import requests

from bs4 import BeautifulSoup

from sentence\_transformers import SentenceTransformer

import faiss

# 1. Crawl and Scrape Website Content

def crawl\_and\_scrape(url):

response = requests.get(url)

soup = BeautifulSoup(response.text, "html.parser")

# Extract textual content (e.g., paragraphs)

content = [p.get\_text() for p in soup.find\_all('p')]

return content

# 2. Chunk Content

def chunk\_text(text\_list, chunk\_size=200):

chunks = []

for text in text\_list:

for i in range(0, len(text), chunk\_size):

chunks.append(text[i:i + chunk\_size])

return chunks

# 3. Create Embeddings and Store in FAISS

def store\_embeddings(chunks):

model = SentenceTransformer("all-MiniLM-L6-v2")

embeddings = model.encode(chunks)

dimension = embeddings.shape[1]

index = faiss.IndexFlatL2(dimension)

index.add(embeddings)

return index