TUNKU ABDUL RAHMAN UNIVERSITY OF MANAGEMENT AND TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

ACADEMIC YEAR 2024/2025

JANUARY EXAMINATION

BACS2003 ARTIFICIAL INTELLIGENCE

TUESDAY, 7 JANUARY 2025   TIME: 2.00 PM – 4.00 PM (2 HOURS)

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) Provide **TWO (2)** applications of AI technologies that enhance business operations. Explain how each application works. (2 + 2 marks)

b) Figure 1 depicts a state space consisting of 11 stations, the values inside the state space show the distance between the stations. Table 1 provides the heuristic costs for each station in relation to the destination, Station K. Use the A* search method to determine the shortest path from Station A to Station K.
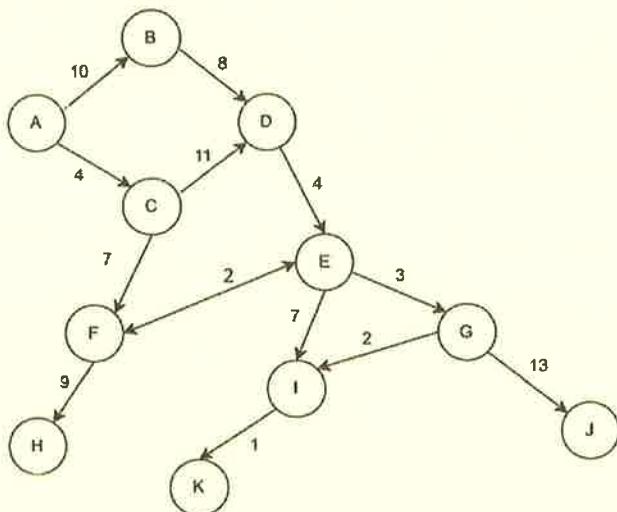

Table 1: Heuristic cost of each station

Station	Cost
A	12
B	14
C	8
D	9
E	4
F	7
G	4
H	8
I	3
J	2
K	0

Figure 1: State space

(i) Formulate the goal, optimal solution, abstraction, initial state, successor function, goal test, step cost and path cost for the problem above. (8 marks)

(ii) Based on the information provided in Figure 1 and Table 1, illustrate the resulting search tree for A* search in alphabetical order. (**Remark**: No repeated state is allowed). (12 marks)

(iii) Identify the returned path by using A* search in **Question 1 b) (ii)**. (1 mark)

[Total: 25 marks]

Question 2

a) The information below describes the process of course enrolment at TAR UMT.

Edwina is a student who just finished her SPM at SMK Jalan Tasek. She attends the open day organised by TAR UMT in the sports complex. Inside, there are many booths, including the engineering faculty, computing faculty, applied science faculty, business faculty, and others. Edwina has no idea which course to choose, so she consults a few counsellors. Finally, she decides to enrol in the computing faculty because she has an interest in coding. She enrols in the course on the same day and goes back to her hometown, Ipoh, with her family.

This question paper consists of 4 questions on 5 printed pages.

Question 1 a)

- A business-designated chatbot can apply NLP to effectively respond the customer's questions within the short time.
- Clustering algorithms can be used to segment customers based on behaviors, purchase history and demographics to create a targeted business strategy.

Question 1 b)

(i) Goal : Reach station K

Optimal solution : Reach station K using the shortest distance

Abstraction : Time

Initial state : Station A
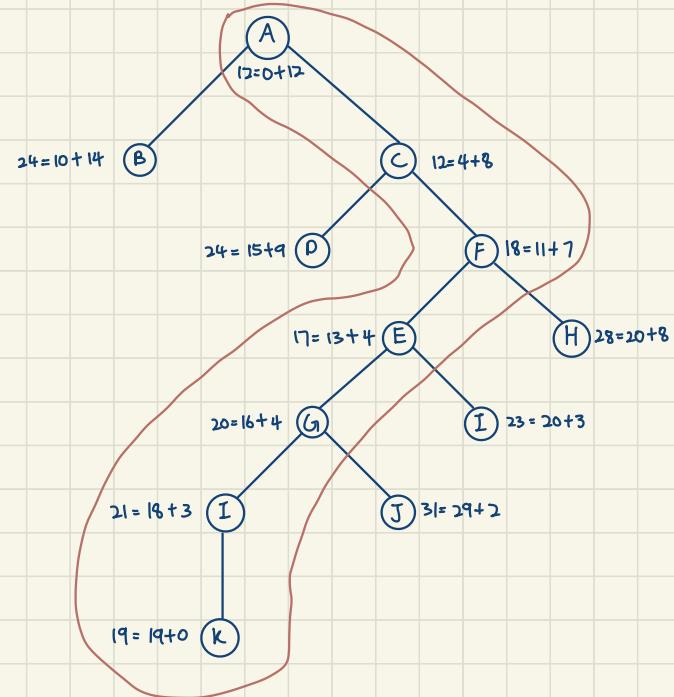
Successor function : Function storing all successor state from current state

Goal test : To test if the current state, is equal to the goal state

Step cost : Cost between two stations

Path cost : Total cost from station A to station K

(ii)

A $12 = 0 + 12$

B $24 = 10 + 14$

C $12 = 4 + 8$

D $24 = 15 + 9$

F $18 = 11 + 7$

E $17 = 13 + 4$

H $28 = 20 + 8$

G $20 = 16 + 4$

I $23 = 20 + 3$

I $21 = 18 + 3$

J $31 = 29 + 2$

K $19 = 19 + 0$
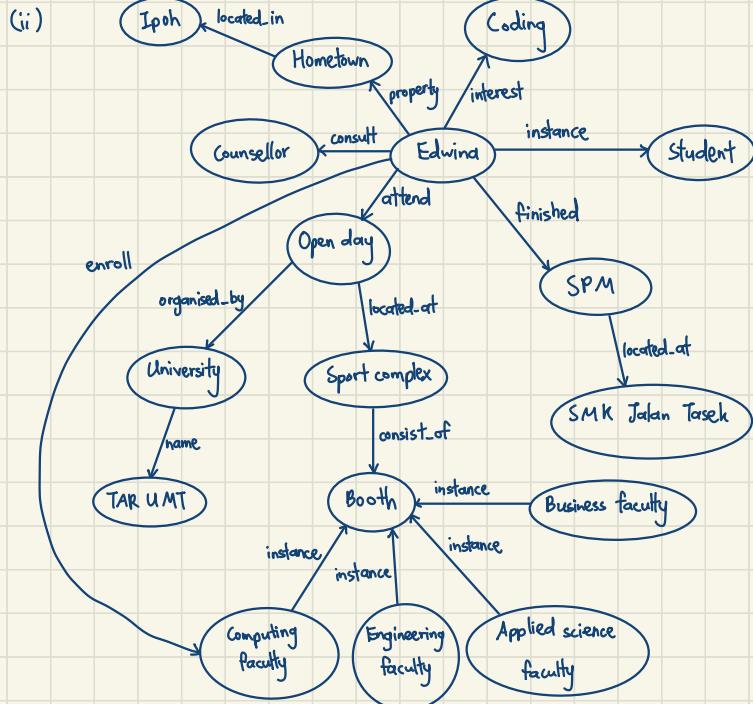
(iii) A - C - F - E - G - I - K

Question 2   a)

(i) - Semantic network is more suitable to describe the relationship between objects and overall concepts.

(ii)

Ipoh —located_in→ Hometown

Hometown —property→ Edwina

Edwina —interest→ Coding

Edwina —consult→ Counsellor

Edwina —instance→ Student

Edwina —attend→ Open day

Edwina —finished→ SPM

Edwina —enroll→ Computing faculty

Open day —organised_by→ University

Open day —located_at→ Sport complex

University —name→ TAR UMT

SPM —located_at→ SMK Jalan Taseh

Sport complex —consist_of→ Booth

Booth ←instance— Business faculty

Computing faculty —instance→ Booth

Engineering faculty —instance→ Booth

Applied science faculty —instance→ Booth

Question 2 a) (Continued)

(i) Frames and semantic networks are among the standard knowledge representation methods to represent information before we design the coding of a system. Explain why semantic networks are more appropriate to represent the information above.

(2 marks)

(ii) Illustrate the information above with semantic network. Ensure that the relationships and concepts in the semantic network are properly labelled. (14 marks)

b) Provide **ONE (1)** application of Natural Language Understanding (NLU) and explain **TWO (2)** challenges to build the application. (5 marks)

c) Provide and explain **TWO (2)** symbolic analyses in NLU. (4 marks)

[Total: 25 marks]

Question 3

a) There are two basic approaches of machine learning which are supervised machine learning and unsupervised machine learning. Compare **TWO (2)** differences between them. (4 marks)

b) Figure 2 shows the data distribution of heart disease patients based on the 2 features which are "Resting blood pressure" and "Serum cholesterol". The data has been classified into 2 different classes: Yes (with heart disease) and No (without heart disease).
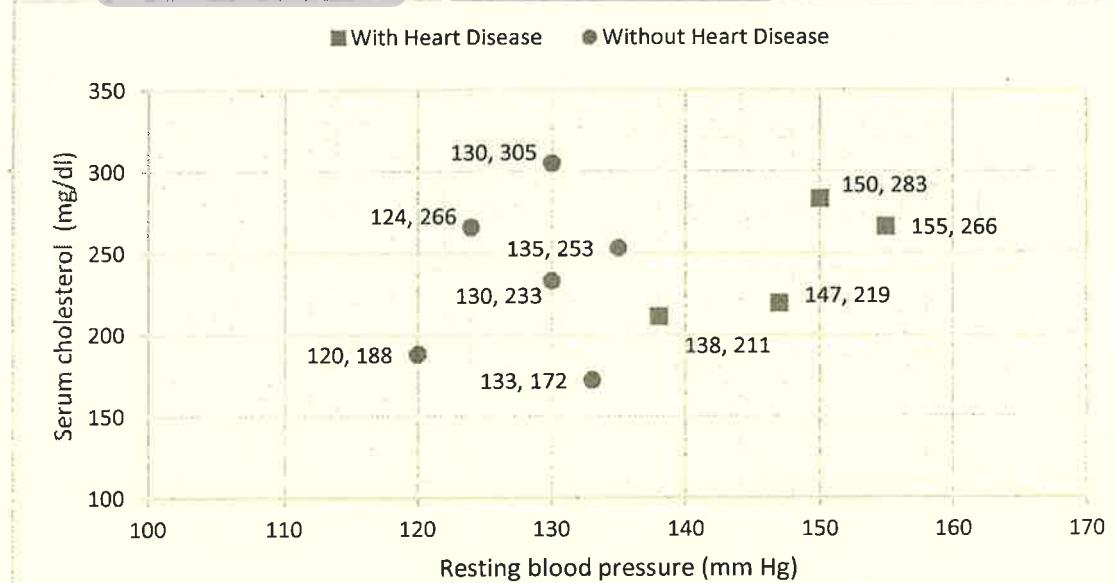

Figure 2: Data distribution for Serum cholesterol against Resting blood pressure

Determine the probability of a patient belonging to each class using the K-Nearest Neighbour (K-NN) algorithm, with given Resting blood pressure of 142 mm Hg and Serum cholesterol of 290 mg/dl. Additionally, analyse the class to which the patient belongs. (**Remark**: K = 5).

(12 + 1 marks)

Question 2 b)

- ChatGPT applies Natural Language Understanding (NLU) to understand the users' queries with different intent and context.
- Human may provides text or speech input what might contain a lot of spelling mistakes, grammar error or various slang The application has to ∧ analyze the actual meaning of the context given by the users while preserving the users' actual intention.
- Human always leave out many information they assume the listener knows when providing input to the application. So, the application requires coreference resolution and shared context inference why are hard to achieve.

correct and

Question c)

Morphology
- It contains the components like prefixes and suffixes
- Those components may change the meaning of the root words.

Syntax
- It is the rule of combining words into phrases and sentences.
- It is related to the grammar.

Question 3 a)

Supervised learning

- It trains model using labeled data where the input provided is paired with the correct output data or target value.

- It learns the mapping from the information in input-output form to predict the output for new and unseen inputs accurately.

Unsupervised learning

- It trains model using unlabeled data where the inputs are not coming with a correct answer or target value.

- It learns to identify the patterns, clusters and relationship in the data to discover hidden patterns, grouping and structures with the data.

Question 3 b)

Resting Blood Pressure (mm Hg)	Serum Cholesterol (mg/dl)	Heart Disease	Distance	Ranking	$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
120	188	0	104.3456		
124	266	0	30	4	
130	233	0	58.2495		
130	305	0	19.2094	2	
133	172	0	118.3427		
135	253	0	37.6563	5	
138	211	1	79.1012		
147	219	1	71.1758		
150	283	1	10.6301	1	
155	266	1	27.2947	3	

$P(\text{Yes}) = \frac{2}{5} = 0.4$

$P(\text{No}) = \frac{3}{5} = 0.6$

$\therefore$ Since $P(No) > P(Yes)$, the patients are most likely belong to class `No (without heart disease)`.

$$Accuracy = \frac{TP+TN}{Total}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

Question 3 (Continued)

c)   Table 2 shows a confusion matrix. Identify the *accuracy, precision* and *recall* for "Yes" case.
(6 marks)

Table 2: Confusion matrix

	Predicted Yes	Predicted No
Actual Yes	96 TP	4 FN
Actual No	8 FP	86 TN

d)   Identify **TWO (2)** methods based on internal information that can be applied to evaluate a clustering algorithm.
(2 marks)

[Total: 25 marks]

Question 4

a)   List and explain the **THREE (3)** properties of an image.
(3 + 3 marks)

b)   Figure 3 shows a part of scanned image which will be recognised by an Optical Character Recognition (OCR) algorithm. However, the detection performance is relatively poor compared to the actual paragraph as shown in Figure 4. Identify **TWO (2)** potential output errors produced by the OCR algorithm. In addition, suggest **TWO (2)** pre-processing techniques and explain how each of them helps to increase the accuracy of the OCR algorithm.

Image processing refers to the manipulation and analysis of images using computer algorithms. It involves various techniques to enhance, transform, or extract useful information from images.

Figure 3: Scanned image

Image processing refers to the manipulation and analysis of images using computer algorithms. It involves various techniques to enhance, transform, or extract useful information from images.

Figure 4: Original text

(4 + 4 marks)

c)   The following is the rule-base for detection of heart disease.

Rule 1
If   Age > 45 AND cholesterol level > 240 mg/dL
Then   High risk of heart disease. (CF=0.95)

BACS2003 ARTIFICIAL INTELLIGENCE

Question 4 c) (Continued)

Rule 2
If   Systolic blood pressure > 140 mmHg OR diastolic blood pressure > 90 mmHg
Then   High risk of heart disease. (CF=0.80)

Rule 3
If   Patient has a family history of heart disease AND is overweight (BMI > 25)
Then   High risk of heart disease. (CF=0.75)

Figure 5 shows the observed conditions of a new patient. Based on the expert rules above and the observed conditions, analyse the risk of heart disease for the new patient by using certainty factor (CF). If necessary, you are required to use the formulae for combination of 2 rules provided in Figure 6.

(11 marks)

1.	Age > 45 (CF=1.0)
2.	Cholesterol level > 240 mg/dL (CF=0.98)
3.	Systolic blood pressure > 140 mmHg (CF=0.77)
4.	Diastolic blood pressure > 90 mmHg (CF=0.85)
5.	Patient has a family history of heart disease (CF=0.40)
6.	Overweight (BMI > 25) (CF=0.95)

Figure 5: Condition of new patient

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

Figure 6: Formulae for combination of 2 rules

[Total: 25 marks]

This question paper consists of 4 questions on 5 printed pages.

Question 3 c)

$$\text{Accuracy} = \frac{96 + 86}{96 + 4 + 8 + 86}$$

$$= \frac{182}{194}$$

$$= 0.9381$$

$$\text{Precision} = \frac{96}{96 + 8}$$

$$= \frac{96}{104}$$

$$= 0.9231$$

$$\text{Recall} = \frac{96}{96 + 4}$$

$$= \frac{96}{100}$$

$$= 0.96$$

Question 3 d)

- Intra-cluster cohesion
- Inter-cluster separation

Question 4 a)

Spatial resolution
- It determines the number of independent pixel values per inch.

Number of channel
- It determines the number of separate or distinct components or bands of information that make up an image.

Intensity resolution
- It shows the ability of imaging system to distinguish between different levels of brightness and intensity in an image.
- It also determines the image color depth and range of color and shades it can display.

Question 4 b)
- The word "and" may be detected as "ano".
- The word "analysis" may be detected as "anaiysis".

- Dilation can be applied to enlarge the object word of "and" to make all the part of the word linked together to ensure the accurate OCR detection.
- Closing can be applied to enlarge the object word of "analysis", then shrink the word back to normal word font to enable the letter "l" being recognized by OCR, instead of letter "i".

Question 4 c)

R1 : CF (Heart disease) = min (1.0, 0.98) × 0.95 = 0.98 × 0.95 = 0.931

R2 : CF (Heart disease) = max (0.77, 0.85) × 0.80 = 0.85 × 0.80 = 0.68

R3 : CF (Heart disease) = min (0.40, 0.95) × 0.75 = 0.40 × 0.75 = 0.3

CF (Heart disease) = CF1 + CF2 − CF1 × CF2

$$= 0.931 + 0.68 − 0.931 × 0.68$$

$$= 0.97792$$

CF (Heart disease) = 0.97792 + 0.3 − 0.97792 × 0.3

$$= 0.984544$$

Conclusion : The new patient has the risk of getting heart disease. (CF 0.984544)