```
import numpy as np
import warnings
warnings.filterwarnings('ignore')
import torch
import transformers
from transformers import AutoTokenizer, AutoModel, AutoModelForSequenceClassification
import semantic_search
```

Input a medical text:

```
input_text = input("""Please insert a medical text to be analyzed by the model: """)
```

```
print_text = input_text.replace('. ', '.\n')
print("You've inserted the following medical text:")
print(print_text)
```

```
You've inserted the following medical text:
The patient was a 34-yr-old man who presented with complaints of fever and a chronic cough.
He was a smoker and had a history of pulmonary tuberculosis that had been treated and cured.
A computed tomographic (CT) scan revealed multiple tiny nodules in both lungs.
A thoracoscopic lung biopsy was taken from the right upper lobe.
The microscopic examination revealed a typical LCH.
The tumor cells had vesicular and grooved nuclei, and they formed small aggregations around the bronchioles (Fig.1).
The tumor cells were strongly positive for S-100 protein, vimentin, CD68 and CD1a.
There were infiltrations of lymphocytes and eosinophils around the tumor cells.
With performing additional radiologic examinations, no other organs were thought to be involved.
He quit smoking, but he received no other specific treatment.
He was well for the following one year.
After this, a follow-up CT scan was performed and it showed a 4 cm-sized mass in the left lower lobe, in addition to the mul
A needle biopsy specimen revealed the possibility of a sarcoma; therefore, a lobectomy was performed.
Grossly, a 4 cm-sized poorly-circumscribed lobulated gray-white mass was found (Fig.3), and there were a few small satellite
Microscopically, the tumor cells were aggregated in large sheets and they showed an infiltrative growth.
The cytologic features of some of the tumor cells were similar to those seen in a typical LCH.
However, many tumor cells showed overtly malignant cytologic features such as pleomorphic/hyperchromatic nuclei and prominen
There were numerous mitotic figures ranging from 30 to 60 per 10 high power fields, and some of them were abnormal.
A few foci of typical LCH remained around the main tumor mass.
Immunohistochemically, the tumor cells were strongly positive for S-100 protein (Fig.5) and vimentin; they were also positiv
The ultrastructural analysis failed to demonstrate any Birbeck granules in the cytoplasm of the tumor cells.
Now, at five months after lobectomy, the patient is doing well with no significant change in the radiologic findings.
```

Named Entity Recognition

```
ner_model_name = "SahuH/distilbert-ner"
ner_tokenizer = AutoTokenizer.from_pretrained(ner_model_name)
ner_model = AutoModel.from_pretrained(ner_model_name)
```

```
tokenizer_config.json:                                  1.32k/1.32k [00:00<00:00,
100%                                                    39.4kB/s]
vocab.txt: 100%                                 232k/232k [00:00<00:00, 586kB/s]
tokenizer.json: 100%                                711k/711k [00:00<00:00, 19.7MB/s]
config.json: 100%                               5.00k/5.00k [00:00<00:00, 148kB/s]
model.safetensors: 100%                                 266M/266M [00:11<00:00, 26.5MB/s]
```

```
from transformers import pipeline
from transformers import AutoTokenizer, AutoModelForTokenClassification

tokenizer = AutoTokenizer.from_pretrained(ner_model_name)
model = AutoModelForTokenClassification.from_pretrained(ner_model_name)

pipe = pipeline("ner", model=model, tokenizer=tokenizer, aggregation_strategy="simple")
output = pipe(input_text)
```

```
!python -m spacy download en_core_web_sm
```

```
import spacy
from spacy import displacy

nlp = spacy.load('en_core_web_sm')
doc = nlp(input_text)

ents = []
for d in output:
    ent = doc.char_span(d['start'], d['end'], label=d['entity_group'])
    if ent is None:
        continue

    ents.append(ent)

doc.ents = ents

displacy.render(doc, style="ent", jupyter=True)
```

The patient was a 34-yr-old **Age** man **Sex** who presented **Clinical_event** with complaints of fever **Sign_symptom** and a chronic cough. He was a smoker **History** and had a history of pulmonary tuberculosis **History** that had been treated and cured. A computed tomographic **Diagnostic_procedure** (**Diagnostic_procedure** CT) **Diagnostic_procedure** scan revealed multiple **Detailed_description** tiny **Detailed_description** nodules **Sign_symptom** in both lungs. A thoracoscopic lung **Biological_structure** biopsy **Diagnostic_procedure** was taken from the right upper lobe. The microscopic examination **Diagnostic_procedure** revealed a typical LCH. The tumor cells **Sign_symptom** had vesicular and grooved nuclei, and they formed small **Detailed_description** aggregations around the bronchioles **Biological_structure** (Fig.1). The tumor **Sign_symptom** cells were strongly positive **Lab_value** for S-100 **Diagnostic_procedure** protein, vimentin, CD68 and CD1a. There were infiltrations of lymphocytes and eosinophils **Biological_structure** around the tumor **Sign_symptom** cells. With performing additional radiologic examinations, no other **Biological_structure** organs were thought to be involved. He quit smoking, but he received no other specific treatment. He was well **Sign_symptom** for the following one year. After this, a follow-up CT **Diagnostic_procedure** scan was performed and it showed a 4 cm **Shape** -sized mass **Sign_symptom** in the left **Biological_structure** lower lobe, in addition to the multiple **Detailed_description** tiny **Detailed_description** nodules **Sign_symptom** in both lungs **Biological_structure** (Fig.2). A needle biopsy **Diagnostic_procedure** specimen revealed the possibility of a sarcoma **Sign_symptom** ; therefore, a lobectomy was performed. Grossly, a 4 cm **Detailed_description** -sized poorly **Detailed_description** - circumscribed **Detailed_description** lobulated **Detailed_description** gray-white **Color** mass **Sign_symptom** was found (Fig.3), and there were a few small **Detailed_description** satellite **Detailed_description** nodules **Sign_symptom** around the main mass. Microscopically, the tumor **Sign_symptom** cells were aggregated in large sheets **Detailed_description** and they showed an infiltrative growth. The cytologic features **Diagnostic_procedure** of some of the tumor **Sign_symptom** cells were similar to those seen in a typical LCH. However, many tumor **Sign_symptom** cells showed overtly malignant cytologic features such as pleomorphic/hyperchromatic nuclei and prominent nucleoli (Fig.4), and multinucleated tumor giant cells **Sign_symptom** were also found. There were numerous mitotic figures **Sign_symptom** ranging from 30 to **Lab_value** 60 per 10 high **Lab_value** power fields, and some of them were abnormal. A few foci of typical LCH **Sign_symptom** remained around the main tumor **Sign_symptom** mass. Immunohistochemically, the tumor **Sign_symptom** cells were strongly positive **Lab_value** for S-100 protein **Diagnostic_procedure** (Fig.5) and vimentin;

Relevant Information retrieval from Medical Text

```
input_question = input("""Please input a query to retrieve relevant information from the given text: """)

print("You've given the following query:")
print(input_question)
```

```
You've given the following query:
Which disease was detected in patient through biopsy?
```

```
relevant_info = semantic_search.semantic_search(input_text, input_question)
```

```
Map: 100%  22/22 [00:01<00:00, 10.16 examples/s]
100%  1/1 [00:00<00:00, 32.03it/s]
```

```
print("The following information was retrieved from the given text based on the query:")
print(relevant_info)
```

```
The following information was retrieved from the given text based on the query:
A needle biopsy specimen revealed the possibility of a sarcoma; therefore, a lobectomy was performed.
With performing additional radiologic examinations, no other organs were thought to be involved.
The microscopic examination revealed a typical LCH.
Now, at five months after lobectomy, the patient is doing well with no significant change in the radiologic findings..
A computed tomographic (CT) scan revealed multiple tiny nodules in both lungs.
Microscopically, the tumor cells were aggregated in large sheets and they showed an infiltrative growth.
After this, a follow-up CT scan was performed and it showed a 4 cm–sized mass in the left lower lobe, in addition to the mul
A thoracoscopic lung biopsy was taken from the right upper lobe.
The cytologic features of some of the tumor cells were similar to those seen in a typical LCH.
The ultrastructural analysis failed to demonstrate any Birbeck granules in the cytoplasm of the tumor cells.
A few foci of typical LCH remained around the main tumor mass.
There were numerous mitotic figures ranging from 30 to 60 per 10 high power fields, and some of them were abnormal.
He quit smoking, but he received no other specific treatment.
There were infiltrations of lymphocytes and eosinophils around the tumor cells.
He was a smoker and had a history of pulmonary tuberculosis that had been treated and cured
```

Information Summarization

Here, we will summarize the relevant information retrieved in previous part

```
from transformers import pipeline

summarizer = pipeline("summarization", model="t5-base")
output = summarizer(relevant_info, min_length=30, max_length=256)
```

```
config.json: 100%                                  1.21k/1.21k [00:00<00:00, 15.8kB/s]
model.safetensors: 100%                            892M/892M [00:11<00:00, 85.4MB/s]
generation_config.json: 100%                       147/147 [00:00<00:00, 6.58kB/s]
spiece.model: 100%                                 792k/792k [00:00<00:00, 16.7MB/s]
tokenizer.json: 100%                               1.39M/1.39M [00:00<00:00, 1.43MB/s]
```

```
final_output = output[0]["summary_text"].replace('. ', '.\n')
print("A concise summary of the relevant information is: ")
print(final_output)
```

```
A concise summary summary of the relevant information is:
a needle biopsy specimen revealed the possibility of a sarcoma .
after lobectomy, no other organs were thought to be involved .
the patient is now doing well with no significant change in the radiologic findings .
```