**Keyword extractor or counter of a pdf**

import re

from collections import Counter

def extract\_keywords(text, num\_keywords=5):

# Remove non-alphabetic characters and convert to lowercase

cleaned\_text = re.sub(r'[^a-zA-Z\s]', '', text.lower())

# Tokenize the text into words

words = cleaned\_text.split()

# Calculate word frequencies using Counter

word\_frequencies = Counter(words)

# Extract the top N keywords

top\_keywords = word\_frequencies.most\_common(num\_keywords)

return top\_keywords

if \_\_name\_\_ == "\_\_main\_\_":

# Example usage

text = "Natural language processing is a subfield of artificial intelligence that focuses on the interaction between computers and humans using natural language."

keywords = extract\_keywords(text, num\_keywords=5)

if keywords:

print("Top Keywords:")

for keyword, count in keywords:

print(f"{keyword}: {count}")

else:

print("No keywords found.")

**OUTPUT :**

C:\Users\civilsys51\PycharmProjects\pythonProject\.venv\Scripts\python.exe C:\Users\civilsys51\PycharmProjects\pythonProject\KeywordExtraction.py

Top Keywords:

natural: 2

language: 2

processing: 1

is: 1

a: 1

Process finished with exit code 0

**RESULT :**

Thus, we created keyword extraction using Python successfully.