**Introduction-**

This code creates a web-based Optical Character Recognition (OCR) and keyword search tool using Gradio and Tesseract OCR. It allows users to upload an image containing English and/or Hindi text, extracts the text from the image using Tesseract, and then enables the user to search for a specific keyword within the extracted text.

**Key features-**

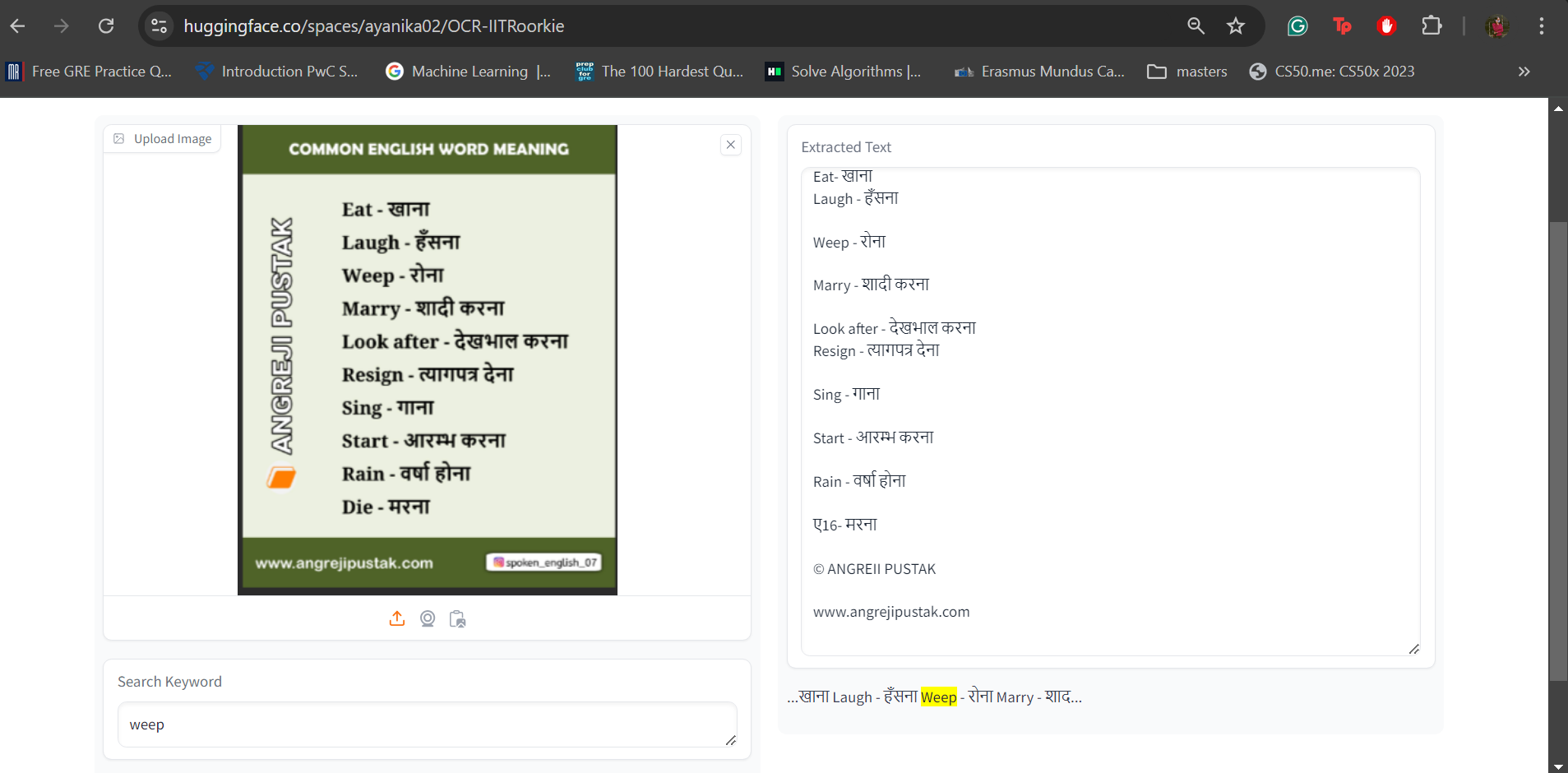
1. **OCR Functionality**:
   * The function ocr\_image() processes an uploaded image, ensures it's in RGB mode, and uses Tesseract to extract text in both English (eng) and Hindi (hin) languages.
2. **Keyword Search**:
   * The function search\_text() performs a case-insensitive search within the extracted text and highlights the matched keywords, showing surrounding context.
3. **Gradio Interface**:
   * The process\_image() function integrates OCR and keyword search.
   * The Gradio interface allows users to upload an image and input a keyword for searching in the extracted text. It returns the extracted text and displays any matching keywords with surrounding context.

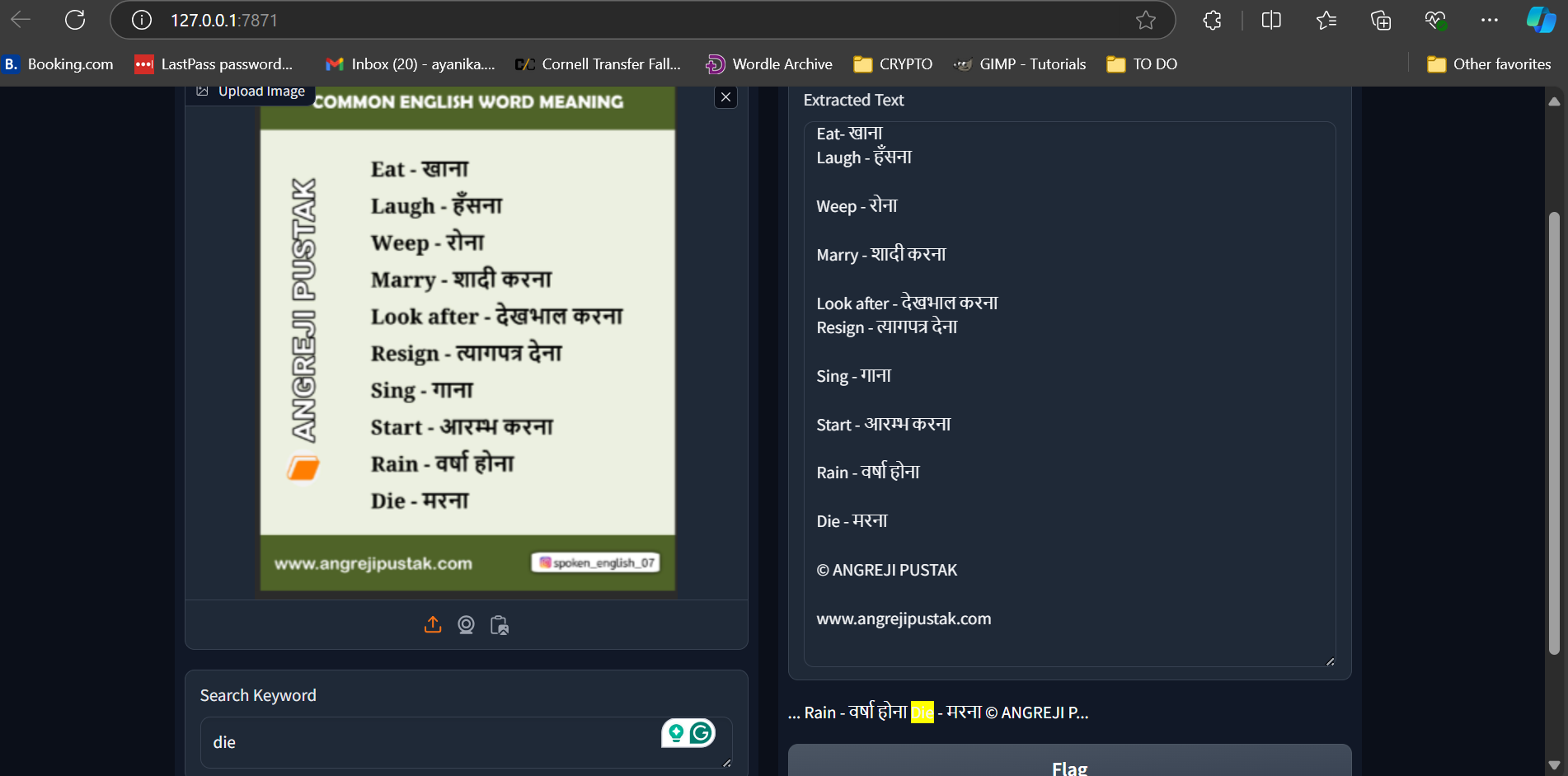
**Final Deployment-**

This code is ready to be deployed as a web app on Hugging Face Spaces, where it is hosted.

<https://huggingface.co/spaces/ayanika02/OCR-IITRoorkie>

**Sample Output 1-**





**Output produced by the web app-**

COMMON ENGLISH WORD MEANING

Eat- खाना

Laugh - हँसना

Weep - रोना

Marry - शादी करना

Look after - देखभाल करना

Resign - त्यागपत्र देना

Sing - गाना

Start - आरम्भ करना

Rain - वर्षा होना

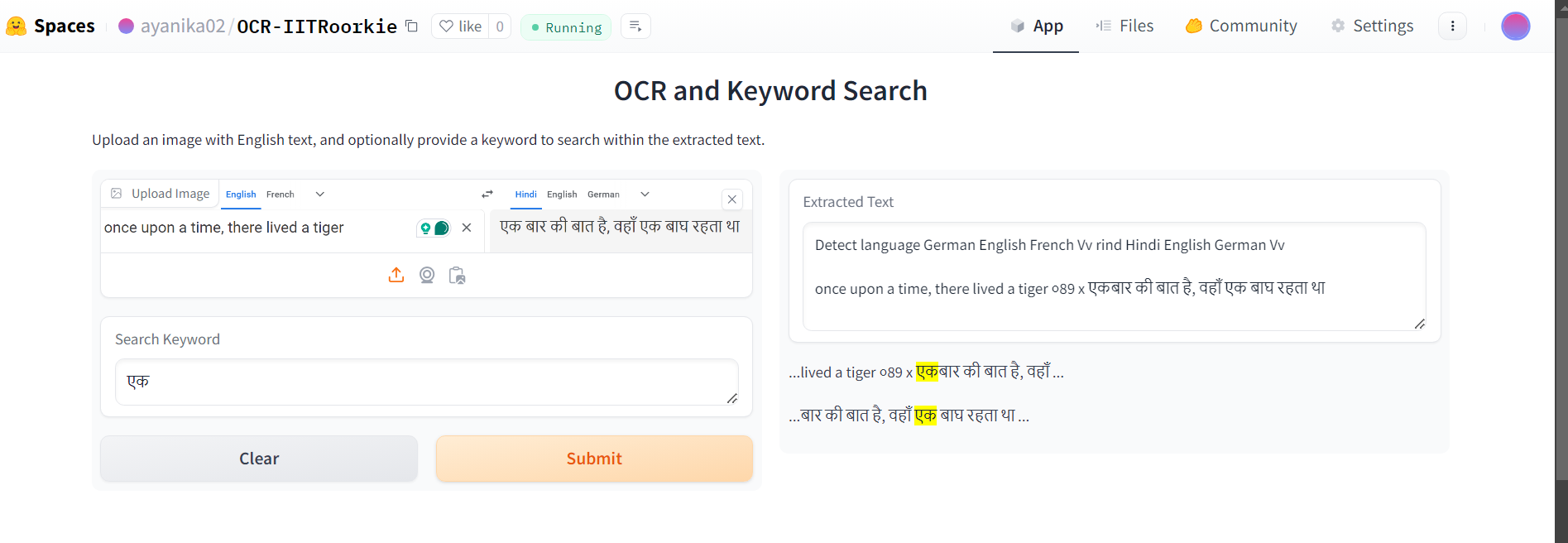
ए16- मरना

© ANGREII PUSTAK

[www.angrejipustak.com](http://www.angrejipustak.com)

**Observation-** The 1st image is of the output generated in Hugging Space and the 2nd image shows the output generated on my local machine. For the same code and input file, there is a slight difference. Everything is correctly extracted in Hugging Space webapp except the word “Die”, which is correctly extracted when locally hosted.

**SAMPLE OUTPUT 2-**



**Output-**

Detect language German English French Vv rind Hindi English German Vv

once upon a time, there lived a tiger ०89 x एकबार की बात है, वहाँ एक बाघ रहता था

**Observation-** The stray “Vv”s are due to unexpected signs in the picture. 1st “Vv” appears due to the swap sign and the 2nd “Vv” appears due to the cross sign. The next stray reading, “०89” is because of Grammarly and the “x” next to it due to cross sign.

**Conclusion-**

The model correctly extracts many Hindi and English words, but often misinterprets when it sees stray signs and symbols.