THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

This exam paper must not be removed from the venue

Venue ____________________

Seat Number ________

Student Number |__|__|__|__|__|__|__|__|

Family Name ____________________

First Name ____________________

School of Information Technology and Electrical Engineering

EXAMINATION

Semester Two Final Examinations, 2020

COMP3702/7702 Artificial Intelligence

This paper is for St Lucia Campus students.

Examination Duration: 90 minutes

Reading Time: 10 minutes

Exam Conditions:

Set start and completion time for all students e.g. a 2 hour exam, starts at 8am, ends at 10am

Paper-based exam (on-campus exam only)

This is a Closed Book examination - specified written materials permitted

Casio FX82 series or UQ approved (labelled)

Materials Permitted In The Exam Venue:

(No electronic aids are permitted e.g. laptops, phones)

One A4 sheet of handwritten or typed notes double sided is permitted

Blank scrap paper permitted - any number of A4 sheets permitted

Materials To Be Supplied To Students:

1 x 6-Page Answer Booklet

Instructions To Students:

Additional exam materials (eg. answer booklets, rough paper) will be provided upon request.

This exam consists of multiple choice and written answer questions. Please answer all questions in the booklet provided.

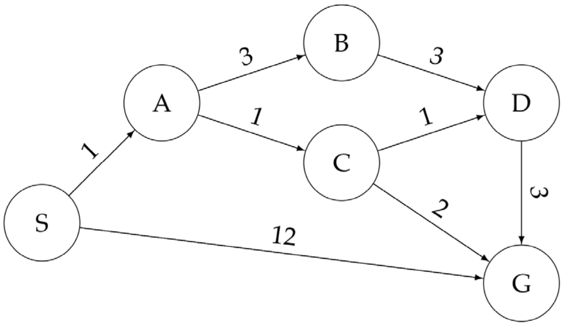
For Examiner Use Only

Question	Mark

Total ________

Question 1. **(20 marks)**

Answer the following questions about the search problem shown below:



The initial state is S and the goal state is G. For the questions that ask for a path, please give your answers in the form ‘S–<state>–…-<state>–G’. Break any priority ties alphabetically.

a) (5 marks) What path does breadth-first search return for this search problem?

b) (5 marks) What path does uniform cost search return for this search problem?

c) (5 marks) What path does an A* graph search, using a consistent heuristic, return for this search problem?

d) Consider the heuristics for this search problem shown in the table below:

State	*h1*	*h2*
S	5	4
A	3	2
B	6	6
C	2	1
D	3	3
G	0	0

i. (2.5 marks) Is *h1* admissible? Yes or no.

ii. (2.5 marks) Is *h2* admissible? Yes or no.

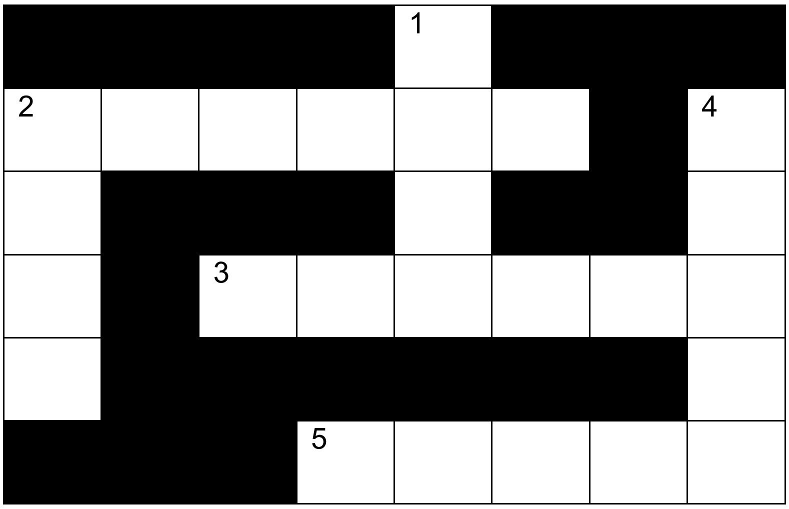
Reminder: Put your answers to all questions in the answer booklet

Question 2. **(5 marks)**

Mr Search says that informed search is always more efficient than blind search. Is Mr Search correct? Give reasons for your answer.

Question 3 **(25 Marks)**

Consider the following crossword puzzle with candidate words, which can be cast as a *constraint satisfaction problem* (CSP):



Words:

ADA
ALEX
ALINA
ARCHIE
JOHN
NICK
RENEE
SERGE
VEKTOR
VINT

Denote the variables: *1D* (1-down), *2A* (2-across), *2D*, *3A*, *4D* and *5A*. Note that words can only be used once in a solution.

a) (2.5 marks) List all binary constraints between variables for this CSP.

b) (3 marks) Apply *domain consistency* to this CSP. List the resulting variable domains.

c) (9 marks) Apply *arc consistency* to the domain-consistent CSP from b). List the resulting variable domains. List the arc-consistent variable domains.

d) (9 marks) Apply *backtracking search* to the domain-consistent CSP from question b). Use the variable ordering (1D, 2A, 2D, 3D, 4A, 5A) and the variable order in the *Words* list to expand nodes in the search graph. List all variable assignment and removal operations, and any backtracking operations.

e) (1.5 marks) What is the solution to this CSP?

Question 4 **(25 marks)**

s0	*s1*	*s2*	*s3*	*s4*	*s5*
5		★			10
			0	0	

Consider the above gridworld. An agent is currently on grid cell *s2*, as indicated by the star, and would like to collect the rewards that lie on both sides of it. If the agent is on a numbered square (0, 5 or 10), the instance terminates and the agent receives a reward equal to the number on the square. On any other (non-numbered) square, its available actions are to move Left and Right. Note that Up and Down are never available actions.

If the agent is in a square with an adjacent square below it, it does not always move successfully: when the agent is in one of these squares and takes a move action, it will only succeed with probability p. With probability $1 - p$, the move action will fail and the agent will instead fall downwards into a trap. If the agent is not in a square with an adjacent space below, it will always move successfully.

For parts a), b) and c), we are using discount factor $\gamma \in [0,1]$.

a) (5 marks) Consider the policy π_{Right}, which is to always move right when possible. For each state $s \in \{s1, s2, s3, s4\}$ in the diagram above, give the value function $V^{\pi_{Right}}$ in terms of γ and p.

b) (5 marks) Consider the policy π_{Left}, which is to always move left when possible. For each state $s \in \{s1, s2, s3, s4\}$ in the diagram above, give the value function $V^{\pi_{Left}}$ in terms of γ and p.

c) (2.5 marks) For what range of values of p is it optimal for the agent to go left from the start state (*s2*, represented by the star)? Express your solution in terms of γ.

For parts d), e) and f), let the discount factor $\gamma = 0.9$, and let the probability of falling into a trap $p = 0.8$ for both *s3* and *s4* (independently).

d) (5 marks) Compute the value function for the gridworld problem above.

e) (5 marks) Compute the Q-function for the gridworld problem above.

f) (2.5 marks) What is the optimal policy in the gridworld problem above when $\gamma = 0.9$ and $p = 0.8$?

Question 5 **(10 Marks)**

We continue to consider the gridworld from the previous page, repeated here:

s0	*s1*	*s2*	*s3*	*s4*	*s5*
5					10
			0	0	

In question 5 we assumed knowledge of the transition function $T(s, a, s')$. Now, in this question, assume we do not know the transition function.

a) (6 marks) Suppose we choose to use Q-learning (in absence of the transition function) and we obtain the following observations:

s_t	a	s_{t+1}	reward
s3	Left	s2	0
s2	Left	s1	0
s1	Left	s0	5
s4	Right	s4-below	0

What values does the Q-function attain if we initialise the Q-values to 0 and replay the experience in the table exactly two times? Use a learning rate $\alpha = 0.6$, discount factor $\gamma = 0.9$ and the probability of falling into a trap of $p = 0.8$ for both $s3$ and $s4$ (independently).

b) (2 marks) Which of the following can be used to obtain a policy if we don't know the transition function? (Choose all correct options. Put your answer in the answer booklet.)
i) Value Iteration followed by Policy Extraction
ii) Approximate Q-learning
iii) TD learning followed by Policy Extraction
iv) Policy Iteration with a learned $T(s, a, s')$

c) (2 marks) Under which conditions would one benefit from using approximate Q-learning over vanilla Q-learning? (Select one only. Put your answer in the answer booklet.)
i) When the state space is very high-dimensional
ii) When the transition function is known
iii) When the transition function is unknown
iv) When the discount factor is small

Question 6 **(15 Marks)**

Anne is a COMP3702 student who enjoys playing a simple game of chance on her phone while waiting for her lecture to begin. The game involves pressing one of two buttons, *a1* or *a2* each round, which return a reward r_t. The goal is to maximise the average reward accumulated over time. The rewards for pushing a button are not know to Anne, but she does know that they follow a fixed distribution. She is investigating two algorithms, ϵ-greedy and the upper-confidence bound (UCB).

In one instance of the game, Anne takes the actions and receives the rewards in the table below:

t	Action	r_t
1	*a1*	2
2	*a2*	1
3	*a1*	2
4	*a2*	0
5	*a1*	0
6	*a1*	0

a) (3 marks) If Anne uses ϵ-greedy, which button is she most likely to press next?

b) (4 marks) Anne is considering a variation on the UCB algorithm that uses the following confidence bound estimate:

$$UB_t(a) = \sqrt{\frac{4}{N_t(a)}}$$

where $N_t(a)$ is the number of times action *a* has been chosen by round *t*. Using this UCB algorithm, which button should Anne press next?

Anne is curious as to how the game works, so inspects the source code This reveals that the rewards are encoded as a repeated game in normal form with the following payoff matrix:

	Game app	
Anne	*b1*	*b2*
a1	(2,-1)	(0,1)
a2	(0, 1)	(1,0)

c) (3 marks) The source code also shows that the game app has a fixed strategy of playing *b1* with probability 0.6. With this knowledge, what strategy should Anne use to play the game?

d) (5 marks) Bing is the app developer. He notices Anne's success (and that she has accessed the game source code). What can Bing do to improve his strategy and maximise the payoff to his game app? Give a suitable strategy and state your reasons for choosing it.

END OF EXAMINATION