eHospital LIS Image OCR API

The objective was to create a microservice to extract text data from a lab report file in such a format that data becomes easily accessible to the doctors in the hospital. So, Tesseract OCR API was used to perform optical character recognition and extract text from the image file and return a Json file so that the data can be accessed easily according to the keys.

Build and Run

To build and run the Lab Report OCR API, follow these steps:

1. Clone the repository from GitHub:

   https://github.com/sudo-sahilroy/imageOCR

2. Navigate to the project directory:

   cd imageOCR

3. Build the project using Maven:

   mvn clean package

4. Run the application:

   java -jar target/imageOCR.jar

The API will be accessible at http://localhost:8080.

Usage

1. Image to Json

Endpoint: /image-to-json

Method: POST

Description: Extracts text from the uploaded image file and converts it into JSON format.

Request Parameters:

Parameter	Type	Description
image	report.jpg	The image file to be processed

Sample Image :

Success Response:

- Status Code: 200 OK
- Content Type: application/json

- Body:

```json
{
    "": "150Year(s) (Male",
    "Age: ": "50 years 10 months 16 days",
    "Lab Reference NO: ": "52",
    "Sample Details :BI0-": "060623129 (Blood) / Bio-Chemistry Sub Centre Report
        Date: 06/06/2023 06:13 PM",
    "HBAIC ": "52 % -6.50 Verification Comment",
    "SERUM CALCIUM ": "7.2 mg/dL 9-11 mg/dL Verification Comment",
    "Sample Received Date: ": "06/06/2023 05:48 PM",
    "Sample Collection Date: ": "06/06/2023 05:47 PM",
    "SERUM POTASSIUM ": "45 mEq/L Verification Comment",
    "SERUM SODIUM ": "142 mEq/L",
    "Government of Tripura ,Agartala, Tripura West , ": "799006",
    "UHID: ": "2022010049138",
    "SERUM UREA ": "35 mg/dL 15-40 mg/dL",
    "SERUM CREATININE ": "204 mg/dL 0.6- 1.5 mg/dL",
    "SERUM URIC ACID ": "6.7 mg/dL 3 7 mg/dL 0Year(s) - Verification Comment",
    "Report Upload Date: ": "06/06/2023 06:13 PM",
    "Order Date: ": "06/06/2023",
    "Reg Date: ": "21/07/2022 11:59 AM"
}
```

1. Image to Text

Endpoint: /image-to-text

Method: POST

Description: Extracts text from the uploaded image file and converts it into plain text format.

Request Parameters:

Parameter	Type	Description
image	report.jpg	The image file to be processed

Success Response:

- Status Code: 200 OK
- Content Type: application/json

- Body

Pretty   Raw   Preview   Visualize   Text ∨

```
 1   Government of Tripura ,Agartala, Tripura West , 799006
 2
 3   g @gflx Government Medical College & Gobind Ballav Pant Hospital
 4
 5   LABORATORY OBSERVATION REPORT
 6
 7   UHID: 2022010049138
 8
 9   Patient Name: - Mr. TEST TEST
10   Sex: Male
11
12   Order By: Dr. IPD Doctor Incharge
13   Unit Name: Unit I
14
15   Sample Collection Date: 06/06/2023 05:47 PM
16
17   Lab Name: BIOCHEMISTRY
18   Department: Medicine
19
20   Reg Date: 21/07/2022 11:59 AM
21   'Ward Name : Male Medicine
22
23   Age: 50 years 10 months 16 days.
24   Order Date: 06/06/2023
25
26   Unit In-charge: Dr. Arunaba Dasgupta
27
28   Sample Received Date: 06/06/2023 05:48 PM
```

```
30   Lab Reference NO: 52
31
32   Report Upload Date: 06/06/2023 06:13 PM
33
34   Sample Details :BI0-060623129 (Blood) / Bio-Chemistry Sub Centre Report Date: 06/06/2023 06:13 PM
35
36   Test Name Result Reference Range Verification Comment
37   HBAIC 52 % -6.50 Verification Comment
38   KIDNEY FUNCTION TEST
39   SERUM UREA 35 mg/dL 15-40 mg/dL
40   SERUM CREATININE 204 mg/dL 0.6- 1.5 mg/dL.
41   OYc:n(s) - T50Year(s)
42   SERUM URIC ACID 6.7 mg/dL 3 7 mg/dL 0Year(s) - Verification Comment
43   150Year(s) (Male)
44   ELECTROLYTE PROFILE
45   SERUM SODIUM 142 mEq/L
46   SERUM POTASSIUM 45 mEq/L Verification Comment
47   SERUM CALCIUM 7.2 mg/dL 9-11 mg/dL Verification Comment
48   Entered By Verified By
49
50   (Mr.HOSPITAL ADMI!
51
52   # This is computer generated report, hence no signature is required.
53
54   (Mr.HOSPITAL ADMIN)
55
```

Failure Response:

1. No file provided
- Status Code: 400 Bad Request
- Body: Image File is Required

2. Non-image file provided
- Status Code: 400 Bad Request
- Body:Invalid Image File

3. Issue with image to text conversion
- Status Code: 500 Internal Server Error
- Body: Error extracting text from image

Libraries and Tools Used

The following libraries and tools are used in this API:
- Language : Java (JDK 17)
- Framework: SpringBoot (3.0.8)
- IDE: IntelliJ

Dependencies

Maven Dependencies used:
- tess4j: The library provides optical character recognition (OCR) support for: TIFF, JPEG, GIF, PNG, and BM(provided in this project)
- org.json: JSON is a light-weight, language independent, data interchange format
- spring-boot-starter-web: Starter for building web, including RESTful, applications using Spring MVC. Uses Tomcat as the default embedded container

License

This API is open source and licensed under the MIT License.

Contributions

Contributions to this project are welcome. If you encounter any issues or have suggestions for improvements, please open an issue or submit a pull request on the project's repository or mail me at suggi.aditya@gmail.com.