally captures and stores receipts, allows the user to search specific receipts, categorize receipts (ie. grocery, transportation, entertainment…), and analyze spending trends and categorically to provide insights.

Some general examples of the project goals are as follows:

- To produce an application that takes jpeg, png, or digital captures and stores them.
- To design an algorithm that categorizes receipts based on spending categories.
- To perform analysis on the receipt data to provide spending insights

B.2 Design Objectives

Some of the objectives for the project are as follows:

- The design will allow for receipt capture.
- The design will digitally store physical receipts.
- The design will categorize receipts based on common spending categories.
- The design will allow the user to search for receipts via date, vendor, etc.
- The design will compare and analyze spending over a period of time and categories.
- The design will analyze spending trends.

B.3 Design Specifications and Constraints

Some of the design specifications for the project are as follows:

- Design must have the ability to receive images (jpg, png, etc.) of receipts for extraction. (Functional constraint)
- Design must have the ability to process receipts of varying layouts and formats. (Functional constraint)
- Design must have accurate OCR results when extracting text from receipts (Functional constraint)
- Design must have the ability to extract key information (Functional constraint) – examples include date, vendor, total, etc.
- Design must be able to automatically and/or manually categorize receipts for organization (Functional constraint) – Categorical examples include grocery, shopping, entertainment, medical, utilities, dining out, transportation, travel, etc.
- Design must have the ability to search for receipts (Functional constraint) – examples include search by date, vendor, cost, category, etc.
- Design must analyze spending based on time periods and categories. (Functional constraint)
- Design must show trend analysis to identify spending patterns. (Functional constraint)
- Design must include at least one open-source project (Cost constraint)

References

Bai R. (2023). Impact of financial literacy, mental budgeting and self control on financial wellbeing: Mediating impact of investment decision making. PloS one, 18(11), e0294466. https://doi.org/10.1371/journal.pone.0294466

Bean, R. (2024, August 12). Capital one: The ongoing story of how one firm has been pioneering data, analytics, & Ai Innovation for over three decades. Forbes. https://www.forbes.com/sites/randybean/2024/08/11/capital-one-the-ongoing-story-of-how-one-firm-has-been-pioneering-data-analytics--ai-innovation-for-over-three-decades/

Hamilton, K. (2022, September 29). What is rocket money and why should you sign up?. What Is Rocket Money? https://www.rocketmoney.com/learn/personal-finance/what-is-rocket-money

Lebonitte, J. (2024, June 10). How capital one innovates using "You build, your data." Capital One. https://www.capitalone.com/software/blog/capital-one-you-build-your-data/

Use the spending & budgeting tool to get a clear view of your finances. Bank of America. (n.d.). https://www.bankofamerica.com/online-banking/mobile-and-online-banking-features/spending-budgeting/?sgm=adv