INFO5731 Assignment 2

In this assignment, you will work on gathering text data from an open data source via web scraping or API. Following this, you will need to clean the text data and perform syntactic analysis on the data. Follow the instructions carefully and design well-structured Python programs to address each question.

Expectations:

- Use the provided *.ipynb* document to write your code & respond to the questions. Avoid generating a new file.
- Write complete answers and run all the cells before submission.
- Make sure the submission is "clean"; *i.e.*, no unnecessary code cells.
- Once finished, allow shared rights from top right corner (*see Canvas for details*).
- **Make sure to submit the cleaned data CSV in the comment section - 10 points**

Total points: 100

Deadline: Tuesday, at 11:59 PM.

Late Submission will have a penalty of 10% reduction for each day after the deadline.

Please check that the link you submitted can be opened and points to the correct assignment.

Question 1 (40 points)

Write a python program to collect text data from **either of the following sources** and save the data into a **csv file:**

(1) Collect all the customer reviews of a product (you can choose any porduct) on amazon. [atleast 1000 reviews]

(2) Collect the top 1000 User Reviews of a movie recently in 2023 or 2024 (you can choose any movie) from IMDB. [If one movie doesn't have sufficient reviews, collect reviews of atleast 2 or 3 movies]

(3) Collect all the reviews of the top 1000 most popular software from G2 or Capterra.

(4) Collect the **abstracts** of the top 10000 research papers by using the query "machine learning", "data science", "artifical intelligence", or "information extraction" from Semantic Scholar.

(5) Collect all the information of the 904 narrators in the Densho Digital Repository.

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

```
import requests
from bs4 import BeautifulSoup
import pandas as pd

imdb_base_url = f'https://www.imdb.com/title/tt15239678/reviews?ref_=tt_urv'  #imdb reviews
total_reviews = int(input("number of reviews to scrap: "))
extracted_reviews = []
reviews_per_page = 25
iterations = (total_reviews // reviews_per_page) + 1
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.121 Safari/537.36'
}

for current_page in range(1, iterations * reviews_per_page, reviews_per_page):
    page_url = f'{imdb_base_url}&start={current_page}'
    response = requests.get(page_url)
    page_content = BeautifulSoup(response.content, 'html.parser')
    review_blocks = page_content.find_all('div', class_='text show-more__control')
    extracted_reviews.extend([block.text.strip() for block in review_blocks])
    if len(extracted_reviews) >= total_reviews:
        break

extracted_reviews = extracted_reviews[:total_reviews]
dune_reviews = extracted_reviews #Scraping user reviews
print(f"Data scraping successful.")
reviews_df = pd.DataFrame( dune_reviews, columns=['reviews'])
```

```
reviews_df
reviews_df.to_csv('dune_reviews.csv', index=False)
print(' File created successfully')
```

```
number of reviews to scrap: 200
Data scraping successful.
 File created successfully
```

```
data_url="https://raw.githubusercontent.com/Mukeshreddy3699/Mukesh_INFO5731/refs/heads/main/Dune_reviews.csv"
df = pd.read_table(data_url,names=['text'])
df
```

	text
0	reviews
1	This is what Hollywood needs. A great story wi...
2	I'm going to write this as a review for both D...
3	Had the pleasure to watch this film in an earl...
4	Phenomenal stuff. I'll probably calm down tomo...
...	...
996	Saw an early screening of this film at the Til...
997	The biggest problem with this movie is one tha...
998	The Bad1. The dialogue is clunky and robotic w...
999	I just watched this in IMAX and it was one of ...
1000	This movie is a visual masterpiece that will b...

1001 rows × 1 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

Question 2 (30 points)

Write a python program to **clean the text data** you collected in the previous question and save the clean data in a new column in the csv file. The data cleaning steps include: [Code and output is required for each part]

(1) Remove noise, such as special characters and punctuations.

(2) Remove numbers.

(3) Remove stopwords by using the stopwords list.

(4) Lowercase all texts

(5) Stemming.

(6) Lemmatization.

```
import pandas as pd
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer, WordNetLemmatizer
import string
import re

nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')
```

Show hidden output

```
#1)Remove Noise function
def remove_noise(text):
    clean_text = re.sub('[^a-zA-Z0-9]', ' ', text)
    return clean_text
df['clean_text'] = df['text'].apply(remove_noise)
print("\nData Frame after removing noise:")   #Displaying the Data Frame after removing noise
df
```

Data Frame after removing noise:

	text	clean_text
0	reviews	reviews
1	This is what Hollywood needs. A great story wi...	This is what Hollywood needs A great story wi...
2	I'm going to write this as a review for both D...	I m going to write this as a review for both D...
3	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...
4	Phenomenal stuff. I'll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...
...	...	...
996	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...
997	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...
998	The Bad1. The dialogue is clunky and robotic w...	The Bad1 The dialogue is clunky and robotic w...
999	I just watched this in IMAX and it was one of ...	I just watched this in IMAX and it was one of ...
1000	This movie is a visual masterpiece that will b...	This movie is a visual masterpiece that will b...

1001 rows × 2 columns

```
#2)Remove Numbers
def remove_numbers(text):
    clean_text = re.sub(r'\d+', '', text)
    return clean_text
df['clean_text_remove_numbers'] = df['clean_text'].apply(remove_numbers)
print("\nData Frame after removing numbers:")
df
```

Data Frame after removing numbers:

	text	clean_text	clean_text_remove_numbers
0	reviews	reviews	reviews
1	This is what Hollywood needs. A great story wi...	This is what Hollywood needs A great story wi...	This is what Hollywood needs A great story wi...
2	I'm going to write this as a review for both D...	I m going to write this as a review for both D...	I m going to write this as a review for both D...
3	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...
4	Phenomenal stuff. I'll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...
...	...	...	...
996	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...
997	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...
998	The Bad1. The dialogue is clunky and robotic w...	The Bad1 The dialogue is clunky and robotic w...	The Bad The dialogue is clunky and robotic wi...
999	I just watched this in IMAX and it was one of ...	I just watched this in IMAX and it was one of ...	I just watched this in IMAX and it was one of ...
1000	This movie is a visual masterpiece that will b...	This movie is a visual masterpiece that will b...	This movie is a visual masterpiece that will b...

1001 rows × 3 columns

```
#3)Remove stopwords by using the stopwords List
def remove_stopwords(text):
    stop_words = set(stopwords.words('english'))
    words = nltk.word_tokenize(text)
    filter_words = [word for word in words if word.lower() not in stop_words]
    return ' '.join(filter_words)
df['clean_text_remove_stopwords'] = df['clean_text_remove_numbers'].apply(remove_stopwords)
print("\nData Frame after removing stopwords without lowercase:")
df
```

```
Data Frame after removing stopwords without lowercase:
```

	text	clean_text	clean_text_remove_numbers	clean_text_remove_stopwords
0	reviews	reviews	reviews	reviews
1	This is what Hollywood needs. A great story wi...	This is what Hollywood needs A great story wi...	This is what Hollywood needs A great story wi...	Hollywood needs great story great director pro...
2	I'm going to write this as a review for both D...	I m going to write this as a review for both D...	I m going to write this as a review for both D...	going write review Dune movies include thought...
3	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	pleasure watch film early screening completely...
4	Phenomenal stuff. I'll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff probably calm tomorrow right ...
...	...	...	...	...
996	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw early screening film Tilton Square Theatre...
997	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	biggest problem movie one crept Hollywood last...
998	The Bad1. The dialogue is clunky and robotic w...	The Bad1 The dialogue is clunky and robotic w...	The Bad The dialogue is clunky and robotic wi...	Bad dialogue clunky robotic much name use shou...
999	I just watched this in IMAX and it was one of ...	I just watched this in IMAX and it was one of ...	I just watched this in IMAX and it was one of ...	watched IMAX one greatest movie experiences ev...
1000	This movie is a visual masterpiece that will b...	This movie is a visual masterpiece that will b...	This movie is a visual masterpiece that will b...	movie visual masterpiece studied decades color...

```
# 4)Lowercase all texts
df['clean_text_lowercase'] = df['clean_text_remove_stopwords'].apply(lambda x: x.lower())
print("\nData Frame after converting texts to lowercase:")
df
```

```
Data Frame after converting texts to lowercase:
```

	text	clean_text	clean_text_remove_numbers	clean_text_remove_stopwords	clean_text_lowercase
0	reviews	reviews	reviews	reviews	reviews
1	This is what Hollywood needs. A great story wi...	This is what Hollywood needs A great story wi...	This is what Hollywood needs A great story wi...	Hollywood needs great story great director pro...	hollywood needs great story great director pro...
2	I'm going to write this as a review for both D...	I m going to write this as a review for both D...	I m going to write this as a review for both D...	going write review Dune movies include thought...	going write review dune movies include thought...
3	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	pleasure watch film early screening completely...	pleasure watch film early screening completely...
4	Phenomenal stuff. I'll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff probably calm tomorrow right ...	phenomenal stuff probably calm tomorrow right ...
...	...	...	...	...	...
996	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw early screening film Tilton Square Theatre...	saw early screening film tilton square theatre...
997	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	biggest problem movie one crept Hollywood last...	biggest problem movie one crept hollywood last...
998	The Bad1. The dialogue is clunky and robotic w...	The Bad1 The dialogue is clunky and robotic w...	The Bad The dialogue is clunky and robotic wi...	Bad dialogue clunky robotic much name use shou...	bad dialogue clunky robotic much name use shou...

```
#5)Stemming
stem = PorterStemmer()
def apply_stemming(text):
    words = nltk.word_tokenize(text)
    stemmed_words = [stem.stem(word) for word in words]
    return ' '.join(stemmed_words)
```

```
df['clean_text_stemmed'] = df['clean_text_lowercase'].apply(apply_stemming)
print("\nData Frame after applying stemming:")
df
```

Data Frame after applying stemming:

	text	clean_text	clean_text_remove_numbers	clean_text_remove_stopwords	clean_text_lowercase	clean_text_stemmed
0	reviews	reviews	reviews	reviews	reviews	review
1	This is what Hollywood needs. A great story wi...	This is what Hollywood needs A great story wi...	This is what Hollywood needs A great story wi...	Hollywood needs great story great director pro...	hollywood needs great story great director pro...	hollywood need great stori great director prod...
2	I'm going to write this as a review for both D...	I m going to write this as a review for both D...	I m going to write this as a review for both D...	going write review Dune movies include thought...	going write review dune movies include thought...	go write review dune movi includ thought dune ...
3	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	pleasure watch film early screening completely...	pleasure watch film early screening completely...	pleasur watch film earli screen complet blown ...
4	Phenomenal stuff. I'll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff probably calm tomorrow right ...	phenomenal stuff probably calm tomorrow right ...	phenomen stuff probabl calm tomorrow right hea...
...	...	...	...	...	...	...
996	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw early screening film Tilton Square Theatre...	saw early screening film tilton square theatre...	saw earli screen film tilton squar theatr new ...
997	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	biggest problem movie one crept Hollywood last...	biggest problem movie one crept hollywood last...	biggest problem movi one crept hollywood last ...
998	The Bad1. The dialogue is clunky and robotic w...	The Bad1 The dialogue is clunky and robotic w...	The Bad The dialogue is clunky and robotic wi...	Bad dialogue clunky robotic much name use shou...	bad dialogue clunky robotic much name use shou...	bad dialogu clunki robot much name use shoutou...
	I just watched	I just watched			watched imax one	watch imax one

```
#6)Lemmatization
lemmatizer = WordNetLemmatizer()
def apply_lemmatization(text):
    words = nltk.word_tokenize(text)
    lemmatized_words = [lemmatizer.lemmatize(word) for word in words]
    return ' '.join(lemmatized_words)
df['clean_text_lemmatized'] = df['clean_text_stemmed'].apply(apply_lemmatization)
print("\nData Frame after applying lemmatization:")
df
```

```
Data Frame after applying lemmatization:
```

	text	clean_text	clean_text_remove_numbers	clean_text_remove_stopwords	clean_text_lowercase	clean_text_stemmed	clean_te
0	reviews	reviews	reviews	reviews	reviews	review	
1	This is what Hollywood needs. A great story wi...	This is what Hollywood needs A great story wi...	This is what Hollywood needs A great story wi...	Hollywood needs great story great director pro...	hollywood needs great story great director pro...	hollywood need great stori great director prod...	holly stori gre
2	I'm going to write this as a review for both D...	I m going to write this as a review for both D...	I m going to write this as a review for both D...	going write review Dune movies include thought...	going write review dune movies include thought...	go write review dune movi includ thought dune ...	go v movi inc
3	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	Had the pleasure to watch this film in an earl...	pleasure watch film early screening completely...	pleasure watch film early screening completely...	pleasur watch film earli screen complet blown ...	pleasu screen
4	Phenomenal stuff. I'll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff I ll probably calm down tomo...	Phenomenal stuff probably calm tomorrow right ...	phenomenal stuff probably calm tomorrow right ...	phenomen stuff probabl calm tomorrow right hea...	phenor cal
...	...	...	...	...	...	...	
996	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw an early screening of this film at the Til...	Saw early screening film Tilton Square Theatre...	saw early screening film tilton square theatre...	saw earli screen film tilton squar theatr new ...	saw earli sq
997	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	The biggest problem with this movie is one tha...	biggest problem movie one crept Hollywood last...	biggest problem movie one crept hollywood last...	biggest problem movi one crept hollywood last ...	biggest p crept
998	The Bad1. The dialogue is clunky and robotic w...	The Bad1 The dialogue is clunky and robotic w...	The Bad The dialogue is clunky and robotic wi...	Bad dialogue clunky robotic much name use shou...	bad dialogue clunky robotic much name use shou...	bad dialogu clunki robot much name use shoutou...	bad dia
999	I just watched this in IMAX and it was one of	I just watched this in IMAX and it was one of	I just watched this in IMAX and it was one of ...	watched IMAX one greatest movie experiences ev...	watched imax one greatest movie experiences ev...	watch imax one greatest movi experi ever every...	watch ir movi ex

```
# Save the cleaned data to a new CSV file
df.to_csv('cleaned_data_.csv', index=False)
print("\nCleaned data saved ")
```

```
Cleaned data saved
```

Question 3 (30 points)

Write a python program to **conduct syntax and structure analysis of the clean text** you just saved above. The syntax and structure analysis includes:

(1) **Parts of Speech (POS) Tagging:** Tag Parts of Speech of each word in the text, and calculate the total number of N(oun), V(erb), Adj(ective), Adv(erb), respectively.

(2) **Constituency Parsing and Dependency Parsing:** print out the constituency parsing trees and dependency parsing trees of all the sentences. Using one sentence as an example to explain your understanding about the constituency parsing tree and dependency parsing tree.

(3) **Named Entity Recognition:** Extract all the entities such as person names, organizations, locations, product names, and date from the clean texts, calculate the count of each entity.

```
# Your code here
import pandas as pd
```

```
import nltk
from collections import Counter
# Download NLTK resources
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
True
```

```
df = pd.read_csv('cleaned_data_.csv')

#1)Parts of Speech (POS) Tagging
def pos_tagging(text):
    tokens = nltk.word_tokenize(text)
    pos_tags = nltk.pos_tag(tokens)
    return pos_tags
for index, row in df.iterrows():         #Iterate through each row and print the POS tagging on a new line
    pos_tags = pos_tagging(row['clean_text'])
    print(f"POS tagging for row {index + 1}:\n{pos_tags}\n")
    noun_count = verb_count = adj_count = adv_count = 0
    for _, pos in pos_tags:
        if pos.startswith('N'):
            noun_count += 1
        elif pos.startswith('V'):
            verb_count += 1
        elif pos.startswith('JJ'):
            adj_count += 1
        elif pos.startswith('RB'):
            adv_count += 1
    print(f"Total Nouns: {noun_count}")
    print(f"Total Verbs: {verb_count}")
    print(f"Total Adjectives: {adj_count}")
    print(f"Total Adverbs: {adv_count}")
```

```
('the', 'DT'), ('first', 'JJ'), ('movie', 'NN'), ('with', 'IN'), ('excellent', 'JJ'), ('action', 'NN'), ('scenes', 'NNS'), ('Casting
```

```
'NNS'), ('for', 'IN'), ('a', 'DT'), ('movie', 'NN'), ('of', 'IN'), ('this', 'DT'), ('caliber', 'NN'), ('I', 'PRP'), ('m', 'VBP'), ('
```

```
!pip install benepar
!pip install tensorflow
!pip install tensorflow==2.8.0
```

```
import benepar
import spacy.cli
benepar.download('benepar_en3')
spacy.cli.download("en_core_web_sm")

import sys
import spacy
from spacy import displacy
parser = benepar.Parser("benepar_en3")
nlp = spacy.load('en_core_web_sm')
options = {'compact': True, 'font': 'Arial black', 'distance': 100}
for sentence in df['clean_text']:
    try:
        tree = parser.parse(sentence)
        print(tree)
    except:
        print("No Parse Tree")
        continue
for sentence in df['clean_text']:     #Printing parse trees using spacy module
    doc = nlp(sentence)
    displacy.render(doc, style='dep', options=options, jupyter=True)
```

```
[nltk_data] Downloading package benepar_en3 to /root/nltk_data...
[nltk_data]   Package benepar_en3 is already up-to-date!
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
⚠ Restart to reload dependencies
If you are in a Jupyter or Colab notebook, you may need to restart Python in
order to load all the package's dependencies. You can do this by selecting the
'Restart kernel' or 'Restart runtime' option.
/usr/local/lib/python3.10/dist-packages/benepar/parse_chart.py:169: FutureWarning: You are using `torch.load` with `weights_only=False
  state_dict = torch.load(
/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:1601: FutureWarning: `clean_up_tokenization_spaces` wa
  warnings.warn(
You're using a T5TokenizerFast tokenizer. Please note that with a fast tokenizer, using the `__call__` method is faster than using a m
/usr/local/lib/python3.10/dist-packages/torch/distributions/distribution.py:55: UserWarning: <class 'torch_struct.distributions.TreeCR
  warnings.warn(
(TOP (NP (NNS reviews)))
Token indices sequence length is longer than the specified maximum sequence length for this model (2137 > 512). Running this sequence
Streaming output truncated to the last 5000 lines.
            (SBAR
              (S
                (SBAR
                  (S
                    (NP (PRP they))
                    (VP
                      (VBD were)
                      (VP
                        (VBG working)
                        (PP
                          (IN on)
                          (NP
                            (NP
                              (NP (NN something))
                              (ADJP (JJ special)))
                            (SBAR
                              (S
                                (NP
                                  (NP
                                    (NNP Denis)
                                    (NNP Villeneuve)
                                    (POS s))
                                  (NN love)
                                  (PP (IN for) (NP (NNP Dune))))
                                (VP
                                  (VBZ shines)
                                  (PP
                                    (IN through)
                                    (NP
                                      (NP
                                        (NP
                                          (DT every)
                                          (JJ stunning)
                                          (NN frame))
                                        (NP
                                          (NP (NN Part) (CD 1))
                                          (PP
                                            (IN of)
                                            (NP
                                              (PRP$ his)
                                              (NN adaptation)))))
                                      (VP
                                        (VBD was)
                                        (ADJP (JJ wonderful))))))))))))))
                (CC but)))
            (ADVP (RB here))
            (NP (PRP he)))
          (VP
            (VBZ improves)
            (PP (IN on) (NP (PRP it)))
            (PP
              (IN in)
              (NP
                (NP (DT every) (NN way))
                (NP (PRP It))
                (ADVP (RB just))
                (VP
                  (VBZ doesn)
                  (RB t)
                  (VP
                    (VB get)
                    (ADJP (ADJP (JJR better)) (IN than) (DT this)))))))))))
No Parse Tree
(TOP
  (S
    (S
      (NP (PRP I))
```

```
    (NP (PRP I))
    (ADVP (RB just))
    (VP
      (VBD got)
      (PP
        (IN out)
        (PP
          (IN of)
          (NP (DT an) (JJ early) (NN access) (NN showing))))))
  (CC and)
  (S
    (NP (PRP it))
    (VP
      (VBD was)
      (ADJP (RB absolutely) (JJ incredible))
      (S
        (VP
          (VB See)
          (PP (IN for) (NP (PRP yourself)))
          (PP (IN in) (NP (NNP IMAX)))
          (NP (DT The) (NNS characters))
          (VP
            (VBG acting)
            (NP (NN screenplay) (NN world))
            (VBG building)
            (NP
              (NN storytelling)
              (NN score)
              (NNS actions)
              (NNS sequences))
            (NN cinematography))
          (CC and)
          (NP
            (NP
              (NP
                (NP
                  (NP (NN everything))
                  (PP (IN in) (PP (IN between))))
                (VBP make))
              (PP
                (IN for)
                (NP (DT a) (JJ cinematic) (NN masterpiece))))
            (NP (NNP Denis) (NNP Villeneuve)))
          (VP
            (VBZ provides)
            (NP
              (NP (DT a) (NN masterclass))
              (PP
                (PP
                  (PP (IN of) (NP (NN filmmaking)))
                  (NP (DT The) (NN casting) (NN continuation)))
                (VP
                  (VBD was)
                  (ADJP (JJ perfect))
                  (ADVP (DT all) (DT the) (NN way) (RB through))
                  (PP
                    (IN with)
                    (NP
                      (NP
                        (NP (JJ great) (JJ new) (NN add) (NNS ons))
                        (NML
                          (SBAR
                            (S
                              (NP (NNP Timothee) (NNP Chalamet))
                              (VP
                                (VBZ is)
                                (ADJP
                                  (JJ believable)
                                  (JJ raw)
                                  (CC and)
                                  (JJ real))
                                (PP
                                  (IN as)
                                  (S
                                    (NP
                                      (NP
                                        (NNP Paul)
                                        (NNP Atriedes))
                                      (NP (PRP He)))
                                    (VP
                                      (VBD was)
                                      (ADJP (JJ flawless))
                                      (SBAR
                                        (IN as)
                                        (S
```