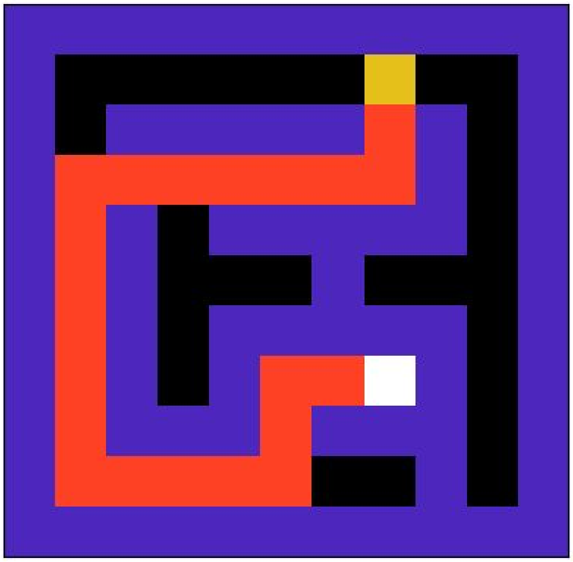
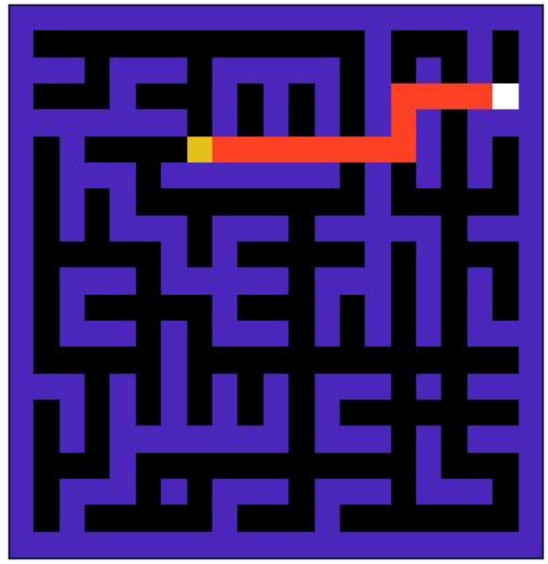
CS747 Assignment - 2 Report

Name - Saurabh Parekh
Roll number - 170100016

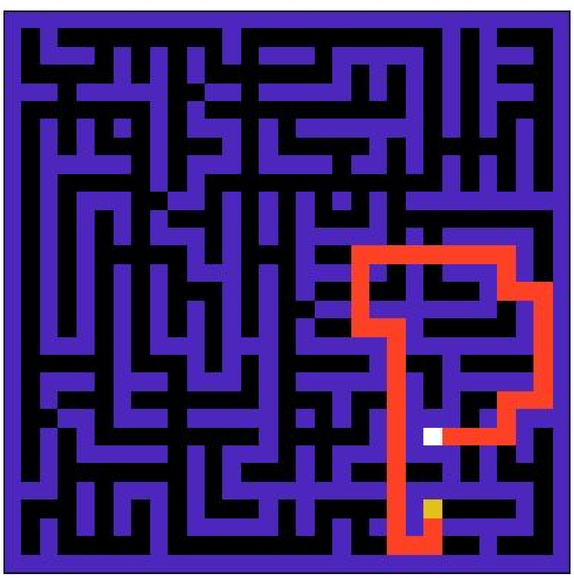
Shortest paths using vi:



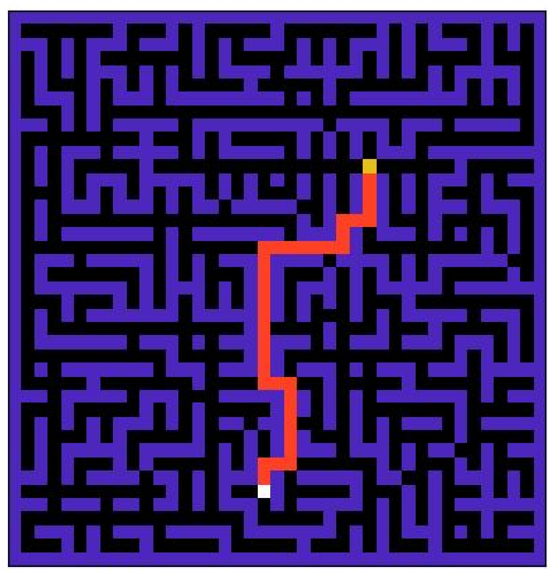
Grid 10



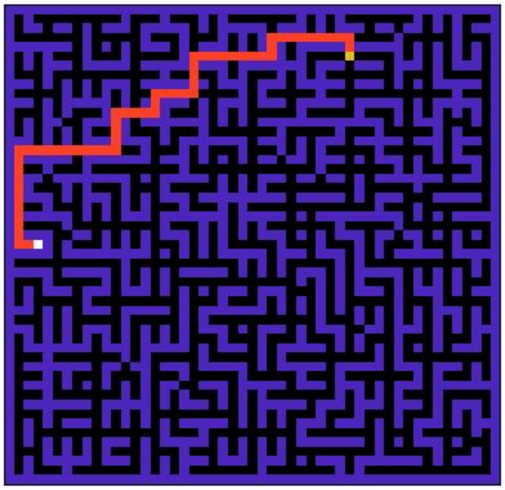
Grid 20



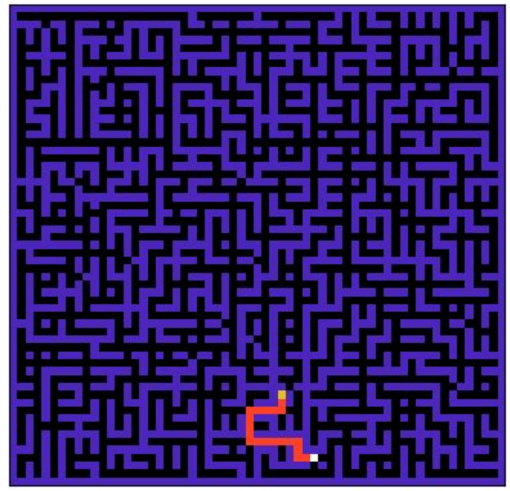
Grid 30



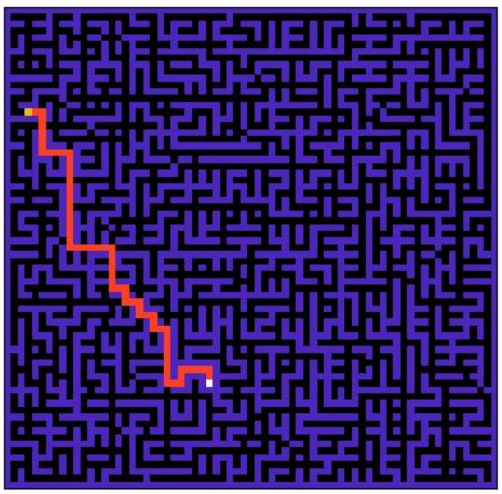
Grid 40



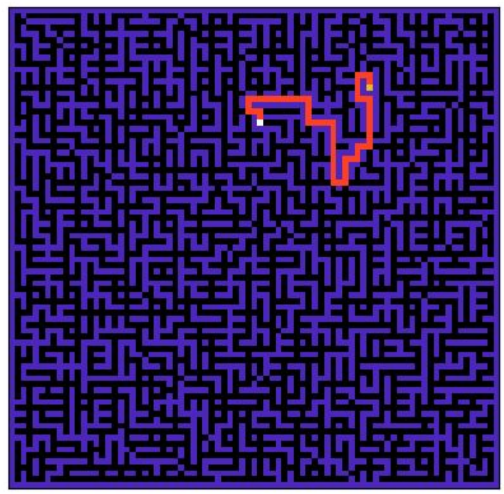
Grid 50



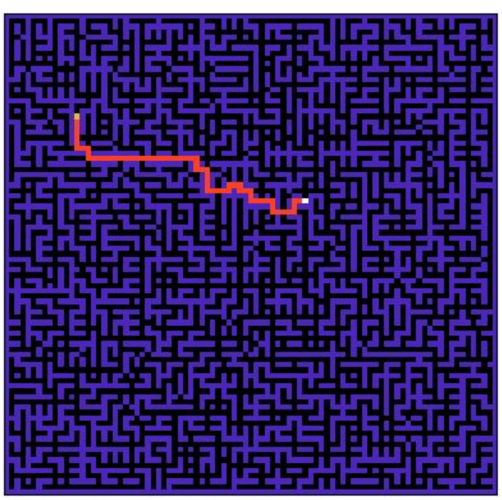
Grid 60



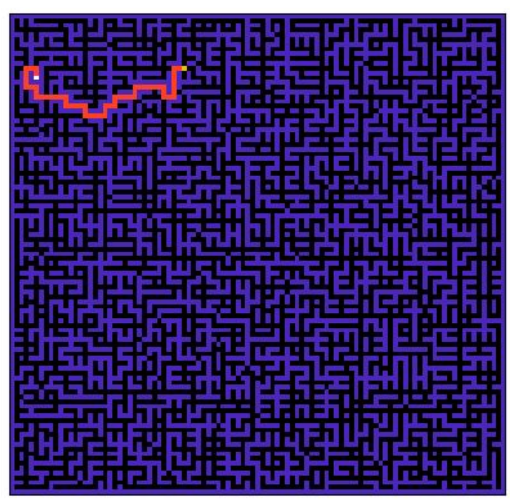
Grid 70



Grid 80



Grid 90



Grid 100

MDP design decisions for the maze problem:

1. Number of states(n) were chosen to be equal to the number of grid cells not having walls, i.e., the number of grid cells with values 0, 2 or 3.
2. Number of actions(k) were chosen to be 4, one for each of the 4 directions.
3. Gamma(γ) was chosen to be 1 to assure that the MDP indeed gives the shortest path always. It is possible since there is always a path from the start to the end so the MDP can be designed as an episodic MDP and gamma can be 1.
4. The transition matrix(T) for every state s and every action a is 1 since the probability of transition from s to s' (where s' is the state just above s) will be 1 if the action is north. So if an action is given, then the next state gets fixed with probability 1. Also a dummy state has been created. If the next state s' is a wall, then the transition probability to the dummy state will be 1 with huge -ve reward.
5. The reward matrix(R) is designed such that if the next state s' is a wall, then it gives huge negative reward of -1000000 and transitions into the dummy state. If the next state s' is the end state, then it gives huge positive reward of 1000000 and transitions into the dummy state. For all other transitions, the reward is -1.

Observations:

1. It was observed that value iteration was the fastest for solving the mazes for shortest paths followed by Howard's policy Iteration whereas Linear Programming was the slowest.
2. Value iteration and Howard's policy iteration gave different shortest paths for the mazes having multiple shortest paths.
3. For mdps with less than 100 states, linear programming takes the shortest time followed by value iteration whereas howard's policy iteration takes the longest time to solve the MDP.