BSD2513 ARTIFICIAL INTELLIGENCE

LAB REPORT 3

NAME : TEAN JIN HE

MATRIC ID : SD21063

SECTION : 02G

Questions 1 : General Knowledge

```
Syntactic analysis deals with the structure and arrangement of words or phrases in a sentence. It primarily focuses on the grammatical rules and relationships between words to determine the syntactic structure of a sentence. This analysis involves tasks such as part-of-speech (POS) tagging, parsing, and dependency parsing. The goal is to understand the grammatical structure of a sentence without considering the precise meaning of the words. Syntactic analysis helps in identifying the roles of words in a sentence, such as subject, object, verb and others.

Semantic analysis focuses on understanding the meaning of words, phrases, and sentences. It aims to capture the deeper meaning and interpretation of the text. This analysis involves tasks such as word sense disambiguation, named entity recognition, semantic role labeling, sentiment analysis, and question answering. Semantic analysis takes into account the context, domain knowledge, and world knowledge to derive the intended meaning of a sentence.

Reference:
Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT press.
```

Question 2:

Python: Search Algorithms

In [4]:
```
import nltk
from nltk.tokenize import sent_tokenize, word_tokenize
```

In [2]:
```
pip install nltk
```

```
Requirement already satisfied: nltk in d:\anaconda\lib\site-packages (3.7)
Requirement already satisfied: click in d:\anaconda\lib\site-packages (from nltk) (8.0.4)
Requirement already satisfied: joblib in d:\anaconda\lib\site-packages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in d:\anaconda\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in d:\anaconda\lib\site-packages (from nltk) (4.64.1)
Requirement already satisfied: colorama in d:\anaconda\lib\site-packages (from click->nltk) (0.4.5)
Note: you may need to restart the kernel to use updated packages.
```

In [5]:
```
# Define input text
input_text = '''Sherlock Holmes (/ˈʃɜːrlɒk ˈhoʊmz/) is a fictional detective created by
British author Sir Arthur Conan Doyle. Referring to himself as a
"consulting detective" in the stories, Holmes is known for his proficiency
with observation, deduction, forensic science and logical reasoning that
borders on the fantastic, which he employs when investigating cases for a
wide variety of clients, including Scotland Yard.

First appearing in print in 1887's A Study in Scarlet, the character's
popularity became widespread with the first series of short stories in
The Strand Magazine, beginning with "A Scandal in Bohemia" in 1891;
additional tales appeared from then until 1927, eventually totalling four
novels and 56 short stories. All but one[a] are set in the Victorian or
Edwardian eras, between about 1880 and 1914. Most are narrated by the
character of Holmes's friend and biographer Dr. John H. Watson, who usually
accompanies Holmes during his investigations and often shares quarters
with him at the address of 221B Baker Street, London, where many of the
stories begin'''
```

In [6]:

```
# Word tokenizer
print("Word tokenizer:")
print(word_tokenize(input_text))
```

```
Word tokenizer:
['Sherlock', 'Holmes', '(', '/ˈʃɜːrlɒk', 'ˈhoʊmz/', ')', 'is', 'a', 'fictional', 'detective', 'created', 'by', 'Brit
ish', 'author', 'Sir', 'Arthur', 'Conan', 'Doyle', '.', 'Referring', 'to', 'himself', 'as', 'a', "'", 'consulting',
'detective', "'", 'in', 'the', 'stories', ',', 'Holmes', 'is', 'known', 'for', 'his', 'proficiency', 'with', 'obser
vation', ',', 'deduction', ',', 'forensic', 'science', 'and', 'logical', 'reasoning', 'that', 'borders', 'on', 'th
e', 'fantastic', ',', 'which', 'he', 'employs', 'when', 'investigating', 'cases', 'for', 'a', 'wide', 'variety', 'o
f', 'clients', ',', 'including', 'Scotland', 'Yard', '.', 'First', 'appearing', 'in', 'print', 'in', '1887', "'s",
'A', 'Study', 'in', 'Scarlet', ',', 'the', 'character', "'s", 'popularity', 'became', 'widespread', 'with', 'the',
'first', 'series', 'of', 'short', 'stories', 'in', 'The', 'Strand', 'Magazine', ',', 'beginning', 'with', '``', 'A',
'Scandal', 'in', 'Bohemia', "'", 'in', '1891', ';', 'additional', 'tales', 'appeared', 'from', 'then', 'until', '19
27', ',', 'eventually', 'totalling', 'four', 'novels', 'and', '56', 'short', 'stories', '.', 'All', 'but', 'one',
'[', 'a', ']', 'are', 'set', 'in', 'the', 'Victorian', 'or', 'Edwardian', 'eras', ',', 'between', 'about', '1880',
'and', '1914', '.', 'Most', 'are', 'narrated', 'by', 'the', 'character', 'of', 'Holmes', "'s", 'friend', 'and', 'bio
grapher', 'Dr.', 'John', 'H.', 'Watson', ',', 'who', 'usually', 'accompanies', 'Holmes', 'during', 'his', 'investiga
tions', 'and', 'often', 'shares', 'quarters', 'with', 'him', 'at', 'the', 'address', 'of', '221B', 'Baker', 'Stree
t', ',', 'London', ',', 'where', 'many', 'of', 'the', 'stories', 'begin']
```

In [1]:

```python
# Alternative method
import nltk

nltk.download('punkt')

text = '''Sherlock Holmes (/ˈʃɜːrlɒk ˈhoʊmz/) is a fictional detective created by
British author Sir Arthur Conan Doyle. Referring to himself as a
"consulting detective" in the stories, Holmes is known for his proficiency
with observation, deduction, forensic science and logical reasoning that
borders on the fantastic, which he employs when investigating cases for a
wide variety of clients, including Scotland Yard.
First appearing in print in 1887's A Study in Scarlet, the character's
popularity became widespread with the first series of short stories in
The Strand Magazine, beginning with "A Scandal in Bohemia" in 1891;
additional tales appeared from then until 1927, eventually totaling four
novels and 56 short stories. All but one[a] are set in the Victorian or
Edwardian eras, between about 1880 and 1914. Most are narrated by the
character of Holmes's friend and biographer Dr. John H. Watson, who usually
accompanies Holmes during his investigations and often shares quarters
with him at the address of 221B Baker Street, London, where many of the
stories begin'''

# Tokenize into sentences
sentences = nltk.sent_tokenize(text)

# Tokenize into words
words = nltk.word_tokenize(text)

# Print the results
print("Sentences:")
for sentence in sentences:
    print(sentence)

print("\nWords:")
for word in words:
    print(word)
```

```
Sentences:
Sherlock Holmes (/ˈʃɜːrlɒk ˈhoʊmz/) is a fictional detective created by
British author Sir Arthur Conan Doyle.
Referring to himself as a
"consulting detective" in the stories, Holmes is known for his proficiency
with observation, deduction, forensic science and logical reasoning that
borders on the fantastic, which he employs when investigating cases for a
wide variety of clients, including Scotland Yard.
First appearing in print in 1887's A Study in Scarlet, the character's
popularity became widespread with the first series of short stories in
The Strand Magazine, beginning with "A Scandal in Bohemia" in 1891;
additional tales appeared from then until 1927, eventually totaling four
novels and 56 short stories.
All but one[a] are set in the Victorian or
Edwardian eras, between about 1880 and 1914.
Most are narrated by the
character of Holmes's friend and biographer Dr. John H. Watson, who usually
accompanies Holmes during his investigations and often shares quarters
with him at the address of 221B Baker Street, London, where many of the
```

```
1 From the code, we can see that it imports the NLTK library and downloads the 'punkt' package, which is a pre-
  trained model for sentence and word tokenization.
2 Then, we will assigns the input text as a string variable named 'text' and we tokenize the text into words
  using nltk.word_tokenize(text). This function splits the text into individual words, including punctuation
  marks.
3 Next, we will using the nltk.sent_tokenize() function to split the text into a list of sentences, and assigns
  it to a variable called 'sentences' which using nltk.sent_tokenize(text) and this function splits the text into
  individual sentences.
4 Moreover, we use the nltk.word_tokenize() function to split the text into a list of words, and assigns it to a
  variable called 'words'.
5 Finally, we print the results by looping through the 'sentences' and 'words' lists and printing each element on
  a new line.
6 The code will output the words and phrases from the input text as independent tokens. The words will be printed
  within quote marks one at a time. The sentences will be displayed line by line until the full stop within the
  quote mark.
```