Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2024-25

Class:	BE-AIDS	**Semester:**	VII
Course Code:		**Course Name:**	Natural language processing

Name of Student:	Mokshad Ketan Sankhe
Roll No. :	67
Assignment No.:	2
Title of Assignment:	
Date of Submission:	
Date of Correction:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Demonstrated knowledge	5	
Legibility	3	
Completeness and timely submission	2	
Total	10	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Demonstrated Knowledge	5	3-4	1-2
Legibility	3	2	1
Completeness and Timely submission	2	1	0

Checked by

Name of Faculty :

Signature :

Date :

NLP Assignment-2

Q1) Given statement:

1) <s> clark Kent conceals his identity. </s>
2) <s> Bruce wayne knows his cards </s>
3) <s> Diana Prince is saviour of week </s>

Soln:-

1) P(his/Knows):-

Occurance of "knows his" = 1

Occurance of "Knows" = 1

$\therefore P(his/Knows) = \frac{count(Knows\ his)}{count(Knows)} = \frac{1}{1} = 1$

2) P(conceals/his):-

Occurance of 'his conceals' = 0

Occurance of 'his' = 1

$\therefore P(conceals/his) = \frac{0}{1} = 0.00$

3) P(Diana/Prince):-

Occurance of "Prince Diana" : 0

Occurance of "Prince" = 1

$\therefore P(Diana/Prince) = \frac{0}{1} = 0.00$

4) P(Prince/Diana):-

Occurance of "Diana Prince" = 1

Occurance of "Diana" = 1

$\therefore P(Prince/Diana) = \frac{1}{1} = 1.00$

Q2) Deterministic Finite Automata (DFA) for "the" in strings

→

- Given: a string containing "the" anywhere within a string of a-z. eg:- "there" but not "those" we need to construct a DFA that accepts strings containing "the".
- DFA construction:
  - States: q_0 (initial), q_1, q_2, q_3 (accepting)
  - Alphabets: [a, b, c .... z]
  - Transitions:
    - From q_0:
      - on 't' → q_1
      - on any other character → stay at q_0.
    - From q_1:
      - on 'h' → q_2
      - on 't' → stay at q_1
      - on any other character → go back to q_0
    - From q_2:
      - on 'e' → q_3 (accepting state)
      - on 't' → q_1
      - on any other character → go back to q_0
    - From q_3:
      - on any character → stay at q_3

't' 'h' 'e'

(q_{10}) → (q_{11}) → (q_{12}) → (q_{13})

any other character | 't' character | any other characters | 't' character

- The DFA start at q_0, and it transitions to q_3 upon encountering the sequence "the" q_3 is the accepting state indicating the presence of "the" in the string.
- With this DFA, any string containing "the" will be accepted and those without it like "those" will not be accepted.