Building Software Systems

Lecture 5.5
Case Study – Tesseract and OpenAI API

SAURABH SRIVASTAVA

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

IIT (ISM) DHANBAD

What will we do today?

We will work on building an Application which answers questions from Posters
- The Posters shall be in an Image Form
- Based on the content of the Poster, we can answer Questions from the same

We will use two distinct steps to do the same
- Perform OCR over the Image to extract Text
- Use the extracted text to answer questions

What is Tesseract?

Tesseract is an open-source Optical Character Recognition (OCR) engine
- Initially developed by Hewlett-Packard in the late 1980s
- Released as open-source in 2005 and has been maintained by Google since 2006

It can read and interpret printed text from images or scanned documents
- It is capable of recognizing more than 100 languages out of the box
- It can understand complex document structures, including columns, footnotes, and headers

It has a CLI Tool
- Tesseract can be used directly from the command line, allowing for scriptable and automated processing

Can also be used in your programs
- Easily integrates into coding projects, offering libraries for many programming languages to embed OCR feature

Using OpenAI APIs

Provides access to OpenAI's AI models through an API interface

◦ Developers can register themselves and use the API to perform tasks such as Chat Completions

You need to create an API key

◦ A unique identifier assigned to each user or application, granting authenticated access to the API's functionalities

Let us see how we can use it !!

Check out the Video Recording of the Lecture

https://www.youtube.com/watch?v=8jdTyu77zhQ

Relevant Commands for the Lecture

```
 sudo apt install tesseract-ocr libtesseract-dev

#Setup Python Env
 python -m venv mlaas
 source mlaas/bin/activate

#Install Tesseract
 pip install pytesseract pip install openai
```