```
Data: enviroment // world
Result: $Q_{table}$ // Final table with all best actions
$\beta \leftarrow 0.001$ // exploartion decreasing decay for exponential decreasing
$exploration \leftarrow 1$ // initialize the exploration probability to 1
$\gamma \leftarrow 0.99$ // discounted factor
$\alpha \leftarrow 0.1$ // learning rate
$Q_{table} \leftarrow 0 \; \forall s, a$ // Initialize the Q-table to 0
// until max number of episodes, here is 0 to 1000
for each episode do
    $s \leftarrow$random state from environment
    // until max number of iteration per episode, here is 0 to 100
    for each iteration do
        // uniform distribution with limits:[0,1]
        if random number from uniform distribution $<$ exploration then
            $a \leftarrow$random action where $a \in \mathcal{A}(s)$
        else
            $a \leftarrow argmax_a(Q_{table}(s, a))$
        end
        $s', r \leftarrow$enviroment$\leftarrow a$
        $Q_{table}(s, a) = Q_{table}(s, a) + \alpha[r + \gamma \; Q_{table}(s', a) - Q_{table}(s, a)]$
        if $s'$ is terminal then
            break
        else
            $s \leftarrow s'$
        end
    end
    // here, 0.01 is the minimal exploration value
    $exploration \leftarrow max(0.01, e^{\beta * episode})$
end
```