NATIONAL UNIVERSITY OF SINGAPORE
SCHOOL OF COMPUTING

CS4248 – Natural Language Processing

Semester 1 AY2022/2023

November 2022 **Time Allowed: 1.5 Hours**

INSTRUCTIONS TO CANDIDATES

1. This assessment paper contains **EIGHT (8)** questions and comprises **TEN (10)** printed pages, including this page.
2. Answer **ALL** questions within the space in this booklet.
3. This is a **CLOSED** book assessment, but one double-sided A4 sized sheet is allowed for notes.
4. A non-programmable calculator is permitted.
5. Please write your Student Number below. Do not write your name.

This portion is for lecturer's use only

Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
Max	20	5	10	20	5	20	15	5	100
Marks									

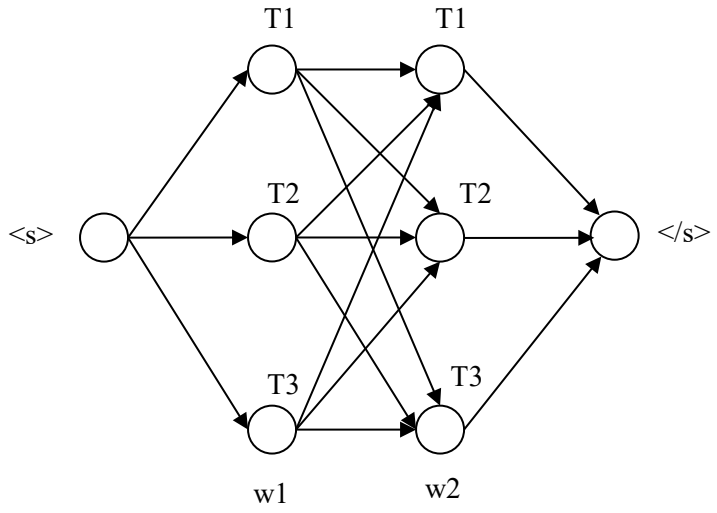
1. (20 marks) Give a trace of the minimum edit distance algorithm (a dynamic programming algorithm) to compute the minimum cost of transforming the string "event" to "vain", by filling out every cell entry in the following table, where each cell entry denotes the minimum cost of transforming the associated substrings. Assume that the cost of inserting a character is 1, the cost of deleting a character is 1, and the cost of substituting a character by a different character is 2. (You do not need to show the optimal path.)

t	5				
n	4				
e	3				
v	2				
e	1				
	0	1	2	3	4
		v	a	i	n

2. (5 marks) Consider the task of translating from German to English, which takes a German sentence as input, and outputs a translated English sentence. When this task is cast as a noisy channel model, what is the input to the noisy channel?

3. (10 marks) An automatic speech recognition (ASR) system recognizes the following words:
deposit, which occurs 1 in 4 times
loan, which occurs 1 in 4 times
operator, which occurs 1 in 4 times
100 different family names, each name occurs 1 in 400 times
Compute the perplexity of the language recognized by this ASR system. Show clearly the steps of your computation.

4. (20 marks) Consider the following HMM:



Suppose this HMM has the following set of parameters:

P(T1|<s>) = 4/5	P(T1|T1) = 1/4	P(T1|T2) = 1/3	P(T1|T3) = 1/8
P(T2|<s>) = 0	P(T2|T1) = 1/8	P(T2|T2) = 1/3	P(T2|T3) = 3/8
P(T3|<s>) = 1/5	P(T3|T1) = 1/2	P(T3|T2) = 1/6	P(T3|T3) = 1/4
	P(</s>|T1) = 1/8	P(</s>|T2) = 1/6	P(</s>|T3) = 1/4

P(w1|T1) = 1/10	P(w2|T1) = 1/10
P(w1|T2) = 1/20	P(w2|T2) = 1/2
P(w1|T3) = 1/5	P(w2|T3) = 1/6

T1, T2, T3 are part-of-speech tags.
Consider the input sentence “w1 w2”, where w1 and w2 are words.
Compute P(<s>, w1, w2, </s>). Show clearly the steps of your computation.

(Additional space for answering question 4, if needed)

5. (5 marks) A perceptron F receives inputs $x_1, \ldots, x_n$ and outputs the following:

$$F(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } w_0 + w_1 x_1 + \cdots + w_n x_n > 0 \\ 0 & \text{otherwise} \end{cases}$$

Give 4 weights w_0, w_1, w_2, w_3 such that F implements the Boolean function

$$(x_1 \vee x_2) \wedge (x_1 \vee x_3), \text{ where } x_i = 0 \text{ or } 1,\ 1 \leq i \leq 3.$$

Clearly justify your answers.

6. Let $P(n)$ be the number of different parses due to prepositional phrase (PP) attachment ambiguity of $(n-1)$ PPs following a noun phrase (NP). For example, given one NP followed by 2 PPs (such as "the meal on flight 123 from Singapore"), there are $P(3) = 2$ different parses, corresponding to the following bracketing:

A [B C]
[A B] C

where A stands for the NP, and B and C stand for the PPs.

(a) (5 marks) Consider $n = 4$, that is, the case of one NP followed by 3 PPs (for example, "the meal on flight 123 from Singapore to Sydney"). What is the number of different parses in this case? List all the parses in the above bracketing format.

(b) (15 marks) Give a recurrence relation for $P(n)$ (i.e., express $P(n)$ in terms of $P(k)$ for $k < n$). Clearly justify your answer.

7. (15 marks) Consider the following English sentence: “There are at most two cats.” Define the following predicate symbols:

$C(x)$: x is a cat
$I(x, y)$: x is the same as y

For each of the following first-order logic formulas, if the formula is a correct meaning representation for the above English sentence, put “Y” in the square bracket before the formula. If the formula is **not** a correct meaning representation for the English sentence, put “N” in the square bracket. 5 marks will be awarded for each correct answer, and 5 marks will be deducted for each wrong answer. The minimum score for this question is 0.

[ ] $\forall x \exists y \exists z((\neg C(x) \vee C(y) \vee I(y, z)) \rightarrow C(z))$

[ ] $\exists x \exists y(C(x) \wedge C(y) \wedge \forall z(C(z) \rightarrow (I(z, x) \vee I(z, y))))$

[ ] $\exists x \exists y((C(x) \wedge C(y)) \rightarrow \forall z(C(z) \rightarrow (I(z, x) \vee I(z, y))))$

8. (5 marks) Give the meaning representation of the noun phrase “a dog” using a lambda expression. The predicate $D(x)$ denotes x is a dog.

(Additional space for answering any questions in this paper, if needed)

(Additional space for answering any questions in this paper, if needed)

END OF PAPER