

Phần 1

```
# Cài đặt spaCy
!pip install spacy

# Tải về mô hình tiếng Anh (kích thước trung bình, có đủ thông tin cho parsing)
!python -m spacy download en_core_web_md

Requirement already satisfied: spacy in c:\users\admin\anaconda3\lib\site-packages (3.8.11)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (1.0.15)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (2.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (3.0.12)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (8.3.10)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (2.5.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (0.4.3)
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (0.20.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (4.67.1)
Requirement already satisfied: numpy>=1.19.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (2.1.3)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (2.32.3)
Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (2.10.3)
Requirement already satisfied: jinja2 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (3.1.6)
Requirement already satisfied: setuptools in c:\users\admin\anaconda3\lib\site-packages (from spacy) (72.1.0)
Requirement already satisfied: packaging>=20.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (24.2)
Requirement already satisfied: annotated-types>=0.6.0 in c:\users\
```

```

admin\anaconda3\lib\site-packages (from pydantic!=1.8,!
=1.8.1,<3.0.0,>=1.7.4->spacy) (0.6.0)
Requirement already satisfied: pydantic-core==2.27.1 in c:\users\
admin\anaconda3\lib\site-packages (from pydantic!=1.8,!
=1.8.1,<3.0.0,>=1.7.4->spacy) (2.27.1)
Requirement already satisfied: typing-extensions>=4.12.2 in c:\users\
admin\anaconda3\lib\site-packages (from pydantic!=1.8,!
=1.8.1,<3.0.0,>=1.7.4->spacy) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\
admin\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0-
>spacy) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\admin\
anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy)
(3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\admin\
anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy)
(2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\admin\
anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy)
(2025.4.26)
Requirement already satisfied: blis<1.4.0,>=1.3.0 in c:\users\admin\
anaconda3\lib\site-packages (from thinc<8.4.0,>=8.3.4->spacy) (1.3.3)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\
admin\anaconda3\lib\site-packages (from thinc<8.4.0,>=8.3.4->spacy)
(0.1.5)
Requirement already satisfied: colorama in c:\users\admin\anaconda3\
lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: click>=8.0.0 in c:\users\admin\
anaconda3\lib\site-packages (from typer-slim<1.0.0,>=0.3.0->spacy)
(8.1.8)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in c:\users\
admin\anaconda3\lib\site-packages (from weasel<0.5.0,>=0.4.2->spacy)
(0.23.0)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in c:\users\
admin\anaconda3\lib\site-packages (from weasel<0.5.0,>=0.4.2->spacy)
(7.5.0)
Requirement already satisfied: wrapt in c:\users\admin\anaconda3\lib\
site-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.4.2-
>spacy) (1.17.0)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\admin\
anaconda3\lib\site-packages (from jinja2->spacy) (3.0.2)
Collecting en-core-web-md==3.8.0
  Downloading
https://github.com/explosion/spacy-models/releases/download/en_core_we
b_md-3.8.0/en_core_web_md-3.8.0-py3-none-any.whl (33.5 MB)
----- 0.0/33.5 MB ? eta
-:--:--
----- 0.0/33.5 MB ? eta
-:--:--

```

```
----- 0.3/33.5 MB ? eta
-:--:--
----- 0.5/33.5 MB 1.0 MB/s eta
0:00:32
----- 0.8/33.5 MB 1.0 MB/s eta
0:00:32
----- 0.8/33.5 MB 1.0 MB/s eta
0:00:32
----- 1.0/33.5 MB 956.9 kB/s
eta 0:00:34
----- 1.6/33.5 MB 1.1 MB/s eta
0:00:29
----- 1.8/33.5 MB 1.2 MB/s eta
0:00:27
----- 2.1/33.5 MB 1.3 MB/s eta
0:00:26
----- 2.4/33.5 MB 1.2 MB/s eta
0:00:26
----- 2.9/33.5 MB 1.3 MB/s eta
0:00:24
----- 3.1/33.5 MB 1.3 MB/s eta
0:00:24
----- 3.4/33.5 MB 1.3 MB/s eta
0:00:23
----- 3.7/33.5 MB 1.3 MB/s eta
0:00:23
----- 4.2/33.5 MB 1.4 MB/s eta
0:00:22
----- 4.5/33.5 MB 1.4 MB/s eta
0:00:22
----- 4.7/33.5 MB 1.4 MB/s eta
0:00:21
----- 5.2/33.5 MB 1.4 MB/s eta
0:00:21
----- 5.5/33.5 MB 1.4 MB/s eta
0:00:20
----- 5.8/33.5 MB 1.4 MB/s eta
0:00:20
----- 6.0/33.5 MB 1.4 MB/s eta
0:00:20
----- 6.6/33.5 MB 1.4 MB/s eta
0:00:19
----- 6.6/33.5 MB 1.4 MB/s eta
0:00:19
----- 6.6/33.5 MB 1.4 MB/s eta
0:00:19
----- 6.6/33.5 MB 1.4 MB/s eta
0:00:19
----- 6.6/33.5 MB 1.4 MB/s eta
```

0:00:19	-----	6.6/33.5 MB	1.4 MB/s	eta
0:00:19	-----	6.6/33.5 MB	1.4 MB/s	eta
0:00:19	-----	6.6/33.5 MB	1.4 MB/s	eta
0:00:19	-----	6.8/33.5 MB	1.1 MB/s	eta
0:00:25	-----	7.1/33.5 MB	1.1 MB/s	eta
0:00:25	-----	7.6/33.5 MB	1.1 MB/s	eta
0:00:24	-----	7.9/33.5 MB	1.1 MB/s	eta
0:00:23	-----	8.4/33.5 MB	1.2 MB/s	eta
0:00:22	-----	8.7/33.5 MB	1.2 MB/s	eta
0:00:22	-----	8.9/33.5 MB	1.2 MB/s	eta
0:00:21	-----	9.2/33.5 MB	1.2 MB/s	eta
0:00:21	-----	9.7/33.5 MB	1.2 MB/s	eta
0:00:20	-----	10.0/33.5 MB	1.2 MB/s	
eta 0:00:20	-----	10.5/33.5 MB	1.2 MB/s	
eta 0:00:19	-----	10.7/33.5 MB	1.2 MB/s	
eta 0:00:19	-----	11.3/33.5 MB	1.3 MB/s	
eta 0:00:18	-----	11.5/33.5 MB	1.3 MB/s	
eta 0:00:18	-----	12.1/33.5 MB	1.3 MB/s	
eta 0:00:17	-----	12.3/33.5 MB	1.3 MB/s	
eta 0:00:17	-----	12.6/33.5 MB	1.3 MB/s	
eta 0:00:17	-----	13.1/33.5 MB	1.3 MB/s	
eta 0:00:16	-----	13.4/33.5 MB	1.3 MB/s	
eta 0:00:16	-----	13.9/33.5 MB	1.3 MB/s	
eta 0:00:15	-----	14.2/33.5 MB	1.3 MB/s	
eta 0:00:15				

	-----	14.7/33.5 MB 1.4 MB/s
eta 0:00:14	-----	
	-----	15.2/33.5 MB 1.4 MB/s
eta 0:00:14	-----	
	-----	15.5/33.5 MB 1.4 MB/s
eta 0:00:14	-----	
	-----	16.0/33.5 MB 1.4 MB/s
eta 0:00:13	-----	
	-----	16.3/33.5 MB 1.4 MB/s
eta 0:00:13	-----	
	-----	16.8/33.5 MB 1.4 MB/s
eta 0:00:12	-----	
	-----	17.0/33.5 MB 1.4 MB/s
eta 0:00:12	-----	
	-----	17.6/33.5 MB 1.4 MB/s
eta 0:00:12	-----	
	-----	17.8/33.5 MB 1.4 MB/s
eta 0:00:11	-----	
	-----	18.4/33.5 MB 1.4 MB/s
eta 0:00:11	-----	
	-----	18.6/33.5 MB 1.4 MB/s
eta 0:00:11	-----	
	-----	19.1/33.5 MB 1.4 MB/s
eta 0:00:10	-----	
	-----	19.4/33.5 MB 1.4 MB/s
eta 0:00:10	-----	
	-----	19.9/33.5 MB 1.5 MB/s
eta 0:00:10	-----	
	-----	20.2/33.5 MB 1.5 MB/s
eta 0:00:10	-----	
	-----	20.2/33.5 MB 1.5 MB/s
eta 0:00:10	-----	
	-----	21.0/33.5 MB 1.5 MB/s
eta 0:00:09	-----	
	-----	21.2/33.5 MB 1.5 MB/s
eta 0:00:09	-----	
	-----	21.8/33.5 MB 1.5 MB/s
eta 0:00:08	-----	
	-----	22.3/33.5 MB 1.5 MB/s
eta 0:00:08	-----	
	-----	22.5/33.5 MB 1.5 MB/s
eta 0:00:08	-----	
	-----	23.1/33.5 MB 1.5 MB/s
eta 0:00:07	-----	
	-----	23.3/33.5 MB 1.5 MB/s
eta 0:00:07	-----	
	-----	23.9/33.5 MB 1.5 MB/s
eta 0:00:07	-----	
	-----	24.1/33.5 MB 1.5 MB/s

```
eta 0:00:07
----- 24.6/33.5 MB 1.5 MB/s
eta 0:00:06
----- 25.2/33.5 MB 1.5 MB/s
eta 0:00:06
----- 25.4/33.5 MB 1.5 MB/s
eta 0:00:06
----- 26.0/33.5 MB 1.5 MB/s
eta 0:00:05
----- 26.2/33.5 MB 1.5 MB/s
eta 0:00:05
----- 26.7/33.5 MB 1.5 MB/s
eta 0:00:05
----- 27.3/33.5 MB 1.5 MB/s
eta 0:00:05
----- 27.5/33.5 MB 1.6 MB/s
eta 0:00:04
----- 27.8/33.5 MB 1.6 MB/s
eta 0:00:04
----- 28.3/33.5 MB 1.6 MB/s
eta 0:00:04
----- 28.8/33.5 MB 1.6 MB/s
eta 0:00:03
----- 29.1/33.5 MB 1.6 MB/s
eta 0:00:03
----- 29.6/33.5 MB 1.6 MB/s
eta 0:00:03
----- 29.9/33.5 MB 1.6 MB/s
eta 0:00:03
----- 30.4/33.5 MB 1.6 MB/s
eta 0:00:02
----- 30.9/33.5 MB 1.6 MB/s
eta 0:00:02
----- 31.5/33.5 MB 1.6 MB/s
eta 0:00:02
----- 31.7/33.5 MB 1.6 MB/s
eta 0:00:02
----- 32.2/33.5 MB 1.6 MB/s
eta 0:00:01
----- 32.5/33.5 MB 1.6 MB/s
eta 0:00:01
----- 32.5/33.5 MB 1.6 MB/s
eta 0:00:01
----- 32.8/33.5 MB 1.6 MB/s
eta 0:00:01
----- 33.0/33.5 MB 1.6 MB/s
eta 0:00:01
----- 33.3/33.5 MB 1.6 MB/s
eta 0:00:01
```

```
----- 33.5/33.5 MB 1.6 MB/s
eta 0:00:00
Installing collected packages: en-core-web-md
Successfully installed en-core-web-md-3.8.0
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_md')
```

Phần 2

2.1

```
import spacy
from spacy import displacy

# Ta'í mô hình tiếng Anh đã cài đặt
# Sử dụng en_core_web_md vì nó chứa các vector từ và cây cú pháp đầy đủ
nlp = spacy.load("en_core_web_md")

# Câu ví dụ
text = "The quick brown fox jumps over the lazy dog."

# Phân tích câu với pipeline của spaCy
doc = nlp(text)
```

2.2

```
# Tùy chọn để hiển thị trong trình duyệt
# options = {"compact": True, "color": "blue", "font": "Source Sans Pro"}

# Khởi chạy server tại http://127.0.0.1:5000
# Bạn có thể truy cập địa chỉ này trên trình duyệt để xem cây phụ thuộc
# Nhấn Ctrl+C trong terminal để dừng server
displacy.serve(doc, style="dep")

c:\Users\Admin\anaconda3\Lib\site-packages\spacy\displacy\
__init__.py:108: UserWarning: [W011] It looks like you're calling
displacy.serve from within a Jupyter notebook or a similar
environment. This likely means you're already running a local web
server, so there's no need to make displaCy start another one.
Instead, you should be able to replace displacy.serve with
displacy.render to show the visualization.
  warnings.warn(Warnings.W011)

<IPython.core.display.HTML object>
```

```
Using the 'dep' visualizer
Serving on http://0.0.0.0:5000 ...
```

```
127.0.0.1 - - [04/Dec/2025 07:37:11] "GET / HTTP/1.1" 200 7538
127.0.0.1 - - [04/Dec/2025 07:37:11] "GET /favicon.ico HTTP/1.1" 200
7538
```

```
Shutting down server on port 5000.
```

- "jumps" là từ gốc của câu.
- "jumps" có các phụ thuộc là "fox" (nsubj) và "over" (prep)
- "fox" là head của các từ "The", "quick" và "brown"

Phần 3

```
# Lấy một câu khác để phân tích
```

```
text = "Apple is looking at buying U.K. startup for $1 billion"
doc = nlp(text)
```

```
# In ra thông tin cu' a từng token
```

```
print(f"{'TEXT':<12} | {'DEP':<10} | {'HEAD TEXT':<12} | {'HEAD POS':<8} | {'CHILDREN'}")
print("-" * 70)
```

```
for token in doc:
```

```
# Trích xuất các thuộc tính
```

```
    children = [child.text for child in token.children]
    print(f"{token.text:<12} | {token.dep_:<10} | {token.head.text:<12} | {token.head.pos_:<8} | {children}")
```

TEXT	DEP	HEAD TEXT	HEAD POS	CHILDREN
Apple	nsubj	looking	VERB	[]
is	aux	looking	VERB	[]
looking	ROOT	looking	VERB	['Apple', 'is', 'at']
at	prep	looking	VERB	['buying']
buying	pcomp	at	ADP	['startup']
U.K.	compound	startup	NOUN	[]
startup	dobj	buying	VERB	['U.K.', 'for']
for	prep	startup	NOUN	['billion']
\$	quantmod	billion	NUM	[]
1	compound	billion	NUM	[]
billion	pobj	for	ADP	['\$', '1']

Phần 4

4.1

```
text = "The cat chased the mouse and the dog watched them."
doc = nlp(text)

for token in doc:
    # Chỉ tìm các động từ
    if token.pos_ == "VERB":
        verb = token.text
        subject = ""
        obj = ""

        # Tìm chủ ngữ (nsubj) và tân ngữ (dobj) trong các con cu'a
        động từ
        for child in token.children:
            if child.dep_ == "nsubj":
                subject = child.text
            if child.dep_ == "dobj":
                obj = child.text
        if subject and obj:
            print(f"Found Triplet: ({subject}, {verb}, {obj})")
```

```
Found Triplet: (cat, chased, mouse)
Found Triplet: (dog, watched, them)
```

4.2

```
text = "The big, fluffy white cat is sleeping on the warm mat."
doc = nlp(text)

for token in doc:
    # Chỉ tìm các danh từ
    if token.pos_ == "NOUN":
        adjectives = []
        # Tìm các tính từ bổ nghĩa (amod) trong các con cu'a danh từ
        for child in token.children:
            if child.dep_ == "amod":
                adjectives.append(child.text)

        if adjectives:
            print(f"Danh từ '{token.text}' được bổ nghĩa bởi các tính từ: {adjectives}")
```

```
Danh từ 'cat' được bổ nghĩa bởi các tính từ: ['big', 'fluffy', 'white']
```

```
Danh từ 'mat' được bổ nghĩa bởi các tính từ: ['warm']
```

Phần 5

```
sentences = [  
    "Yesterday I set the house on fire",  
    "Mom beat me almost half-dead",  
    "Today Mom went up to the fields",  
    "So I burned it again, all by myself",  
    "The forest smells of smoke",  
    "The stream turns pitch black",  
    "My whole village burns down",  
    "I'm sentenced to life in prison"  
]
```

Bài 1

```
def find_main_verb(sentence):  
    doc = nlp(sentence)  
    for token in doc:  
        if token.dep_ == "ROOT" and token.pos_ == "VERB":  
            return token.text  
    return None  
  
for sentence in sentences:  
    print(f"Sentence: {sentence}")  
    main_verb = find_main_verb(sentence)  
    print(f"Main Verb: {main_verb}")
```

```
Sentence: Yesterday I set the house on fire  
Main Verb: set  
Sentence: Mom beat me almost half-dead  
Main Verb: beat  
Sentence: Today Mom went up to the fields  
Main Verb: went  
Sentence: So I burned it again, all by myself  
Main Verb: burned  
Sentence: The forest smells of smoke  
Main Verb: smells  
Sentence: The stream turns pitch black  
Main Verb: turns  
Sentence: My whole village burns down  
Main Verb: burns  
Sentence: I'm sentenced to life in prison  
Main Verb: sentenced
```

Bài 2

```
def get_noun_chunks(sentence):  
    doc = nlp(sentence)  
    noun_chunks = []
```

```

    for token in doc:
        # Check if token is a noun and not already part of a recorded
        noun chunk
        if token.pos_ == "NOUN" and not any(token.i in chunk for chunk
in noun_chunks):
            chunk = [token.text]
            # Include modifiers (amod, det, compound)
            for child in token.children:
                if child.dep_ in ("amod", "det", "compound"):
                    chunk.insert(0, child.text)
            noun_chunks.append(chunk)
    return [" ".join(chunk) for chunk in noun_chunks]

for sentence in sentences:
    chunks = get_noun_chunks(sentence)
    print(f"Sentence: {sentence}")
    print(f"Noun Chunks: {chunks}")

```

```

Sentence: Yesterday I set the house on fire
Noun Chunks: ['Yesterday', 'the house', 'fire']
Sentence: Mom beat me almost half-dead
Noun Chunks: ['Mom']
Sentence: Today Mom went up to the fields
Noun Chunks: ['Today', 'the fields']

```

```

Sentence: So I burned it again, all by myself
Noun Chunks: []
Sentence: The forest smells of smoke
Noun Chunks: ['forest', 'smoke']
Sentence: The stream turns pitch black
Noun Chunks: ['The stream', 'pitch']
Sentence: My whole village burns down
Noun Chunks: ['whole village']
Sentence: I'm sentenced to life in prison
Noun Chunks: ['life', 'prison']

```

Bài 3

```

def get_path_to_root(token):
    path = []
    current = token
    while current.head != current:
        path.append(current.text)
        current = current.head
    path.append("ROOT")
    return " -> ".join(path)

doc = nlp("The quick brown fox jumps over the lazy dog.")
for token in doc:
    path = get_path_to_root(token)

```

```
print(f"Token: {token.text}")  
print(f"Path to ROOT: {path}")
```

```
Token: The  
Path to ROOT: The -> fox -> ROOT  
Token: quick  
Path to ROOT: quick -> fox -> ROOT  
Token: brown  
Path to ROOT: brown -> fox -> ROOT  
Token: fox  
Path to ROOT: fox -> ROOT  
Token: jumps  
Path to ROOT: ROOT  
Token: over  
Path to ROOT: over -> ROOT  
Token: the  
Path to ROOT: the -> dog -> over -> ROOT  
Token: lazy  
Path to ROOT: lazy -> dog -> over -> ROOT  
Token: dog  
Path to ROOT: dog -> over -> ROOT  
Token: .  
Path to ROOT: . -> ROOT
```