

TUNKU ABDUL RAHMAN UNIVERSITY OF MANAGEMENT AND TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

ACADEMIC YEAR 2023/2024

JANUARY EXAMINATION

COMPUTER SCIENCE BACS2003
ARTIFICIAL INTELLIGENCE

TUESDAY, 9 JANUARY 2024

TIME: 2.00 PM – 4.00 PM (2 HOURS)

BACHELOR OF SOFTWARE ENGINEERING (HONOURS)

Instructions to Candidates:

Answer **ALL** questions. All questions carry equal marks.

Question 1

- a) The Turing Test evaluates a machine's capacity to exhibit intelligent behaviour in natural language conversations that is indistinguishable from human behaviour. Despite its popularity, the Turing Test has been criticised for its ineffectiveness in assessing artificial intelligence. John Searle's Chinese Room argument is a well-known objection.
- (i) Explain how the Turing Test works. (3 marks)
 - (ii) Outline the criticism presented by the Chinese Room argument. (2 marks)
- b) A team of researchers has created an autonomous driving vehicle named Agent X. An upcoming experiment is designed to evaluate Agent X's performance as it navigates a map illustrated in Figure 1. This map consists of a total of 9 nodes, labelled by the letters A to I, and with the distances (in kilometers) between these nodes are illustrated along the arcs.

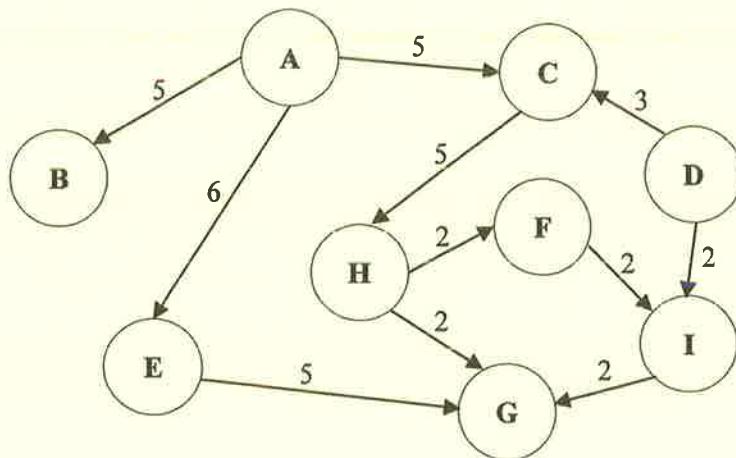


Figure 1: Map

- (i) Formulate the goal, optimal solution, abstraction, initial state, successor function, goal test, step cost and path cost in the problem above. (8 marks)
- (ii) Illustrate resulting search trees by using Breadth First Search (BFS) and Depth First Search (DFS).
(Note: The search process should be carried out alphabetically). (6 marks)
- (iii) For each search algorithm, identify its solution path returned. (2 marks)
- (iv) Compare the performance between the BFS and DFS in terms of completeness and optimality. (4 marks)

[Total: 25 marks]

Question | a)

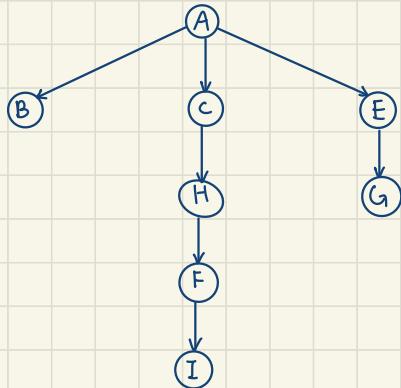
- (i) - A human judge (interrogator) will ask a question to both human and machine through a text-based interface during Turing Test.
 - Machine's goal is to produce a response which is indistinguishable from human response, while human's responses will act as a baseline for comparison with machine's response.
 - If the interrogator cannot consistently tell which is the machine, it indicates that the machine has passed the Turing Test.
-
- (ii) - Chinese Room is a thought experiment that challenges the notion that machine can possess genuine understanding and consciousness, despite potentially exhibiting intelligent behavior.
 - However, the criticism states that the machine does not have actual understanding in the context even though it can produce a responses that seem intelligent.

Question 1 b) (i)

- Goal : Reach node I
- Optimal solution : Reach node I using the shortest distance
- Abstraction : Time
- Initial state : Node A
- Successor function : Function that stores all the possible successor states from the current state
- Goal test : Test if the current state equal to the goal state (node I)
- Step cost : The cost between two nodes
- Path cost : Total cost from current state to goal state

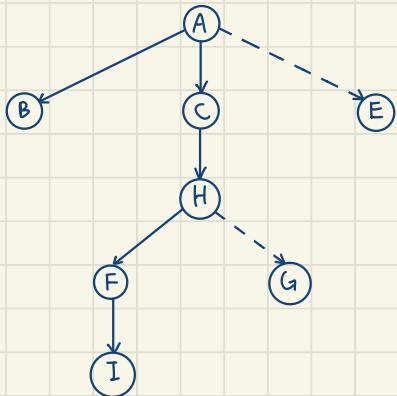
Question b) (ii)

Breadth First Search (BFS)



Open	Closed
A	
B	A
	B
	C
E, H	
H, G	C
F	E
	H
	G
I	F

Depth First Search (DFS)



Open	Closed
A	
B, C, E	A
	B
H, E	C
F, G, E	H
I	F

Question 1 b) (iii)

BFS : A - C - H - F - I

DFS : A - C - H - F - I

Question 1 b) (iv)

	BFS	DFS
Completeness	It is complete since it has successfully found out the solution path.	It is complete since it has successfully found out the solution path.
Optimality	It is less optimality compared to DFS because it will explore all the nodes level by level and store all nodes at current level. It may not suitable and take longer time when the goal state is at a deep level.	It is more optimal compared to BFS because it can explore the nodes as deep as possible and store current path only. It is suitable when the goal node is at a deep level.

BACS2003 ARTIFICIAL INTELLIGENCE**Question 2**

- a) Natural Language Processing (NLP) comprises two distinct components: Natural Language Understanding (NLU) and Natural Language Generation (NLG). Explain on the functions of each of these components. (4 marks)
- b) Consider a sentence $S = I \text{ shot an elephant in my pyjamas}$. Explain TWO (2) types of ambiguities found in the sentence S . (2 + 4 marks)
- c) Illustrate TWO (2) parse trees for the sentence S from Question 2 b) using the grammar rules outlined in Figure 2. (11 marks)

```

noun_phrase(NP)
verb_phrase(VP)
preposition_phrase(PP)
determiner(D)
noun(N)
verb(V)
S -> NP VP
NP -> Det N PP | Det N | 'I'
VP -> VP PP | V NP
PP -> P NP
N -> [elephant, pyjamas]
V -> [shot]
P -> [in]
Det -> [an, my]

```

Figure 2: Grammar rules

- d) Provide and explain TWO (2) analyses in NLP. (4 marks)

[Total: 25 marks]

Question 2 a)

- Natural Language Understanding (NLU) makes sense of human language input by understanding the intent, context and the meaning.
- Natural Language Generation (NLG) generates human-like language from structured data and text by converting machine readable information into human readable sentences.

Question 2 b)

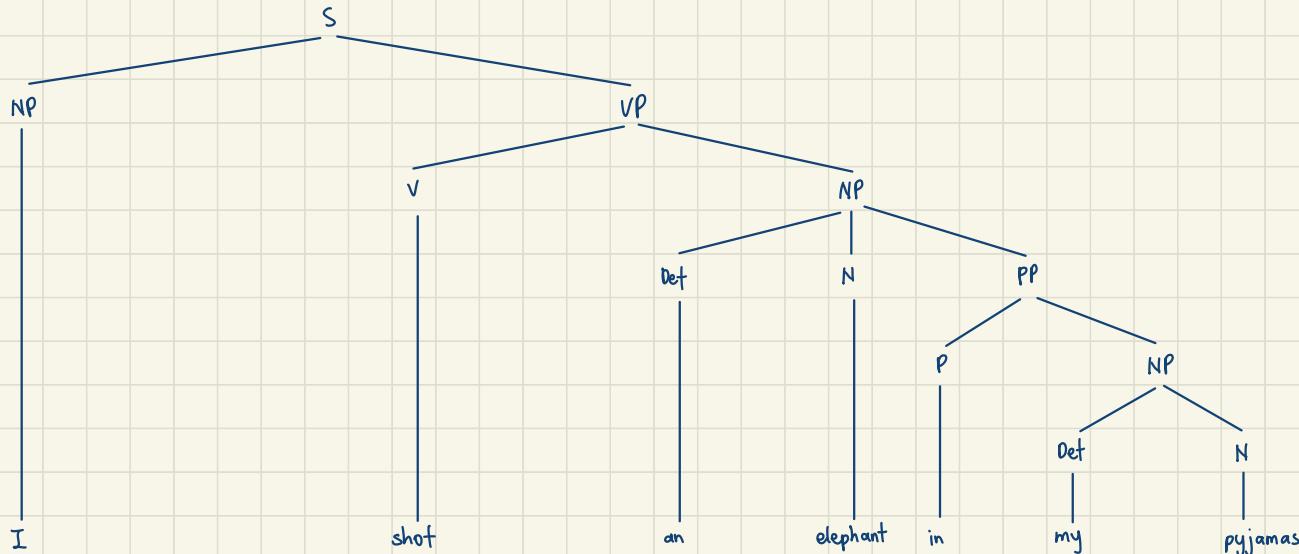
Semantic ambiguity

- The word "shot" may have two different meanings.
- It can be the meaning of shooting something or capturing something.

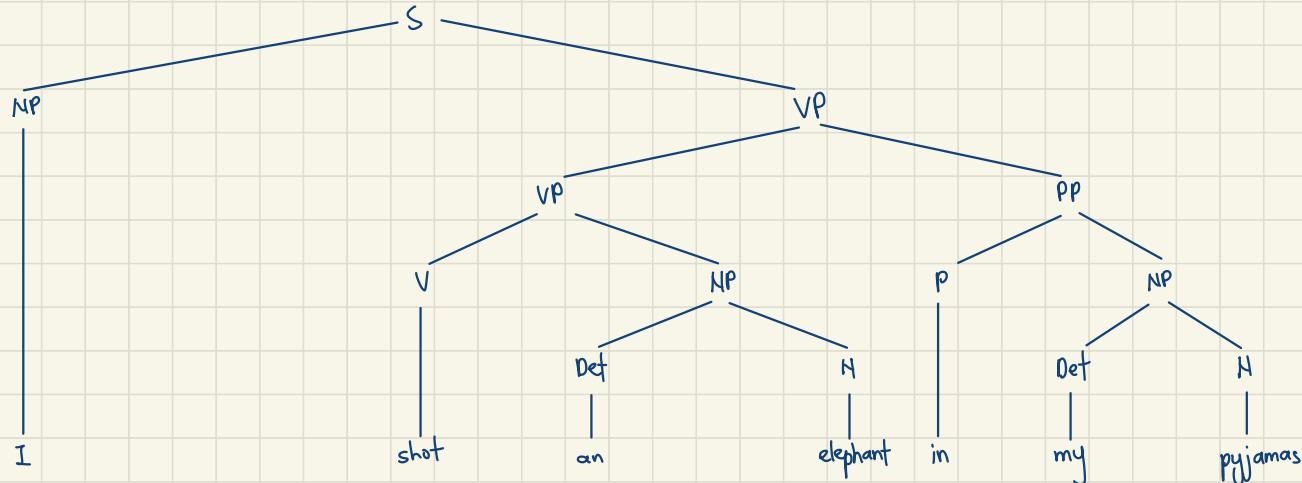
Syntactic ambiguity

- The sentence may have two different syntactic structures although they have the same sequence.
- It can be I shot the elephant which is placed in my pyjamas.
- It can also be I shot the elephant when I am currently in my pyjamas.

Question 2 c)



Question c)



Question 2 d)

Morphology

- It contains the element such as prefixes and suffixes.
- The presence of the elements may alter the meaning of root word.

Semantic

- It represents the meaning of words, phrases and sentences.

BACS2003 ARTIFICIAL INTELLIGENCE**Question 3**

- a) Contrast **TWO (2)** differences between supervised and unsupervised machine learnings. Also, identify and explain **ONE (1)** application for each type of machine learning in academic area. (4 + 4 marks)
- b) K-means is used to cluster the data in Table 1. Two cluster centres ($k=2$) are randomly selected; $C1 = (7.0, 3.0)$ and $C2 = (5.0, 3.0)$.

Table 1: Sample of data

Data point	Feature 1	Feature 2
1	3.2	3.6
2	6.5	7.6
3	4.3	4.2
4	7.8	2.5
5	5.5	6.6
6	5.2	4.3
7	4.6	5.7
8	8.6	1.7

- (i) Compute the clusters that each data point will be assigned to in the first iteration.
 (Note: You are required to demonstrate the step-by-step working step for assigning each data point to its respective cluster). (12 marks)
- (ii) Compute the new centres after the first iteration. (3 marks)
- c) Explain why evaluating the performance of unsupervised machine learning is more challenging than supervised machine learning. (2 marks)

[Total: 25 marks]

Question 3 a)

Supervised machine learning

- It trains model using labeled data which is in form of input-output pairs.
- It requires more human intervention to label the training data with correct output data.
- It can be applied in detecting spam email.

Unsupervised machine learning

- It trains model using unlabeled data which is not provided with correct answer.
- It requires less human intervention. It only requires human interpreted result for validation and understanding.
- It can be applied in segmenting customers based on purchase history, behavior and demographics.

Question 3 b) (i)

Data point	C1	C2
1	3.9471	1.8974
2	4.6271	4.8384
3	2.9547	1.3892
4	0.9434	2.8443
5	3.9	3.6346
6	2.2204	1.3153
7	3.6125	2.7295
8	2.0616	3.8275

We obtain 2 clusters containing: $\{2, 4, 8\}$ and $\{1, 3, 5, 6, 7\}$

Question 3 b) (ii)

$$\begin{aligned} C1_{\text{new}} &= \left(\frac{1}{3}(6.5 + 7.8 + 8.6), \frac{1}{3}(7.6 + 2.5 + 17) \right) \\ &= (7.6333, 3.9333) \end{aligned}$$

$$\begin{aligned} C2_{\text{new}} &= \left(\frac{1}{5}(3.2 + 4.3 + 5.5 + 5.2 + 4.6), \frac{1}{5}(3.6 + 4.2 + 6.6 + 4.3 + 5.7) \right) \\ &= (4.56, 4.88) \end{aligned}$$

Question 3 c)

- Supervised machine learning can be evaluated using various metrics such as precision, accuracy, recall and F1 score.
- Unsupervised machine learning do not have a specific metrics to be evaluated, it can only be evaluated based on quality of clusters, patterns and structure discovered in the data.

BACS2003 ARTIFICIAL INTELLIGENCE**Question 4**

- a) Figure 3 shows a part of scanned image which will be recognised by an optical character recognition (OCR) algorithm. Nevertheless, the detection performance is relatively poor compared to the actual sentence:

"Image processing is a field of study and practice that involves the manipulation of digital images through various algorithms and techniques. It encompasses the".

Identify **FOUR (4)** potential output errors produced by the OCR algorithm. In addition, identify **ONE (1)** pre-processing technique to increase the accuracy of the OCR for each potential output error. (4 + 4 marks)

Image processing is a field of study and practice that involves the manipulation of digital images through various algorithms and techniques. It encompasses the

Figure 3: Scanned image

- b) The Covid-19 diagnosis system incorporates the following rules:

Rule 1: If a patient exhibits loss of taste or smell,
Then the likelihood of the patient being infected with Covid-19 is 0.6

Rule 2: If a patient experiences difficulty in breathing or shortness of breath,
Then the likelihood of the patient being infected with Covid-19 is 0.9

Rule 3: If a patient reports severe headache,
Then the likelihood of the patient being infected with Covid-19 is 0.75

Calculate the inference using both expert rules and observed conditions (Figure 4), employing certainty factors (CF), to determine whether the patient should undergo quarantine. The formulae for combining two rules are shown in Figure 5. (10 + 1 marks)

The patient has loss of taste or smell (CF=0.3)
The patient has difficulty in breathing or shortness of breath (CF=0.67)
The patient has severe headache (CF=0.56)

Figure 4: Observed conditions

$\begin{array}{l} \text{CF1+CF2-CF1*CF2} \\ \text{CF1+CF2+CF1*CF2} \\ \hline \text{CF1+CF2} \\ 1-\min(CF1 , CF2) \end{array}$	$\begin{array}{l} \text{if CF1 and CF2 are positive,} \\ \text{if CF1 and CF2 are negative,} \\ \text{otherwise} \end{array}$
--	---

Figure 5: Formulae for combination of 2 rules

- c) Explain the difference between *crisp set* and *fuzzy set*. (2 marks)
- d) Name **ONE (1)** application of fuzzy logic in home appliance. In addition, design **ONE (1)** fuzzy rules with for the mentioned application. (2 + 2 marks)

[Total: 25 marks]

Question 4 a)

Potential output errors

- "processing" may be detected as "prcessing".
- "study" may be detected as "studv"
- "through" may be detected as "throughb".
- "algorithms" may be detected as "atgorithms".

Pre-processing technique

- Dilation may be applied to the word "processing" to enlarge the object to recover the letter "o".
- Dilation can be applied on the word "study" to enlarge the word "y" so the OCR would detect "y" instead of "v".
- Erosion can be applied on the word "through" to shrink the object for removing or minimizing the "noise" for the letter "h".
- Erosion can be applied on the word "algorithms" to remove the "noise" at the top of the letter "l"

Question 4 b)

$$R1 : CF(\text{Quarantine}) = 0.3 \times 0.6 = 0.18$$

$$R2 : CF(\text{Quarantine}) = 0.67 \times 0.9 = 0.603$$

$$R3 : CF(\text{Quarantine}) = 0.56 \times 0.75 = 0.42$$

$$CF(\text{Quarantine}) = 0.18 + 0.603 - 0.18 \times 0.603 = 0.67446$$

$$CF(\text{Quarantine}) = 0.67446 + 0.42 - 0.67446 \times 0.42 = 0.8111868 \approx 0.8112$$

∴ The patient should undergo quarantine. (CF 0.8112)

Question 4 c)

Crisp set

- It is a set where elements is either belongs to the set or not belong to the set , no in-between.

Fuzzy set

- It is a set where the element can partially belongs to the set.

Question 4 d)

- Smart air conditioner which can adjust the temperature based on the room condition.

R1 IF room is very hot
 THEN temperature of air conditioner is low

R2 IF room is very cold
 THEN temperature of air conditioner is high