Web Image Retrieval

Project Documentation

Contents

[Introduction](#_bookmark0) [Project Process](#_bookmark1)

[Process Diagram](#_bookmark2) [Define Scope](#_bookmark3) [Inventory](#_bookmark4) [Analyze](#_bookmark5)

[Design](#_bookmark6) [Prototype](#_bookmark7)

[Prototype Validation](#_bookmark8) [Iterate/Refine](#_bookmark9)

[Business Definition and Data Source](#_bookmark10) [Test](#_bookmark11)

[Functional Testing](#_bookmark12)

[UAT (User Acceptance Testing)](#_bookmark13) [Go Live](#_bookmark14)

[Post Production Support](#_bookmark15) [Project Documentation](#_bookmark16)

[Scope Statement](#_bookmark17) [Data Inventory List](#_bookmark18) [Report List](#_bookmark19)

[Business Requirements](#_bookmark20) [Data Dictionary](#_bookmark21)

[Functional Specification Document](#_bookmark22) [Functional Test Document](#_bookmark23)

[UAT Document](#_bookmark24) [Go Live Checklist](#_bookmark25)

[High-Level Project Plan](#_bookmark26) [Short-Term Project Plan](#_bookmark27) [Issue Log](#_bookmark28)

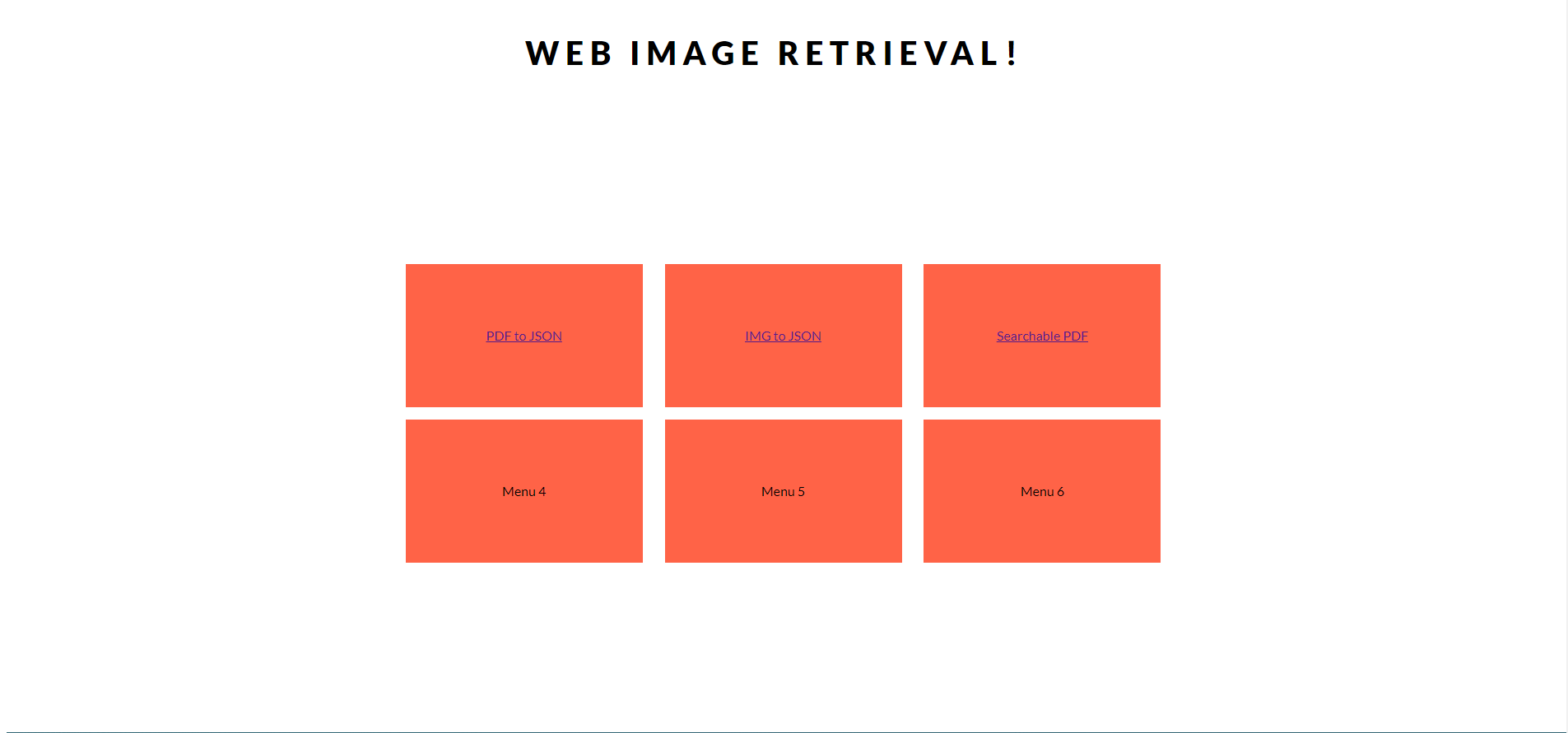
[Bug Tracker](#_bookmark29) [Enhancement Requests](#_bookmark30) [Security Request](#_bookmark31)

# Introduction

Web Image retrieval is a web application that aims to retrieve some of the information contained in a PDF file, then convert it into a JSON file for further processing. In this web application, we use Optical Character Recognition (OCR) to retrieve information on files, as for the additional feature, namely Searchable PDF which is useful converting PDF to scanning word to make that searchable.

# Project Documentation

## Homepage



***PDF Information Retrieval (PDF to JSON)***

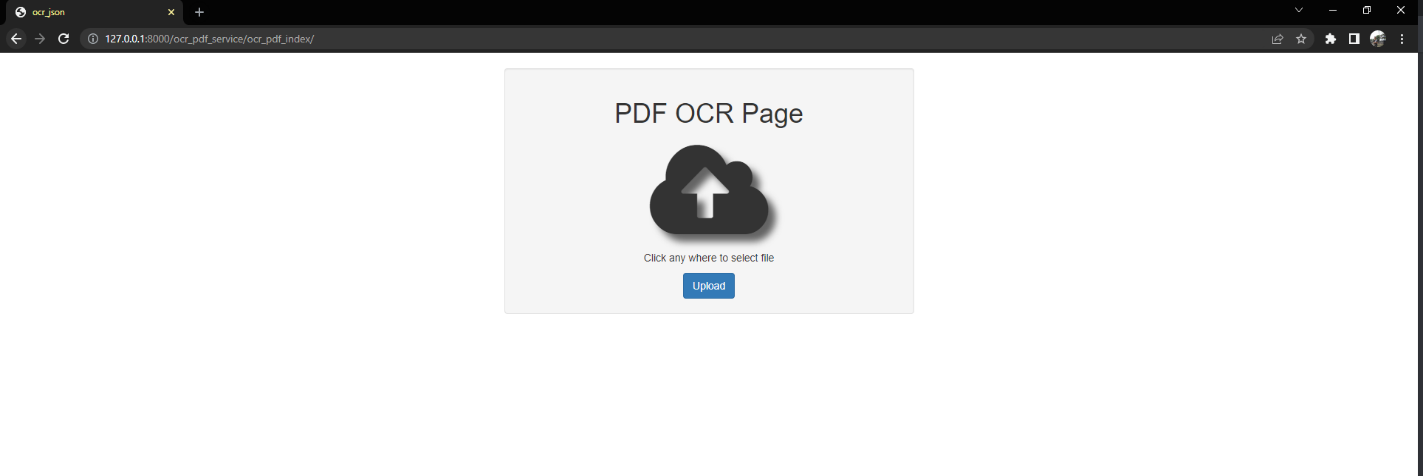
PDF Information Retrieval is a feature that accepts input PDF files and converts them into JSON files so that they can be processed into a management system or NLP Analysis. In the flow, from input PDF File we convert it into a images format (JPG). After that OCR can proceed to retrieve the information in the image. Figure one will be shown the PDF to JSON page looks like.

Figure 1. PDF File to JSON File page.

In the PDF File to JSON File page, you can interact with clicking on the cloud icon, then you can choose the PDF file on the popup menu, if you using window the pop up will be just like in figure 2.

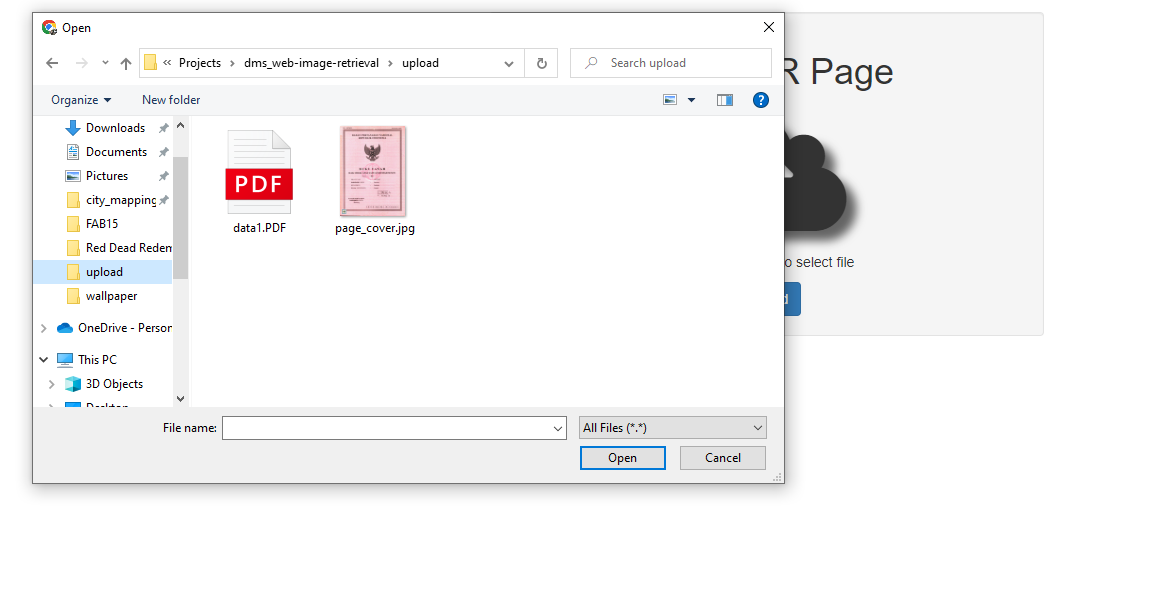


Figure 2. Chose the PDF File

Next after you choose the PDF file, all you need to do is just click on the upload button, and the information in the PDF file will scan by the OCR then the result will be extracted as JSON File.

## Image Information Retrieval (IMG to JSON)

Once the project scope has been defined and agreed upon, the Business Owner will create an inventory of the data elements and reports required to support the defined scope. The Business Owner should try to identify the source system and data domains as well as the needs and the gaps of the project.

## Searchable PDF

Review all of the data elements and the reports. Engage in detailed discussions to profile the data and determine the business rules around the data. Identify any data issues, challenges and try to identify any risks. Document and define the business and data requirements in the Business Requirements Document. Determine the security requirements as well; who should be able to view the data, and who should not.

The Business Requirements document should be reviewed and signed off by the core team so everyone is aware of the actual requirements.