UNIVERSITY OF COLOMBO, SRI LANKA

UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

BACHELOR OF SCIENCE HONOURS IN COMPUTER SCIENCE
BACHELOR OF SCIENCE HONOURS IN SOFTWARE ENGINEERING

Fourth Year Examination – Semester I – 2021

SCS4209 – Natural Language Processing – (Part A)

TWO (2) HOURS (for both Part A and Part B)

To be completed by the candidate

Examination Index No: ....

Important Instructions to candidates:

1. The medium of instruction and question is **English**.

2. If a page or a part of this question paper is not printed, please inform the supervisor immediately.

3. Note that questions appear on both sides of the paper. If a page is not printed, please inform the supervisor immediately.

4. Write your index number on each and every page of the Question paper.

5. **This paper consists of two parts, Part A (Question No 1 and Question No 2) and Part B (Question No 3 and Question No 4) and submit separately.**

6. This Part (A) has **02** questions and **07** pages.

7. Answer **ALL** questions. All questions carry equal marks (25 marks).

8. Any electronic device capable of storing and retrieving text including electronic dictionaries and mobile phones are **not allowed**.

9. **Non-Programmable** calculators are **allowed**.

Part A

Question 1

(a) Briefly describe the difference between *Acoustic Phonetics* and *Articulatory Phonetics*.

[4 Marks]

| |
| |

(b) Write the articulatory description (*voicing, place of articulation and manner of articulation*) for the each of the following phones.

[6 Marks]

Phone	Place of Articulation	Manner of Articulation	Voiced / Voiceless
/l/			
/m/			
/c/			
/f/			

(c) Both *stemming* and *lemmatization* are word normalization techniques. Briefly explain why lemmatization is always a better option from a qualitative perspective compared to stemming. Justify your answer by providing examples in English. [5 Marks]

(d) Identify the stems, derivational morphemes and inflectional morphemes separately of all the following words. [10 Marks]

Word	Stem	Derivational Morphemes (if any)	Inflectional Morphemes (if any)
buyers			
unhappily			
beautiful			
smallest			

3

Question 2

(a) Define **Part of Speech Tagging**. Briefly explain why *Part of Speech Tagging* is essential for language processing. [4 Marks]

(b) Consider the following fragment of English Grammar.

S → NP VP

NP → D N | PropN

VP → V | V NP | V PP

PP → P NP

D → a | the | an

N → man | dog | cat | tree | boy | hat | forest

P → with | on | into | under | at | in

PropN → John | Mary | Bob

V → ran | walked | barked | chased | climbed | saw

(i) Write down two (2) structurally different and grammatical sentences and one (1) grammatical but senseless sentence generated by this grammar. [3 Marks]

4

```
┌─────────────────────────────────────────────────────────────┐
│                                                             │
│                                                             │
│                                                             │
│                                                             │
│                                                             │
└─────────────────────────────────────────────────────────────┘
```

(ii) What additional rule(s) / lexical items would you include to the above grammar to accommodate the following sentences?

[6 Marks]

```
┌─────────────────────────────────────────────────────────────┐
│  (a) A man in the forest climbed the tree.                  │
│                                                             │
│                                                             │
│                                                             │
│  (b) John saw a boy with a hat.                             │
│                                                             │
│                                                             │
│                                                             │
│  (c) Mary said the dog ate a cake.                          │
│                                                             │
│                                                             │
│                                                             │
│                                                             │
└─────────────────────────────────────────────────────────────┘
```

(c) What kind of ambiguity is there in the following sentences? Justify your answers by drawing parse trees or using any other method.

    (i)   I invited the person with the microphone

    (ii)   I saw him at the bank

[6 Marks]

(d) Consider the following small corpus with four (4) sentences. It defines the tags in order to represent the beginning and end of the sentence.

        <S> Buddhi is from Horana </S>
        <S> Buddhi is a research student </S>
        <S> Students are good and are from various towns </S>
        <S> Students from Horana do Computer Science </S>

Calculate the Bigram probability of the following test sentence using the above corpus.

Test Sentence: <S> Students are from Horana </S>   [6 Marks]

******