Natural Language Processing
Assignment 5
Type of Question: MCQ

Number of Questions: 8 Total Marks: (6×1)+(2×2)=10

==

Question 1.
Which of the following are true?

A) Given a CFG and its corresponding CNF, they produce different languages.
B) It requires ‘2n-1’ productions or steps in CNF to generate a string w of length ‘n’.
C) For a given grammar, there can be one CNF.
D) All of the above

Answer: B
Solution: Given a CFG and its corresponding CNF, they produce the same language.
Let n be the length of a string. We start with the (non-terminal) symbol S which has length n=1. Using (n−1) rules of form NT→NT NT (where NT represents a non−terminal) we can construct a string containing ‘n’ non-terminal symbols. Then on each NT symbol of said string of length ‘n’ we apply a rule of form NT→T. i.e. we apply n rules. In total we will have applied
(n−1) + n = 2n−1 rules.

==

Question 2.
Consider the CFG given below:
$S \rightarrow xSy|V$
$V \rightarrow Vz|\epsilon$
How many non-terminals should be added to convert the CFG into CNF?

A) 2
B) 4
C) 5
D) 3

Answer: C
Solution: The final CNF is:
$S^{'} \rightarrow AE|AB|VC|z$
$S \rightarrow AE|AB|VC|z$

$E \rightarrow SB$
$V \rightarrow VC|z$
$A \rightarrow x$
$B \rightarrow y$
$C \rightarrow z$

The correct answer is **C) 5**.

To convert the given Context-Free Grammar (CFG) into Chomsky Normal Form (CNF), you need to add 5 non-terminals. In CNF, production rules must either be of the form:
1. A --> BC (two non-terminals), or
2. A --> a (a single terminal).

The conversion requires breaking down rules that involve sequences of terminals and/or multiple non-terminals. For example, new non-terminals are introduced to replace terminals like `x`, `y`, and `z`, as well as for intermediary non-terminals to create compliant production rules.

==

Question 3:
the CFG to CNF converted form?
In the above Q. 2) How many different numbers of Null productions in

A) 0
B) 1
C) 2
D) 3

The answer to this question is **A) 0**.

A **null production** in a CFG is a production of the form A -> eta, where A is a non-terminal and represents the empty string.

In the given CFG:
- V -> Vz | eta contains the null production V -> eta.
- However, in the process of converting the grammar to CNF, all null productions are eliminated. This is because CNF does not allow null productions (except for the start symbol if needed, but even this can be avoided).

When converting to CNF, eta - productions are removed by adjusting other rules to account for the possibility of the null production. Hence, the resulting CNF for the given CFG does not include any null productions.

Answer: A (0)

Please check the Q. 2. Solution

==

Question 4:
In the above Q. 2) How many different numbers of production rules/steps in the CFG to CNF converted form??

A) 2
B) 4
C) 7
D) 10

Count no. of steps in Q2

Answer: C (7)

Please check the Q. 2. Solution

==

For **Question 5 to 7** consider the following PCFG fragment:

$S \rightarrow$ NN VP	0.50	$S \rightarrow$ VP NN	0.50
$NP \rightarrow$ NN PB	0.40	$PB \rightarrow$ PP NN	0.30
$VP \rightarrow$ VB NN	0.30	$VP \rightarrow$ VB NP	0.20
$VP \rightarrow$ NN VB	0.25	$VP \rightarrow$ NN PB	0.15
$PP \rightarrow$ with	0.10	$PP \rightarrow$ without	0.10
$VB \rightarrow$ play	0.30	$VB \rightarrow$ *enjoy/like*	0.20

$VB \rightarrow$ watch/enjoy	0.25	$NN \rightarrow$ children/students	0.15
$NN \rightarrow$ cricket/football	0.15	$NN \rightarrow$ friends	0.20
$NN \rightarrow$ football/cricket	0.10	$NN \rightarrow$ music/painting	0.12

For a sentence S = w1w2w3w4, assume that the cells in the table are indexed as follows:

	1	2	3	4	
w_1	11	12	13	14	1
	w_2	22	23	24	2
		w_3	33	34	3
			w_4	44	4

Question 5:
Using CKY algorithm, find the probability score for the most probable tree for the sentence S_1 = “**students play football with friends**”.

A) 6.06×10^{-4}
B) 1.62×10^{-6}
C) 2.73×10^{-3}
D) 4.33×10^{-6}

Answer: B
Solution: Calculate the probability using the Bottom-Up method as explained in the lecture. 2

==\

Question 6:
Using CKY algorithm, find the number of parse trees for the sentence S_2 = **students like painting** and the probability score for “at least one of the probable trees”.

A) 1, 4.95×10^{-3}
B) 3, 0.36×10^{-3}
C) 2, 0.99×10^{-3}
D) 2, 0.54×10^{-3}

Answer: C
Solution:
There are two parse trees.
$S \rightarrow NN_{11}\, VP_{23} = 0.5 \times 0.15 \times (0.3 \times 0.2 \times 0.12) = 0.54 \times 10^{-3}$
$S \rightarrow VP_{12}\, NN_{33} = 0.5 \times (0.25 \times 0.15 \times 0.2) \times 0.12 = 0.45 \times 10^{-3}$

Total is sum of both of the above = $0.54 \times 10^{-3} + 0.45 \times 10^{-3} = 0.99 \times 10^{-3}$

We get the above probabilities with the CKY algorithm.

==

Question 7:

Consider the expression below:

P("students enjoy cricket like painting", $N_{34|}G$) = P_j P("students enjoy cricket like painting" |N^j_{34}, G)

What does the L.H.S. represent?

A) Probability of the sentence "students enjoy cricket like painting", given a grammar G.
B) Probability of the sentence "students enjoy cricket like painting", given a grammar G and that there is some consistent spanning of the segment "cricket like", i.e. from word 3 to 4.
C) Probability of the sentence "students enjoy cricket like painting", given a grammar G and some rule which derives the segment "cricket like".
D) None of the above

Answer: B
Solution: Refer to Inside-Outside Probabilities.

==

Question 8:
Which of the following grammars are valid CNF?

1. a) A → B 2. A → BCD 3. A → BC
   b) B → CD B → b B → ϵ
   c) C → c C → c C → c
   d) A → BC   A → a

2. a
3. b
4. c
5. D
6. None

Answer: d
Solution: Valid CNF form is as follows:
A → BC
A → a

A non-terminal generating a terminal or A non-terminal generating two non-terminals

==