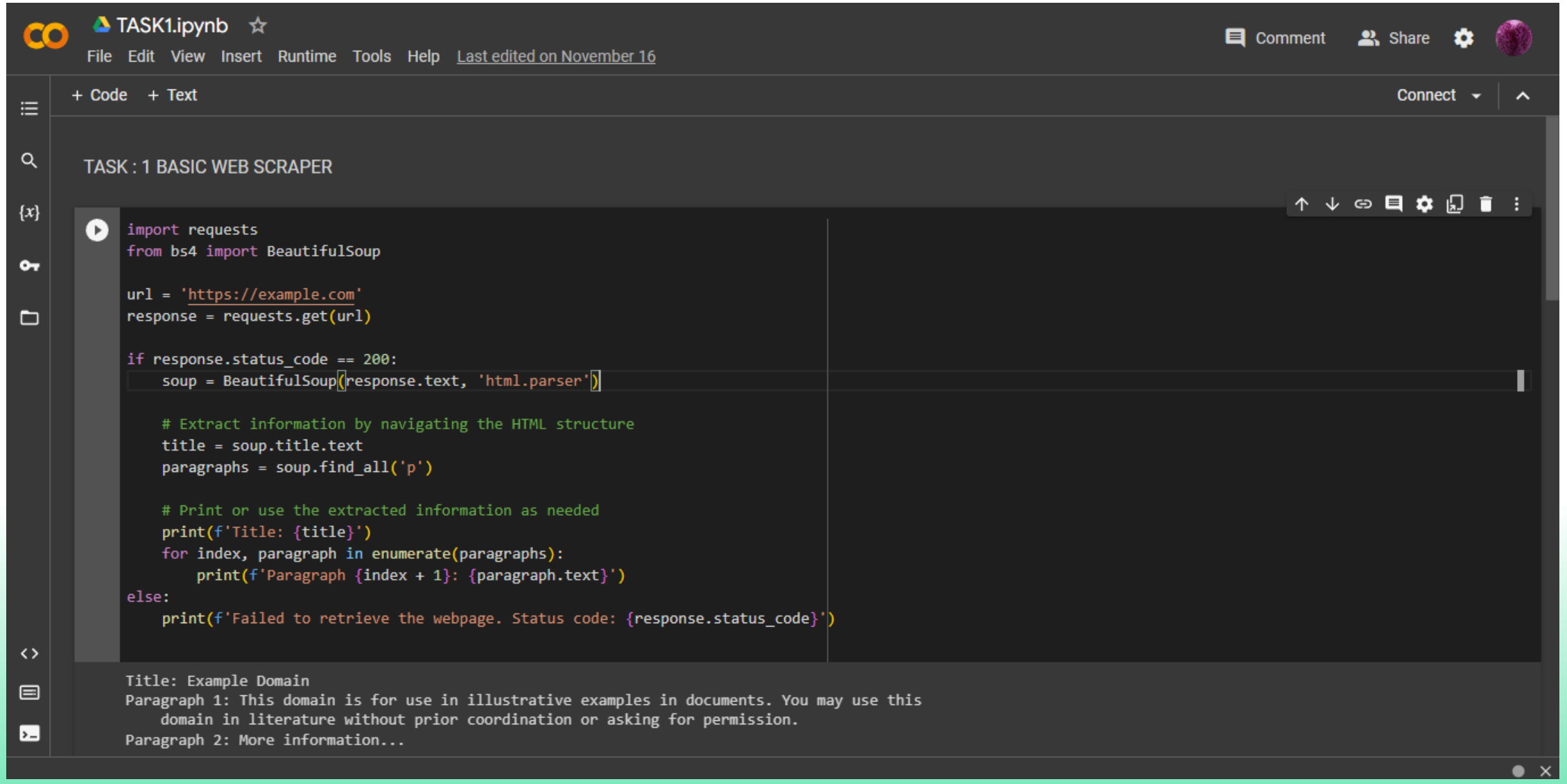
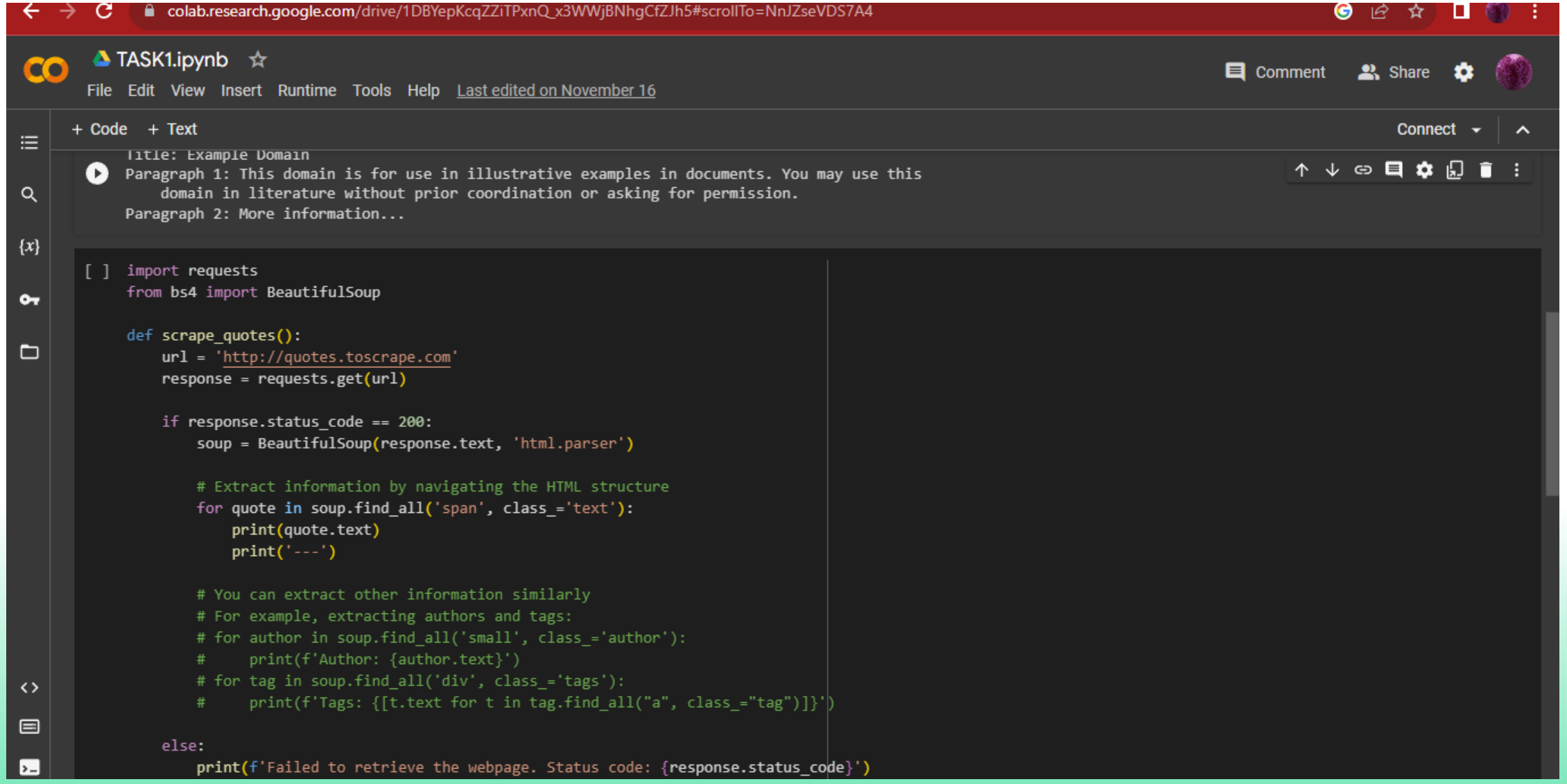
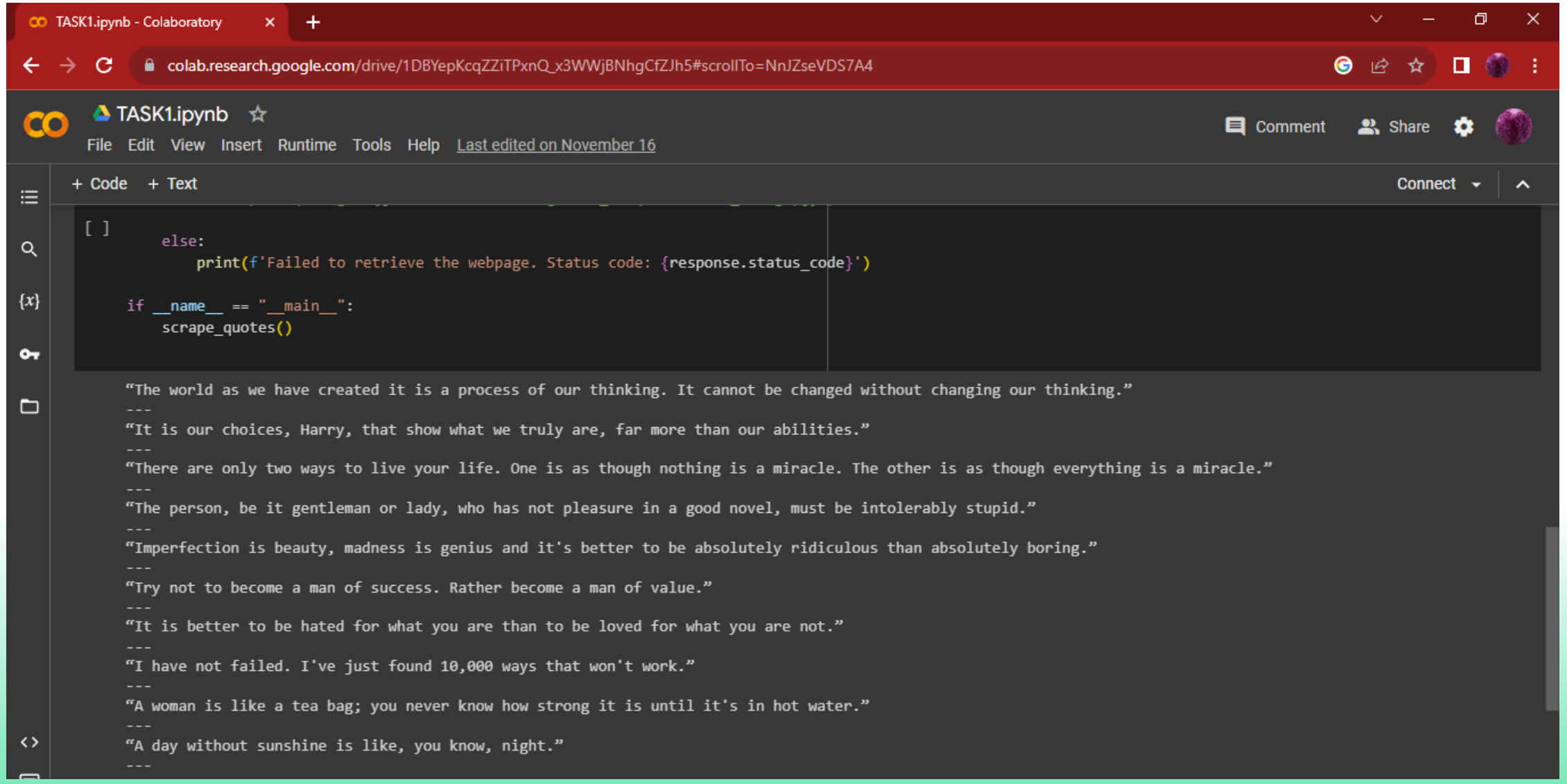
BASIC WEB SCRAPER

PYTHON DEVELOPMENT


TASK1.ipynb
File Edit View Insert Runtime Tools Help Last edited on November 16
Comment
Share
+ Code + Text
Connect
TASK : 1 BASIC WEB SCRAPER
import requests
from bs4 import BeautifulSoup
url = 'https://example.com'
response = requests.get(url)
if response.status_code == 200:
soup = BeautifulSoup(response.text, 'html.parser')
# Extract information by navigating the HTML structure
title = soup.title.text
paragraphs = soup.find_all('p')
# Print or use the extracted information as needed
print(f'Title: {title}')
for index, paragraph in enumerate(paragraphs):
print(f'Paragraph {index + 1}: {paragraph.text}')
else:
print(f'Failed to retrieve the webpage. Status code: {response.status_code}')
Title: Example Domain
Paragraph 1: This domain is for use in illustrative examples in documents. You may use this
domain in literature without prior coordination or asking for permission.
Paragraph 2: More information...


colab.research.google.com/drive/1DBYepKcqZZiTPxnQ_x3WWjBNhgCfZJh5#scrollTo=NnJZseVDS7A4
TASK1.ipynb
File Edit View Insert Runtime Tools Help Last edited on November 16
Comment
Share
+ Code + Text
Connect
Title: Example Domain
Paragraph 1: This domain is for use in illustrative examples in documents. You may use this
domain in literature without prior coordination or asking for permission.
Paragraph 2: More information...
[ ] import requests
from bs4 import BeautifulSoup
def scrape_quotes():
url = 'http://quotes.toscrape.com'
response = requests.get(url)
if response.status_code == 200:
soup = BeautifulSoup(response.text, 'html.parser')
# Extract information by navigating the HTML structure
for quote in soup.find_all('span', class_='text'):
print(quote.text)
print('---')
# You can extract other information similarly
# For example, extracting authors and tags:
# for author in soup.find_all('small', class_='author'):
# print(f'Author: {author.text}')
# for tag in soup.find_all('div', class_='tags'):
# print(f'Tags: {[t.text for t in tag.find_all("a", class_="tag")]}')
else:
print(f'Failed to retrieve the webpage. Status code: {response.status_code}')


TASK1.ipynb - Colaboratory
colab.research.google.com/drive/1DBYepKcqZZiTPxnQ_x3WWjBNhgCfZJh5#scrollTo=NnJZseVDS7A4
TASK1.ipynb
File Edit View Insert Runtime Tools Help Last edited on November 16
Comment
Share
+ Code + Text
Connect
[ ]
else:
print(f'Failed to retrieve the webpage. Status code: {response.status_code}')
if __name__ == "__main__":
scrape_quotes()
“The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking.”

“It is our choices, Harry, that show what we truly are, far more than our abilities.”

“There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.”

“The person, be it gentleman or lady, who has not pleasure in a good novel, must be intolerably stupid.”

“Imperfection is beauty, madness is genius and it's better to be absolutely ridiculous than absolutely boring.”

“Try not to become a man of success. Rather become a man of value.”

“It is better to be hated for what you are than to be loved for what you are not.”

“I have not failed. I've just found 10,000 ways that won't work.”

“A woman is like a tea bag; you never know how strong it is until it's in hot water.”

“A day without sunshine is like, you know, night.”

THANK YOU