# Reinforcement Learning on ZC Campus | Project

---


Muhammed Khalid 201901493

Samaa Khair 201901481

**University of Science and Technology at Zewail City**


**2022**


**Supervised by:**
**Dr. Doaa Shawky**

## I. Introduction

The project's primary goal is for an agent to start off at ZC's main gate. Its objective is to move things inside ZC in the fewest possible steps by picking them up and moving them from one place to another. The NB and HB are the pick-up sites, and the AB and One-stop-shop are the drop-off destinations.

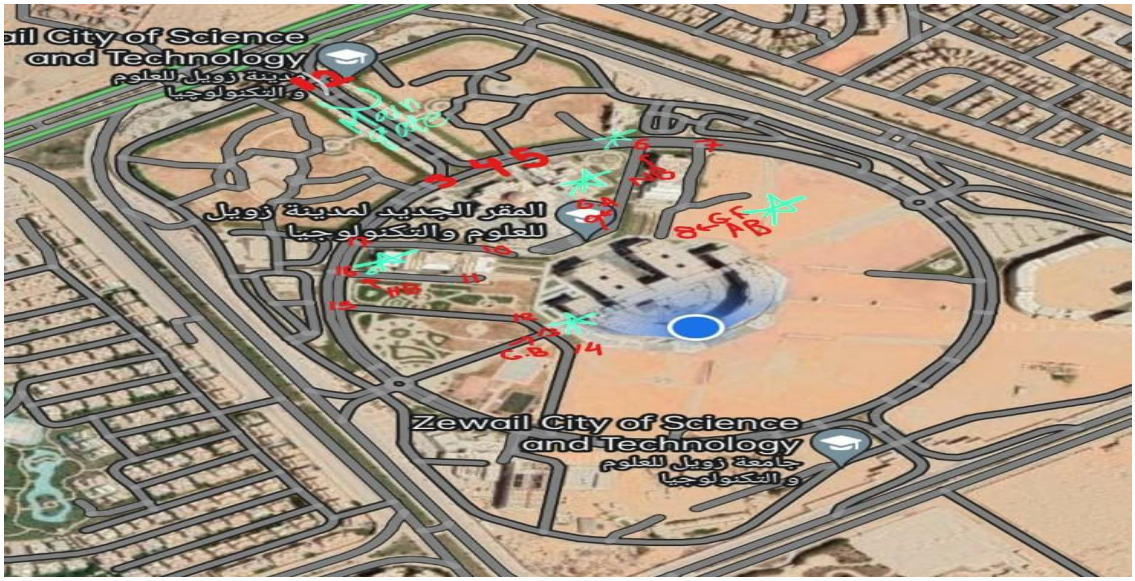
assuming that there are 4 things at each pickup location and a maximum of 4 items at each delivery location. There is a limitation that the agent can only carry one thing at a time. Your objective is to use Q-learning to create a route that allows the agent to send all things in the least amount of time from pick-up sites to delivery locations.

**A. Problem Statement:** Finding the optimal path to deliver items optimally on ZC Campus, navigating between buildings and stops.

**B. Problem Abstraction:**

- **Initial State:** It is defined to be at the main gate (either road 1 or 2), the agent is not carrying any items, and four items are at each pick-up location; HB and NB. Zero items are at both drop locations; OSS and AB.
- **Goal:** To deliver the four items at each pickup location successfully to both drop locations with the optimal path and logistics.
- **States:** Cross Roads. Where each crossroad leads to other crossroads and weather the agent is carrying anything or not and the current number of items in HB , NB , AB and OSS
- **Actions:** 8 degrees of freedom compass (Example: North, East, South, West, South north 1, south-north 2 , … ), picking up or delivering an item.
- **Solutions:** Sequence of actions; sequence of directions to move through along with decisions to pick-up or deliver items.
- **Path Cost:** Measured distances in Kilometers.

**Map with defined states is for reference:**



C. **Reward System**

Merely navigating through paths while not delivering or picking items results in a negative reward while delivering or picking up items results in a positive large reward.

D. **Rates and Factors**

As required in the description the learning rate is set to be 0.5 while the discount rate is set to be 0.3.

So the Q-value equation followed is:

**Q(state, action) = Q(state, action) + 0.5* (reward + 0.3* max(Q(next_state, action)) - Q(state, action))**

## II. Work Load

| Name | Work Distribution |
|---|---|
| Samaa Khair | **Problem Formulation, Actions Function, Random Q-Learning, Report writing** |
| Muhammed Khalid | **Problem Formulation, Apply Function, Greedy Q Learning, Exploring Q Learning** |

## III. Results and Analysis

A. **Random Q-Learning**

Results weren't efficient. The agent got stuck between two roads at some point.

B. **Greedy Q- Learning and Exploring Q-Learning**

Results were very efficient. For the Greedy Q-Learning, the agent managed to deliver all items in just 62 Actions after learning. For the Exploring Q-Learning, the Agent managed to deliver the items after 65 Actions after the learning process was done.