Training Day 4 Report:

Website Summarization with BeautifulSoup and an API

Website summarization means taking the content of a webpage and creating a short and meaningful summary from it. This is done in two main steps: first, we extract the content of the website, and then we use an AI tool or API to summarize it.

- **Step 1 – Getting Website Content:**
  We use a tool called **BeautifulSoup** in Python to collect information from a website. It reads the webpage's structure and helps us take out important text like paragraphs and headings.

- **Step 2 – Extracting Links and Images:**
  Along with text, BeautifulSoup can also collect other useful elements from a webpage, such as hyperlinks and image URLs. This can help us understand the sources, references, or visual content related to the article. We can extract all links using HTML tags like `<a>` and images using `<img>` tags.

- **Step 3 – Cleaning the Text:**
  The content from websites often contains ads, links, or unnecessary code. BeautifulSoup helps us remove these unwanted parts so that we are left with clean, readable text.

- **Step 4 – Sending Text to an API:**
  Once we have the clean content, we send it to a summarization API like **OpenAI**, **HuggingFace**, or **Gemini**. These tools use artificial intelligence to understand the text and generate a short summary.

- **Step 5 – Generate the Summary:**
  The API reads the full content and replies with a few short sentences that explain the main idea of the article. This summary can then be displayed, stored, or shared.

- **Tools Used:**

  - **BeautifulSoup** – To read and clean the content, and extract links and images from the website.
  - **Requests** – To connect to the website and get its data.

– **LLM** – To generate the summary.

- **Why It's Useful:**
  This process saves time. Instead of reading long articles, people can quickly understand the main points. It's useful for news readers, researchers, and students. Also, extracting links and images can help in collecting references and visual data for deeper analysis.

In summary, website summarization using BeautifulSoup and an API is a smart way to read only the important parts of a webpage. It combines web scraping and artificial intelligence to help us learn faster and more efficiently.