

# quick\_LLM\_webscrapping\_and\_analysis

November 6, 2024

```
[ ]: # A Quick Exploration of the Huggingface LLM for Scrapping and Querying a Website
```

```
[9]: # install llamaindex packages and other needed libraries
```

```
'pip install llama-index llama-index-readers-web html2text IPython
```

Requirement already satisfied: llama-index in ./anaconda3/lib/python3.11/site-packages (0.11.21)

Requirement already satisfied: llama-index-readers-web in ./anaconda3/lib/python3.11/site-packages (0.2.4)

Requirement already satisfied: html2text in ./anaconda3/lib/python3.11/site-packages (2024.2.26)

Requirement already satisfied: IPython in ./anaconda3/lib/python3.11/site-packages (8.20.0)

Requirement already satisfied: llama-index-agent-openai<0.4.0,>=0.3.4 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.3.4)

Requirement already satisfied: llama-index-cli<0.4.0,>=0.3.1 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.3.1)

Requirement already satisfied: llama-index-core<0.12.0,>=0.11.20 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.11.21)

Requirement already satisfied: llama-index-embeddings-openai<0.3.0,>=0.2.4 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.2.5)

Requirement already satisfied: llama-index-indices-managed-llama-cloud>=0.3.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.4.0)

Requirement already satisfied: llama-index-legacy<0.10.0,>=0.9.48 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.9.48.post3)

Requirement already satisfied: llama-index-llms-openai<0.3.0,>=0.2.10 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.2.16)

Requirement already satisfied: llama-index-multi-modal-llms-openai<0.3.0,>=0.2.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.2.3)

Requirement already satisfied: llama-index-program-openai<0.3.0,>=0.2.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.2.0)

Requirement already satisfied: llama-index-question-gen-openai<0.3.0,>=0.2.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.2.0)

Requirement already satisfied: llama-index-readers-file<0.3.0,>=0.2.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.2.2)

Requirement already satisfied: llama-index-readers-llama-parse>=0.3.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index) (0.3.0)

Requirement already satisfied: nltk>3.8.1 in ./anaconda3/lib/python3.11/site-

packages (from llama-index) (3.9.1)  
 Requirement already satisfied: aiohttp<4.0.0,>=3.9.1 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (3.9.3)  
 Requirement already satisfied: beautifulsoup4<5.0.0,>=4.12.3 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (4.12.3)  
 Requirement already satisfied: chromedriver-autoinstaller<0.7.0,>=0.6.3 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (0.6.4)  
 Requirement already satisfied: newspaper3k<0.3.0,>=0.2.8 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (0.2.8)  
 Requirement already satisfied: playwright<2.0,>=1.30 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (1.48.0)  
 Requirement already satisfied: requests<3.0.0,>=2.31.0 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (2.31.0)  
 Requirement already satisfied: selenium<5.0.0,>=4.17.2 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (4.26.1)  
 Requirement already satisfied: spider-client<0.0.28,>=0.0.27 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web) (0.0.27)  
 Requirement already satisfied: urllib3>=1.1.0 in  
 ./anaconda3/lib/python3.11/site-packages (from llama-index-readers-web)  
 (1.26.18)  
 Requirement already satisfied: decorator in ./anaconda3/lib/python3.11/site-  
 packages (from IPython) (5.1.1)  
 Requirement already satisfied: jedi>=0.16 in ./anaconda3/lib/python3.11/site-  
 packages (from IPython) (0.18.1)  
 Requirement already satisfied: matplotlib-inline in  
 ./anaconda3/lib/python3.11/site-packages (from IPython) (0.1.6)  
 Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in  
 ./anaconda3/lib/python3.11/site-packages (from IPython) (3.0.43)  
 Requirement already satisfied: pygments>=2.4.0 in  
 ./anaconda3/lib/python3.11/site-packages (from IPython) (2.15.1)  
 Requirement already satisfied: stack-data in ./anaconda3/lib/python3.11/site-  
 packages (from IPython) (0.2.0)  
 Requirement already satisfied: traitlets>=5 in ./anaconda3/lib/python3.11/site-  
 packages (from IPython) (5.7.1)  
 Requirement already satisfied: pexpect>4.3 in ./anaconda3/lib/python3.11/site-  
 packages (from IPython) (4.8.0)  
 Requirement already satisfied: aiosignal>=1.1.2 in  
 ./anaconda3/lib/python3.11/site-packages (from aiohttp<4.0.0,>=3.9.1->llama-  
 index-readers-web) (1.2.0)  
 Requirement already satisfied: attrs>=17.3.0 in ./anaconda3/lib/python3.11/site-  
 packages (from aiohttp<4.0.0,>=3.9.1->llama-index-readers-web) (24.2.0)  
 Requirement already satisfied: frozenlist>=1.1.1 in  
 ./anaconda3/lib/python3.11/site-packages (from aiohttp<4.0.0,>=3.9.1->llama-  
 index-readers-web) (1.4.0)  
 Requirement already satisfied: multidict<7.0,>=4.5 in  
 ./anaconda3/lib/python3.11/site-packages (from aiohttp<4.0.0,>=3.9.1->llama-  
 index-readers-web) (6.0.4)  
 Requirement already satisfied: yarl<2.0,>=1.0 in

./anaconda3/lib/python3.11/site-packages (from aiohttp<4.0.0,>=3.9.1->llama-index-readers-web) (1.9.3)

Requirement already satisfied: soupsieve>1.2 in ./anaconda3/lib/python3.11/site-packages (from beautifulsoup4<5.0.0,>=4.12.3->llama-index-readers-web) (2.5)

Requirement already satisfied: packaging>=23.1 in ./anaconda3/lib/python3.11/site-packages (from chromedriver-autoinstaller<0.7.0,>=0.6.3->llama-index-readers-web) (24.1)

Requirement already satisfied: parso<0.9.0,>=0.8.0 in ./anaconda3/lib/python3.11/site-packages (from jedi>=0.16->IPython) (0.8.3)

Requirement already satisfied: openai>=1.14.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-agent-openai<0.4.0,>=0.3.4->llama-index) (1.54.0)

Requirement already satisfied: PyYAML>=6.0.1 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (6.0.1)

Requirement already satisfied: SQLAlchemy>=1.4.49 in ./anaconda3/lib/python3.11/site-packages (from SQLAlchemy[asyncio]>=1.4.49->llama-index-core<0.12.0,>=0.11.20->llama-index) (2.0.25)

Requirement already satisfied: dataclasses-json in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (0.6.7)

Requirement already satisfied: deprecated>=1.2.9.3 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (1.2.14)

Requirement already satisfied: dirtyjson<2.0.0,>=1.0.8 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (1.0.8)

Requirement already satisfied: fsspec>=2023.5.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (2023.10.0)

Requirement already satisfied: httpx in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (0.26.0)

Requirement already satisfied: nest-asyncio<2.0.0,>=1.5.8 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (1.6.0)

Requirement already satisfied: networkx>=3.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (3.1)

Requirement already satisfied: numpy<2.0.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (1.26.4)

Requirement already satisfied: pillow>=9.0.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (10.2.0)

Requirement already satisfied: pydantic<3.0.0,>=2.7.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (2.9.2)

Requirement already satisfied: tenacity!=8.4.0,<9.0.0,>=8.2.0 in ./anaconda3/lib/python3.11/site-packages (from llama-index-core<0.12.0,>=0.11.20->llama-index) (8.2.2)

Requirement already satisfied: tiktoken>=0.3.3 in

```

./anaconda3/lib/python3.11/site-packages (from llama-index-
core<0.12.0,>=0.11.20->llama-index) (0.8.0)
Requirement already satisfied: tqdm<5.0.0,>=4.66.1 in
./anaconda3/lib/python3.11/site-packages (from llama-index-
core<0.12.0,>=0.11.20->llama-index) (4.66.6)
Requirement already satisfied: typing-extensions>=4.5.0 in
./anaconda3/lib/python3.11/site-packages (from llama-index-
core<0.12.0,>=0.11.20->llama-index) (4.12.2)
Requirement already satisfied: typing-inspect>=0.8.0 in
./anaconda3/lib/python3.11/site-packages (from llama-index-
core<0.12.0,>=0.11.20->llama-index) (0.9.0)
Requirement already satisfied: wrapt in ./anaconda3/lib/python3.11/site-packages
(from llama-index-core<0.12.0,>=0.11.20->llama-index) (1.14.1)
Requirement already satisfied: llama-cloud>=0.0.11 in
./anaconda3/lib/python3.11/site-packages (from llama-index-indices-managed-
llama-cloud>=0.3.0->llama-index) (0.1.4)
Requirement already satisfied: pandas in ./anaconda3/lib/python3.11/site-
packages (from llama-index-legacy<0.10.0,>=0.9.48->llama-index) (2.1.4)
Requirement already satisfied: pypdf<5.0.0,>=4.0.1 in
./anaconda3/lib/python3.11/site-packages (from llama-index-readers-
file<0.3.0,>=0.2.0->llama-index) (4.3.1)
Requirement already satisfied: striprtf<0.0.27,>=0.0.26 in
./anaconda3/lib/python3.11/site-packages (from llama-index-readers-
file<0.3.0,>=0.2.0->llama-index) (0.0.26)
Requirement already satisfied: llama-parse>=0.5.0 in
./anaconda3/lib/python3.11/site-packages (from llama-index-readers-llama-
parse>=0.3.0->llama-index) (0.5.13)
Requirement already satisfied: cssselect>=0.9.2 in
./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-
index-readers-web) (1.2.0)
Requirement already satisfied: lxml>=3.6.0 in ./anaconda3/lib/python3.11/site-
packages (from newspaper3k<0.3.0,>=0.2.8->llama-index-readers-web) (4.9.3)
Requirement already satisfied: feedparser>=5.2.1 in
./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-
index-readers-web) (6.0.11)
Requirement already satisfied: tldextract>=2.0.1 in
./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-
index-readers-web) (3.2.0)
Requirement already satisfied: feedfinder2>=0.0.4 in
./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-
index-readers-web) (0.0.4)
Requirement already satisfied: jieba3k>=0.35.1 in
./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-
index-readers-web) (0.35.1)
Requirement already satisfied: python-dateutil>=2.5.3 in
./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-
index-readers-web) (2.8.2)
Requirement already satisfied: tinysegmenter==0.3 in

```

./anaconda3/lib/python3.11/site-packages (from newspaper3k<0.3.0,>=0.2.8->llama-index-readers-web) (0.3)

Requirement already satisfied: click in ./anaconda3/lib/python3.11/site-packages (from nltk>3.8.1->llama-index) (8.1.7)

Requirement already satisfied: joblib in ./anaconda3/lib/python3.11/site-packages (from nltk>3.8.1->llama-index) (1.2.0)

Requirement already satisfied: regex>=2021.8.3 in ./anaconda3/lib/python3.11/site-packages (from nltk>3.8.1->llama-index) (2023.10.3)

Requirement already satisfied: ptyprocess>=0.5 in ./anaconda3/lib/python3.11/site-packages (from pexpect>4.3->IPython) (0.7.0)

Requirement already satisfied: greenlet==3.1.1 in ./anaconda3/lib/python3.11/site-packages (from playwright<2.0,>=1.30->llama-index-readers-web) (3.1.1)

Requirement already satisfied: pyee==12.0.0 in ./anaconda3/lib/python3.11/site-packages (from playwright<2.0,>=1.30->llama-index-readers-web) (12.0.0)

Requirement already satisfied: wcwidth in ./anaconda3/lib/python3.11/site-packages (from prompt-toolkit<3.1.0,>=3.0.41->IPython) (0.2.5)

Requirement already satisfied: charset-normalizer<4,>=2 in ./anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.31.0->llama-index-readers-web) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in ./anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.31.0->llama-index-readers-web) (3.4)

Requirement already satisfied: certifi>=2017.4.17 in ./anaconda3/lib/python3.11/site-packages (from requests<3.0.0,>=2.31.0->llama-index-readers-web) (2024.8.30)

Requirement already satisfied: trio~=0.17 in ./anaconda3/lib/python3.11/site-packages (from selenium<5.0.0,>=4.17.2->llama-index-readers-web) (0.27.0)

Requirement already satisfied: trio-websocket~=0.9 in ./anaconda3/lib/python3.11/site-packages (from selenium<5.0.0,>=4.17.2->llama-index-readers-web) (0.11.1)

Requirement already satisfied: websocket-client~=1.8 in ./anaconda3/lib/python3.11/site-packages (from selenium<5.0.0,>=4.17.2->llama-index-readers-web) (1.8.0)

Requirement already satisfied: executing in ./anaconda3/lib/python3.11/site-packages (from stack-data->IPython) (0.8.3)

Requirement already satisfied: asttokens in ./anaconda3/lib/python3.11/site-packages (from stack-data->IPython) (2.0.5)

Requirement already satisfied: pure-eval in ./anaconda3/lib/python3.11/site-packages (from stack-data->IPython) (0.2.2)

Requirement already satisfied: six in ./anaconda3/lib/python3.11/site-packages (from feedfinder2>=0.0.4->newspaper3k<0.3.0,>=0.2.8->llama-index-readers-web) (1.16.0)

Requirement already satisfied: sgmlib3k in ./anaconda3/lib/python3.11/site-packages (from feedparser>=5.2.1->newspaper3k<0.3.0,>=0.2.8->llama-index-readers-web) (1.0.0)

Requirement already satisfied: anyio in ./anaconda3/lib/python3.11/site-packages (from httpx->llama-index-core<0.12.0,>=0.11.20->llama-index) (4.2.0)

Requirement already satisfied: httpcore==1.\* in ./anaconda3/lib/python3.11/site-packages (from httpx->llama-index-core<0.12.0,>=0.11.20->llama-index) (1.0.2)

Requirement already satisfied: sniffio in ./anaconda3/lib/python3.11/site-packages (from httpx->llama-index-core<0.12.0,>=0.11.20->llama-index) (1.3.0)

Requirement already satisfied: h11<0.15,>=0.13 in ./anaconda3/lib/python3.11/site-packages (from httpcore==1.\*->httpx->llama-index-core<0.12.0,>=0.11.20->llama-index) (0.14.0)

Requirement already satisfied: distro<2,>=1.7.0 in ./anaconda3/lib/python3.11/site-packages (from openai>=1.14.0->llama-index-agent-openai<0.4.0,>=0.3.4->llama-index) (1.8.0)

Requirement already satisfied: jiter<1,>=0.4.0 in ./anaconda3/lib/python3.11/site-packages (from openai>=1.14.0->llama-index-agent-openai<0.4.0,>=0.3.4->llama-index) (0.7.0)

Requirement already satisfied: annotated-types>=0.6.0 in ./anaconda3/lib/python3.11/site-packages (from pydantic<3.0.0,>=2.7.0->llama-index-core<0.12.0,>=0.11.20->llama-index) (0.7.0)

Requirement already satisfied: pydantic-core==2.23.4 in ./anaconda3/lib/python3.11/site-packages (from pydantic<3.0.0,>=2.7.0->llama-index-core<0.12.0,>=0.11.20->llama-index) (2.23.4)

Requirement already satisfied: requests-file>=1.4 in ./anaconda3/lib/python3.11/site-packages (from tldextract>=2.0.1->newspaper3k<0.3.0,>=0.2.8->llama-index-readers-web) (1.5.1)

Requirement already satisfied: filelock>=3.0.8 in ./anaconda3/lib/python3.11/site-packages (from tldextract>=2.0.1->newspaper3k<0.3.0,>=0.2.8->llama-index-readers-web) (3.13.1)

Requirement already satisfied: sortedcontainers in ./anaconda3/lib/python3.11/site-packages (from trio~0.17->selenium<5.0.0,>=4.17.2->llama-index-readers-web) (2.4.0)

Requirement already satisfied: outcome in ./anaconda3/lib/python3.11/site-packages (from trio~0.17->selenium<5.0.0,>=4.17.2->llama-index-readers-web) (1.3.0.post0)

Requirement already satisfied: wsproto>=0.14 in ./anaconda3/lib/python3.11/site-packages (from trio-websocket~0.9->selenium<5.0.0,>=4.17.2->llama-index-readers-web) (1.2.0)

Requirement already satisfied: mypy-extensions>=0.3.0 in ./anaconda3/lib/python3.11/site-packages (from typing-inspect>=0.8.0->llama-index-core<0.12.0,>=0.11.20->llama-index) (1.0.0)

Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in ./anaconda3/lib/python3.11/site-packages (from urllib3[socks]<3,>=1.26->selenium<5.0.0,>=4.17.2->llama-index-readers-web) (1.7.1)

Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in ./anaconda3/lib/python3.11/site-packages (from dataclasses-json->llama-index-core<0.12.0,>=0.11.20->llama-index) (3.23.1)

Requirement already satisfied: pytz>=2020.1 in ./anaconda3/lib/python3.11/site-packages (from pandas->llama-index-legacy<0.10.0,>=0.9.48->llama-index) (2022.7.1)

Requirement already satisfied: tzdata>=2022.1 in

./anaconda3/lib/python3.11/site-packages (from pandas->llama-index-legacy<0.10.0,>=0.9.48->llama-index) (2023.3)

```
[11]: # import libraries
from llama_index.core import SummaryIndex
from llama_index.readers.web import SimpleWebPageReader
from IPython.display import Markdown, display
from llama_index.core import VectorStoreIndex, SimpleDirectoryReader
import os
```

```
[13]: # access huggingface llm
# load huggingface token after setting up a huggingface account
from getpass import getpass
from huggingface_hub import login

# Prompt to securely enter the Hugging Face token
HF_TOKEN = getpass("Enter your Hugging Face token: ")

# Log in using the token
login(token=HF_TOKEN)
```

Enter your Hugging Face token: .....

The token has not been saved to the git credentials helper. Pass  
`add\_to\_git\_credential=True` in this function directly or `--add-to-git-credential` if using via `huggingface-cli` if you want to set the git credential as well.

Token is valid (permission: fineGrained).

Your token has been saved to /home/ijmg/.cache/huggingface/token

Login successful

```
[119]: # load url of website to be scraped and analyzed by llm
documents = SimpleWebPageReader(html_to_text=True).load_data(
    ["https://en.wikipedia.org/wiki/Twenty_Thousand_Leagues_Under_the_Seas"]
)
```

```
[121]: # index website
index = SummaryIndex.from_documents(documents)
```

```
[123]: # create llm model
from llama_index.llms.huggingface import HuggingFaceInferenceAPI
llm = HuggingFaceInferenceAPI(model_name="mistralai/Mixtral-8x7B-Instruct-v0.
    ↪1", token=HF_TOKEN)
llm
```

/tmp/ipykernel\_342557/1769814756.py:3: DeprecationWarning: Call to deprecated class HuggingFaceInferenceAPI. (Deprecated in favor of `HuggingFaceInferenceAPI` from `llama-index-llms-huggingface-api` which should be used instead.)

llm =

```
HuggingFaceInferenceAPI(model_name="mistralai/Mixtral-8x7B-Instruct-v0.1",
token=HF_TOKEN)
```

```
[123]: HuggingFaceInferenceAPI(callback_manager=<llama_index.core.callbacks.base.Callba
ckManager object at 0x7218fd0e8bd0>, system_prompt=None,
messages_to_prompt=<function messages_to_prompt at 0x7219b7550c20>,
completion_to_prompt=<function default_completion_to_prompt at 0x7219b74800e0>,
output_parser=None, pydantic_program_mode=<PydanticProgramMode.DEFAULT:
'default'>, query_wrapper_prompt=None,
model_name='mistralai/Mixtral-8x7B-Instruct-v0.1',
token='hf_CLGQUUn0nfOIQwQoxkCVDyEdJIL0eVqHWP', timeout=None, headers=None,
cookies=None, task=None, context_window=3900, num_output=256,
is_chat_model=False, is_function_calling_model=False)
```

```
[125]: # set up query engine using indexed website and huggingface llm
query_engine = index.as_query_engine(llm=llm)
```

```
[25]: # Task #1: Use huggingface llm to query website contents
response = query_engine.query("What is the Captain's name?")
print(response)
```

Captain Nemo, whose real name is Prince Dakkar, is a character in the 2012 film "Journey 2: The Mysterious Island". He is an Indian freedom fighter who leads a life of solitude beneath the seas and is the commander of the Nautilus, an advanced submarine. He is portrayed as a radical environmentalist in the film, taking action against the consequences of the destruction of the Earth. He has established a ZAD (Zone A Défendre) in the abysses of the seas, according to the film. Captain Nemo is also a character in Jules Verne's novel "Twenty Thousand Leagues Under the Sea" (1870). In the novel, he is a mysterious and complex character, who is both a scientific genius and a vengeful enemy of society. He is commander of the Nautilus, which uses to travel the seas and explore the depths of the ocean. His motives for his actions are unclear, and he remains an enigmatic figure throughout the novel.

Captain Nemo is a fictional character in Jules Verne's novel "Twenty Thousand

```
[27]: # Task #2: Use huggingface llm to query website contents
response = query_engine.query("What is the name of the vessel in the story?")
print(response)
```

The name of the vessel in the story is the Nautilus.

```
[127]: # Task #3: Use huggingface llm to query website contents
summary = query_engine.query("Which oceans or seas were visited in the story?")
summary_text = summary.text if hasattr(summary, 'text') else str(summary)
print(summary)
```



The story visits the Mediterranean Sea, Red Sea, Atlantic Ocean, Indian Ocean, China Sea, and the Arctic Ocean. The novel, *Twenty Thousand Leagues Under the Sea* by Jules Verne, is the source of the story. The expedition sails around Crete during the Cretan Revolt of 1866-1869, encounters Indian pearl divers in the Indian Ocean, and visits the Arctic Ocean during the search for the lost Franklin expedition. Captain Nemo, the commander of the Nautilus, engages in the protection of marine life. The novel has been adapted into various films, TV series, games, and series, including *Mysterious Island* (1961), *Mysterious Island* (2005), *Journey 2: The Mysterious Island* (2012), *Jules Verne's Mysterious Island* (2012), *La isla misteriosa y el capitán Nemo* (1973), *Captain Nemo* (1975), *Mysterious Island* (2012), and *\_2*

```
[29]: # Task #4: Use huggingface llm to summarize website contents
summary = query_engine.query("Summarize the contents of the document")
summary_text = summary.text if hasattr(summary, 'text') else str(summary)
print(summary)
```

*Twenty Thousand Leagues Under the Sea* is a science fiction novel by French author Jules Verne, first published in 1870. The novel follows Captain Nemo and his submarine, the Nautilus, as they travel 20,000 leagues under the sea. The story is divided into three parts, each focusing on different aspects of the journey, wonders of the deep sea, technology, and freedom. The 1871 Chelebourg edition of the text served as the basis for the translation, and includes 42 illustrations, published for the first time with Hetzel's illustrations and includes notes and a new introduction. Despite criticisms of technical accuracy, the novel is widely regarded as a classic of science fiction and has been translated into numerous languages. Jules Verne is also known for other works such as *The Mutineers of the Bounty* and the novel has inspired numerous adaptations, including films and games.

```
[129]: # Task #5: Use huggingface llm to query website contents
response = query_engine.query("What is the final resting place of the vessel in_
↳the story?")
print(response)
```

In the novel *"The Mysterious Island"* (1874-7) by Jules Verne, the Nautilus, the submarine captained by Captain Nemo, is scuttled by Nemo himself in the lagoon of the island where the submarine had been hidden for some time. This event occurs after Nemo's death, when the protagonists discover the submarine and its secrets. Nemo had previously mentioned that the Nautilus was his home and his grave, and he fulfills this prophecy by destroying it before his own death.

The Nautilus is scuttled in the lagoon of the island, which is located in the

Pacific Ocean, near the volcanic island of Tabor. The exact coordinates are not provided in the novel, but it is suggested that the island is near the Line Islands, which are part of Kiribati.

The Nautilus is scuttled in the lagoon of the island, which is located in the Pacific Ocean, near the volcanic island of Tabor. The exact coordinates are not provided in the novel, but it is suggested that the island is near the Line Islands,

[ ]: