SCRAPING TEXT,LINKS AND IMAGES FROM WEB PAGES

```
import requests
from bs4 import BeautifulSoup


import requests
from bs4 import BeautifulSoup

# Function to fetch HTML content from a URL
def fetch_html(url):
    try:
        response = requests.get(url)
        response.raise_for_status()  # Raise an error for bad status codes
        return response.content
    except requests.RequestException as e:
        print(f"Failed to fetch {url}: {e}")
        return None

# Function to extract paragraphs
def extract_paragraphs(soup):
    paragraphs = soup.find_all('p')
    print("Paragraphs:")
    for paragraph in paragraphs:
        print(paragraph.get_text())
    print("\n")

# Function to extract headings
def extract_headings(soup):
    headings = soup.find_all(['h1', 'h2', 'h3', 'h4', 'h5', 'h6'])
    print("Headings:")
    for heading in headings:
        print(heading.get_text())
    print("\n")

# Function to extract links
def extract_links(soup):
    links = soup.find_all('a', href=True)
    print("Links:")
    for link in links:
        print(link['href'])
    print("\n")

# Function to extract image sources
def extract_images(soup):
    images = soup.find_all('img', src=True)
    print("Image Sources:")
    for image in images:
        print(image['src'])
    print("\n")

# Function to extract metadata
def extract_metadata(soup):
    metadata = soup.find_all('meta')
    print("Metadata:")
    for meta in metadata:
        print(meta.attrs)
    print("\n")

# Function to extract script tags
def extract_scripts(soup):
    scripts = soup.find_all('script')
    print("Script Tags:")
    for script in scripts:
        print(script.get('src'))
    print("\n")

# Function to extract styles
def extract_styles(soup):
    styles = soup.find_all('style')
    print("Styles:")
    for style in styles:
        print(style.get_text())
    print("\n")
```

```
# Function to scrape data from a webpage
def scrape_webpage(url):
    html_content = fetch_html(url)
    if html_content:
        soup = BeautifulSoup(html_content, 'html.parser')

        extract_paragraphs(soup)
        extract_headings(soup)
        extract_links(soup)
        extract_images(soup)
        extract_metadata(soup)
        extract_scripts(soup)
        extract_styles(soup)
    else:
        print(f"Failed to fetch {url}")

# URL of the webpage to scrape (Netflix)
url = 'https://www.netflix.com/'

# Scrape data from the webpage
scrape_webpage(url)
```

```
Paragraphs:
Watch anywhere. Cancel anytime.
Watch on Smart TVs, Playstation, Xbox, Chromecast, Apple TV, Blu-ray players, and more.
Save your favorites easily and always have something to watch.
Stream unlimited movies and TV shows on your phone, tablet, laptop, and TV.
Send kids on adventures with their favorite characters in a space made just for them—free with your membership.


Headings:
Unlimited movies, TV shows, and more
Ready to watch? Enter your email to create or restart your membership.
Enjoy on your TV
Download your shows to watch offline
Watch everywhere
Create profiles for kids
Frequently Asked Questions
What is Netflix?
How much does Netflix cost?
Where can I watch?
How do I cancel?
What can I watch on Netflix?
Is Netflix good for kids?
Ready to watch? Enter your email to create or restart your membership.


Links:
/tw-en/login
https://help.netflix.com/contactus
https://help.netflix.com/support/412
https://help.netflix.com
/youraccount
https://media.netflix.com/
http://ir.netflix.com/
https://jobs.netflix.com/jobs
/watch
https://help.netflix.com/legal/termsofuse
https://help.netflix.com/legal/privacy
#
https://help.netflix.com/legal/corpinfo
https://help.netflix.com/contactus
https://fast.com
https://help.netflix.com/legal/notices
https://www.netflix.com/tw-en/browse/genre/839338


Image Sources:
https://assets.nflxext.com/ffe/siteui/vlv3/8728e059-7686-4d2d-a67a-84872bd71025/2b5319f6-801c-421e-a84c-01280688d7f7/TW-en-20240708-P
https://assets.nflxext.com/ffe/siteui/acquisition/ourStory/fuji/desktop/tv.png
https://assets.nflxext.com/ffe/siteui/acquisition/ourStory/fuji/desktop/mobile-0819.jpg
https://assets.nflxext.com/ffe/siteui/acquisition/ourStory/fuji/desktop/boxshot.png
https://assets.nflxext.com/ffe/siteui/acquisition/ourStory/fuji/desktop/device-pile.png
https://occ-0-395-325.1.nflxso.net/dnm/api/v6/19OhWN2dO19C9txTON9tvTFtefw/AAAABejKYujIIDQciqmGJJ8BtXkYKKTi5jiqexltvN1YmvXYIfX8B9CYwoo
https://assets.nflxext.com/ffe/siteui/acquisition/common/transparent_1x1.png


Metadata:
{'http-equiv': 'Content-Type', 'content': 'text/html; charset=UTF-8'}
```