## Exercise: Value Iteration (ca. 2h)

In the lecture you have seen a simple example of a Markov Decision Process Problem: a 4x3 grid world. We further discussed a simple algorithm that computes a utility for each state of this grid world - Value Iteration - which can be used to directly infer the optimal policy for a given grid world / Markov Decision Problem.

Write a simple simulator that reads in a grid world description from a text file. For each grid cell the text file shall specify whether we can move to the cell or not and which reward we will get if we move to this cell. Also specify "epsiode end" cells in the text file, i.e. in the example in the lecture these were the special fields "goal" (with reward +1) and "trap" (with reward -1) which terminated the current episode.

After reading in the description of the grid world, the simulator shall execute the value iteration algorithm till convergence. Let the user observe the utility for each grid cell after each Value Iteration update step.

Experiment with different rewards for the grid cells and develop an intuition how value iteration leads to utilities of the grid cells (states) such that we can use them for finding the optimal policy.

After value iteration has converged: compute and then visualize the optimal action for each state, i.e., visualize the optimal policy for the grid world.