TUNKU ABDUL RAHMAN UNIVERSITY OF MANAGEMENT AND TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

ACADEMIC YEAR 2023/2024

MAY/JUNE EXAMINATION

BACS2003 ARTIFICIAL INTELLIGENCE

FRIDAY, 31 MAY 2024 TIME: 3.00 PM – 5.00 PM (2 HOURS)

BACHELOR OF INFORMATION TECHNOLOGY (HONOURS) IN INTERNET TECHNOLOGY
BACHELOR OF INFORMATION TECHNOLOGY (HONOURS) IN SOFTWARE SYSTEMS DEVELOPMENT

BACHELOR OF SOFTWARE ENGINEERING (HONOURS)

Instructions to Candidates:

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

BACS2003 ARTIFICIAL INTELLIGENCE

Question 1

a) Figure 1 depicts the state space of a map, comprising 11 stations. An Artificial Intelligence (AI) search robot employs two search algorithms, Depth First Search (DFS) and A* Search, to determine the shortest travel distance between stations A and K. The values inside the state space show the distance from one station to another in *km*. Table 1 shows the heuristic cost of the stations to Station K.

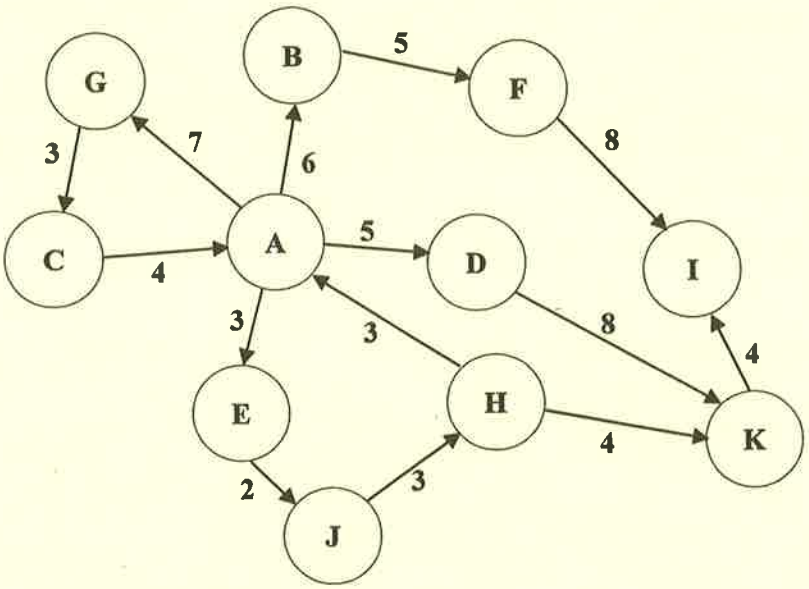


Figure 1: State space

Table 1: Heuristic cost

Station	A	B	C	D	E	F	G	H	I	J	K
Heuristic cost *(km)*	8	9	10	5	7	8	13	4	3	4	0

(i) Formulate the *goal, optimal solution, abstraction, initial state, successor function, goal test* and *step cost* in the problem above. (7 marks)

(ii) For each search algorithm, illustrate a resulting search tree.
(Note: The search should be conducted in alphabetical order). (10 marks)

(iii) For each search algorithm, identify its search path and the solution path returned. (2 + 2 marks)

(iv) Compare the performance between the DFS and A* in terms of completeness and optimality. (4 marks)

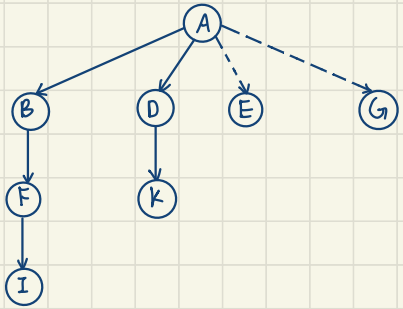
[Total: 25 marks]

Question 1 a) (i)

- Goal : Reach Station K
- Optimal solution : Reach Station K using the shortest distance
- Abstraction : Time
- Initial state : Station A
- Successor function : Function that stores all possible successor state from current state
- Goal test : Test if the current state is equal to the goal state (Station K)
- Step cost : Cost between two station.

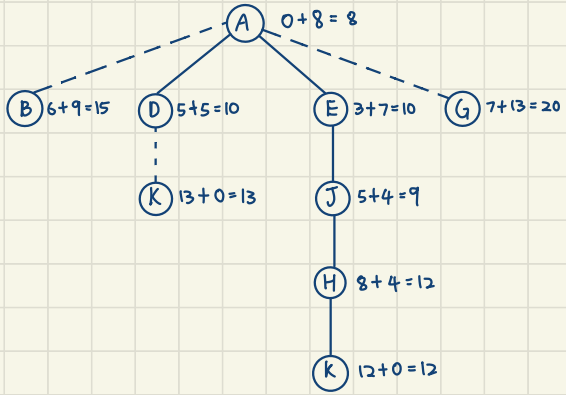
Question 1 a) (ii)

Depth - First Search



Open	Closed
~~A~~	
~~B~~, D, E, G	A
~~F~~, D, E, G	B
~~I~~, ~~∅~~, E, G	F
	I
(K), E, G	D

A* Search



Question 1 a) (iv)

	DFS	A*
Completeness	It is complete since it can successfully find the solution.	It is complete since it has successfully found out the solution.
Optimality	It is not optimal because it does not guarantee the shortest path to achieve the goal even though it can find the solution faster.	It is optimal because it uses Euclidean distance heuristic to find out the shortest path to reach the goal.

Question 2

a) Below shows the segment of the knowledge.

Cars can be broadly classified into three types based on their power source: petrol (or gasoline), electric, and hybrid. Petrol cars, powered by spark-ignited internal combustion engines, tend to emit more CO_2, contributing to environmental pollution. They exhibit the highest noise levels compared to the other two types of cars, but they boast the longest driving range. On the other hand, electric cars run on electricity, producing zero tailpipe emissions and maintaining a lowest noise level. Their driving range is shorter and depends on battery capacity. As an example, the Tesla Model 3 represents an electric car. Hybrid cars, utilising both petrol (gasoline) and electricity, typically emit lower emissions than petrol cars due to reduced fuel consumption. The noise level in hybrid cars is generally lower than in petrol cars, especially when the vehicle operates on electric power alone. Hybrid cars often have a longer driving range than electric cars because they can switch to the petrol engine when the battery is low.

(i) Frames and semantic networks are two commonly used methods for representing knowledge before system coding. Explain why frames are a more suitable choice for representing the aforementioned knowledge. (2 marks)

(ii) Illustrate the knowledge above with frames. (11 marks)

b) Consider a sentence *S* = *The chicken is ready to eat on the table.*
Explain semantic and syntactic ambiguities found in the sentence *S*. (4 marks)

c) Illustrate **ONE (1)** parse tree for the sentence *S* provided in **Question 2 b)**, with the use of the grammar rules provided in Figure 2. (8 marks)

```
noun_phrase(NP)
verb_phrase(VP)
preposition_phrase(PP)
determiner(Det)
noun(N)
verb(V)
preposition(P)
adjective(ADJ)
S -> NP VP
NP -> Det N | N PP | Det NP
VP -> V PP | V ADJ | VP PP
PP -> P NP | P VP
N -> [chicken, table]
V -> [is, eat]
P -> [to, on]
Det -> [the]
ADJ -> [ready]
```

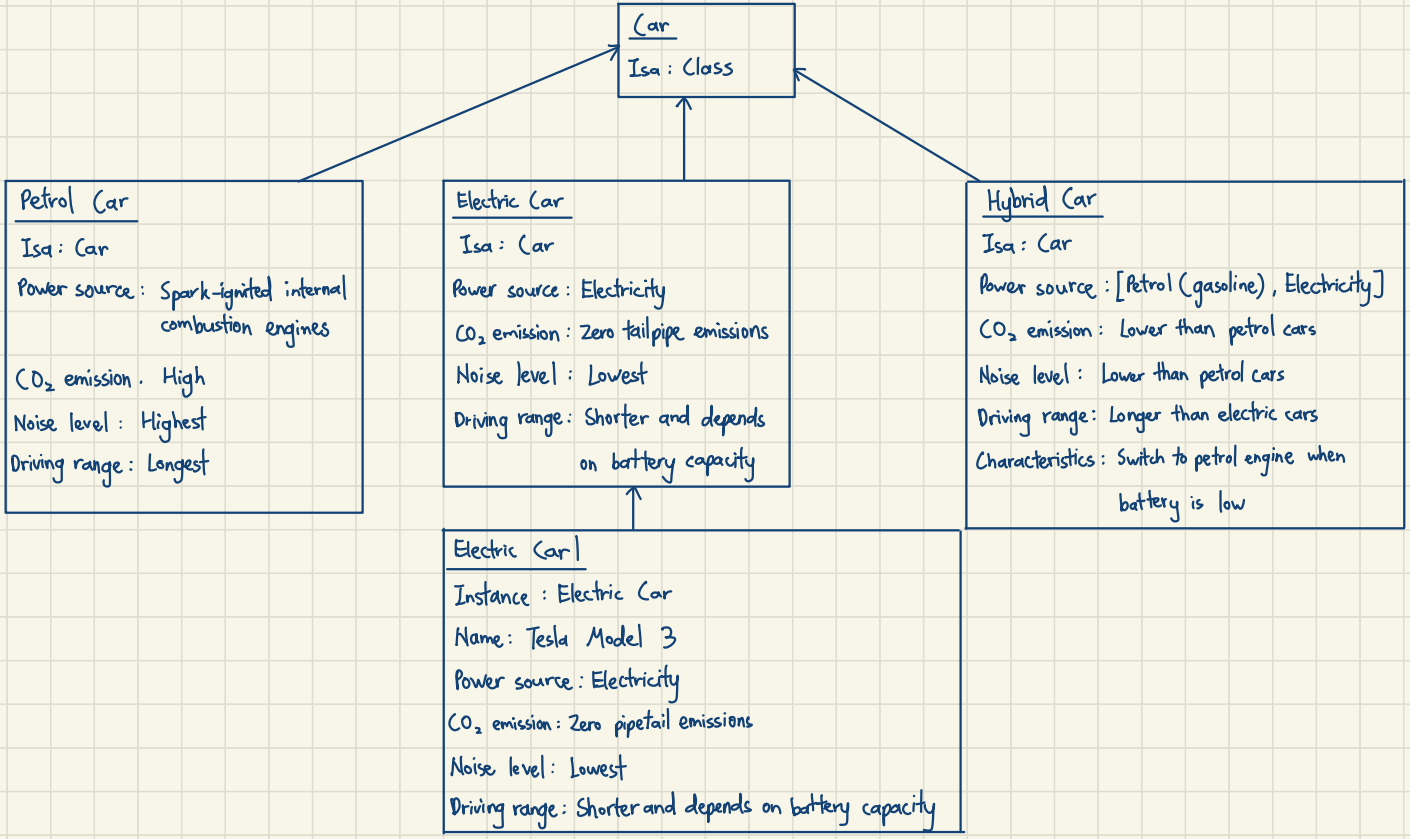
Figure 2: Grammar rules

[Total: 25 marks]

Question 2 a) (i)

- Frames are more suitable to show the details of objects and their relationship.

Question 2 a) (ii)



Question 2 b)

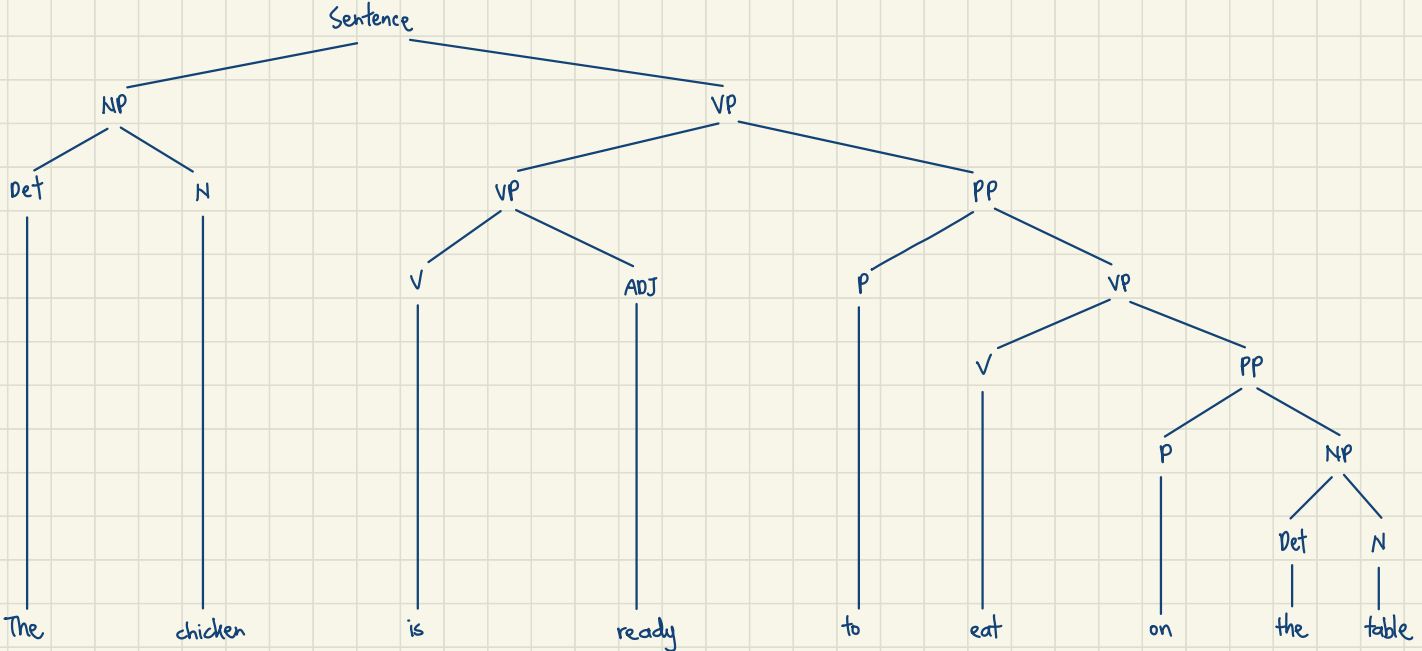
Semantic ambiguity

- The meaning of the word "table" can be the furniture used for putting things on or a virtual table with rows and columns.

Syntactic ambiguity

- The sentence can mean that the chicken is ready to be eaten on the table.
- It can also be the chicken has get ready to eat something on the table.

Question 2 c)



BACS2003 ARTIFICIAL INTELLIGENCE

Question 3

a) Table 2 shows the partial data collection for stroke patients. The data is classified into two different classes which are Yes (1), and No (0). Identify the class that a new patient belongs to by using K-Nearest Neighbour (K-NN) algorithm given that the new patient's avg_glucose_level and bmi are 200.7 and 30.5, respectively (**Note:** K = 5).

Table 2: Attributes of the stroke dataset

patient_id	avg_glucose_level	bmi	stroke
9046	228.69	36.6	1
38047	100.98	28.2	1
56669	186.21	29.0	1
2374	213.37	36.0	0
15528	223.36	41.5	0
70374	122.41	40.3	0
1665	174.12	24.0	1
29908	103.26	25.4	0
11091	75.39	37.8	0
53882	70.09	27.4	1

Source: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset?resource=download

(13 marks)

b) A team of researchers aims to predict risk of stroke using a multilayer Neural Network (NN). The dataset that the researchers use is shown in Table 3. Design an architecture for the multilayer NN based on the attributes in the Table 3 and label the structure clearly. (6 marks)

Table 3: Stroke attributes for multilayer NN

No	Label Name	Unit of Measure	Data Type
1	Patient ID	-	Numeric
2	Gender	Categorical	Numeric
3	Smoking Status	Categorical	Numeric
4	Hypertension	Categorical	Numeric
5	BMI	kg/m^2	Numeric
6	Age	years	Numeric
7	Stroke	Categorical	Numeric

c) Table 4 shows a confusion matrix. Calculate the accuracy, precision and recall for "Yes" case. (6 marks)

Table 4: Confusion matrix

	Predicted Yes	Predicted No
Actual Yes	74	8
Actual No	9	54

[Total: 25 marks]

Question 3 a)

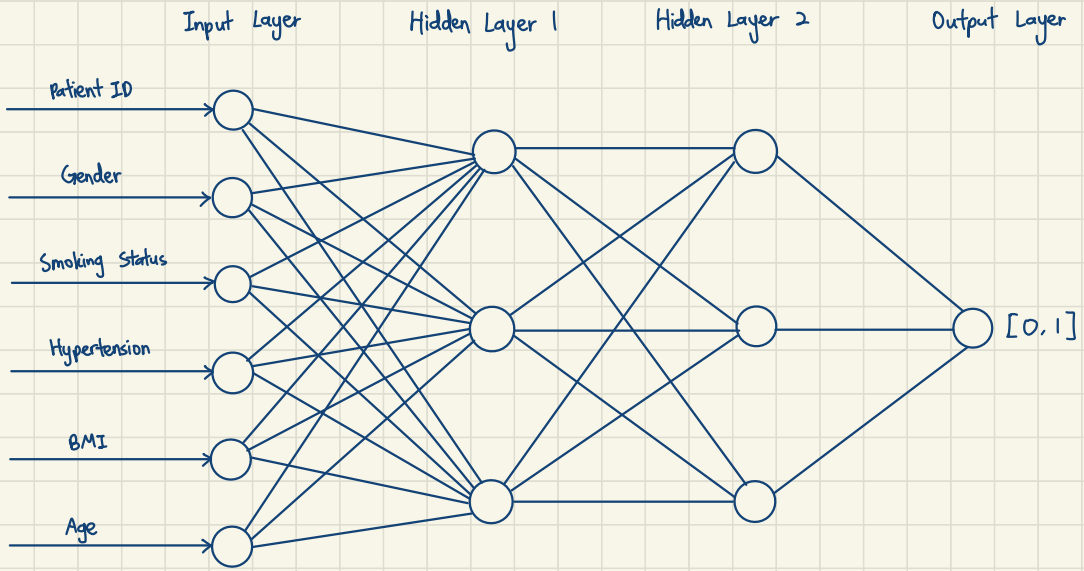
patient_id	avg_glucose_level	bmi	stroke	Distance	Ranking
9046	228.69	36.6	1	28.647	5
38047	100.98	28.2	1	99.7465	
56669	186.21	29.0	1	14.5674	2
2374	213.37	36.0	0	13.8123	1
15528	223.36	41.5	0	25.1888	3
70374	122.41	40.3	0	78.901	
1665	174.12	24.0	1	27.3632	4
29908	103.26	25.4	0	97.5734	
11091	75.39	37.8	0	125.5225	
53882	70.09	27.4	1	130.6468	

$P(\text{Yes} \mid \text{New patient}) = \frac{3}{5} = 0.6$

$P(\text{No} \mid \text{New patient}) = \frac{2}{5} = 0.4$

$\therefore$ Since $P(\text{Yes} \mid \text{New patient}) > P(\text{No} \mid \text{New patient})$, the new patients are most probably belongs to class 'Yes (1)'.

Question 3 b)


Input Layer
Hidden Layer 1
Hidden Layer 2
Output Layer
Patient ID
Gender
Smoking Status
Hypertension
BMI
Age
[0, 1]

Question 3 c)

$$\text{Accuracy} = \frac{74 + 54}{74+8+9+54} = \frac{128}{145} = 0.8828$$

$$\text{Precision} = \frac{74}{74 + 9} = \frac{74}{83} = 0.8916$$

$$\text{Recall} = \frac{74}{74 + 8} = \frac{74}{82} = 0.9024$$

Question 4

a) Figure 3 shows the pixel values of an image. Due to the presence of noise, this image will undergo a filtering process known as median filtering. Explain the workings of median filtering. Also, perform median filtering onto the image in Figure 3 by using 3×3 kernel.
(**Note:** Replace the pixel value to 0 for border problem due to not enough neighbouring pixels to perform the filtering).

(3 + 11 marks)

115	145	145	109	135	146
163	178	200	109	157	138
163	198	201	118	144	145
163	198	201	109	135	124
157	198	201	118	144	109
143	153	171	173	170	168
137	144	168	143	161	141

Figure 3: Pixel values of an image

b) The following is the rule base for hypertension detection used in an expert system.

R1: IF systolic blood pressure is consistently greater than 140 mmHg
AND diastolic blood pressure is consistently greater than 90 mmHg
THEN there is a high likelihood of hypertension. (0.9)

R2: IF the patient has a family history of hypertension
AND body mass index (BMI) is above 30
OR age is above 60
THEN there is a high likelihood of hypertension. (0.85)

R3: IF there is a history of chronic kidney disease
AND regular consumption of high-sodium diet
OR has symptoms such as headaches and dizziness
THEN there is a high likelihood of hypertension. (0.75)

Figure 4 shows the observed conditions of a new patient. Based on the expert rules above and the observed conditions, calculate the inference and conclude the likelihood of hypertension for the new patient by using certainty factor (CF). The formulae for combination of two rules are provided in Figure 5. (10 + 1 marks)

Question 4 a)

- Median filtering operates over a window by selecting the median intensity in the window.
- It determines the value of a central pixel (in a field kernel size 3X3) and supported by all its neighbors pixel values.

0	0	0	0	0	0
0	163	145	144	138	0
0	198	198	144	135	0
0	198	198	144	124	0
0	171	173	170	135	0
0	157	168	168	144	0
0	0	0	0	0	0

Question 4 b)

R1 : $CF(Hypertension) = \min(0.85, 0.85) \times 0.9 = 0.85 \times 0.9 = 0.765$

R2 : $CF(Hypertension) = \max(\min(-0.8, 1), 1.0) \times 0.85 = \max(-0.8, 1.0) \times 0.85 = 1.0 \times 0.85 = 0.85$

R3 : $CF(Hypertension) = \max(\min(1.0, 0.8), 0.75) \times 0.75 = \max(0.8, 0.75) \times 0.75 = 0.8 \times 0.75 = 0.6$

$CF(Hypertension) = 0.765 + 0.85 - 0.765 \times 0.85 = 0.96475$

$CF(Hypertension) = 0.96475 + 0.6 - 0.96475 \times 0.6 = 0.9859$

Conclusion : The new patient may get the hypertension. (CF 0.9859)

Question 4 b) (Continued)

Systolic blood pressure is consistently above 140 mmHg	= 0.85
Diastolic blood pressure is consistently above 90 mmHg	= 0.85
The patient has a family history of hypertension	= -0.8
BMI is above 30	= 1.0
Patient's age is above 60	= 1.0
History of chronic kidney disease	= 1.0
Regular consumption of high-sodium diet	= 0.8
Patient reports symptoms such as headaches and dizziness	= 0.75

Figure 4: Observed condition of a new patient

$CF1+CF2-CF1*CF2$	if CF1 and CF2 are positive,
$CF1+CF2+CF1*CF2$	if CF1 and CF2 are negative,
$\frac{CF1+CF2}{1-\min(\lvert CF1\rvert,\lvert CF2\rvert)}$	otherwise

Figure 5: Formulae for combination of two rules

[Total: 25 marks]