# **Comp2208 Assignment: Search Methods**

Investigation of scalability with problem difficulty of depth first, breadth first, iterative deepening, and A\* heuristic search

Valeriu Andrei Cristea

ID: 28335988 / vac1u16

## **Contents**

- 1. Approach
- 2. Evidence of the 4 search methods in operation
- 3. Scalability Study
- 4. Extras and Limitations
- 5. Reference
- 6. Code
- 7. Output for 4X4 and 3 blocks

# 1. Approach

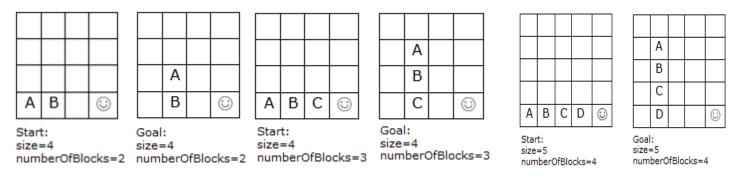
I chose Java for this assignment as the most intuitive approach involves object oriented programming and Java is the object-oriented language I have the most experience with.

I consider the starting state to be A in the bottom left corner and following letter are placed on the right of the previous letter until the desired number of blocks is reached. The agent starts from the bottom right corner. The number of blocks is always smaller than the size of the grid.

I consider the goal state to be all blocks on the second column, ordered alphabetically, if read top to bottom, with the last letter sitting on the bottom row; agent can be anywhere.

I consider a search to fail if it takes longer than 15 minutes or it runs out of memory.

## Examples:



## **Implementation:**

#### Description of classes and their methods:

#### 1. Block:

- Represents the blocks (A, B, etc.) and the agent
- Stores their position and their name (agent or A, B, etc.)

#### 2. Node:

- Represents a node for the tree search
- Stores the blocks in an ArrayList for scalability, the move used to get to it from the parent node, its parent node, the size of the grid and how many blocks there are

#### 3. Main:

- Contains the search methods and methods to write the outputs they produce to a file
- the main(String args[]) method acts as a test harness for the methods

#### Methods:

- Constructor for when the node is the root and when it is a child
- void makeMove(char move) , used in the child constructor, moves the agent if possible and swaps positions with a block if they occupy the same position
- boolean isSolution(), checks if the node is a solution, general implementation for scalability
- ArrayList<Node> getMoves() , returns the nodes created after taking each possible move

# 2. Evidence of the 4 search methods in operation

#### 2.1 BFS

Uses a Queue to store the nodes that are going to be checked. It starts by offering the starting node to the Queue. While the Queue is not empty and a solution has not been found it polls a node from the Queue and checks if it's a solution. If it is it prints the time and nodes expanded and stops the BFS, otherwise offers all the nodes created by using getMoves() on the current node to the Queue.

Could be improved by making it a graph search by adding a HashSet, adding all the nodes explored to the HashSet, and checking if a node already belongs to the HashSet before offering it to the Queue.

#### **2.2 DFS**

Similar implementation to the BFS except it uses a Stack instead of the Queue and shuffles the ArrayList of nodes returned by getMoves() before pushing the nodes from the ArrayList to the Stack.

Since it is not a graph search it does not check if a node has already been explored so the Stack is unnecessary as the next node it checks will always be the first element of the shuffled ArrayList.

Could be improved the same way as the BFS by making it a graph search, then the Stack would be needed as well.

## 2.3 Iterative deepening search

Algorithm adopted form pseudocode (En.wikipedia.org, 2017). It starts with a maximum depth of 0 and the starting Node. It calls the deepening function on the starting Node and it increases the maximum depth until a solution is found.

The deepening function checks if it is allowed to go deeper. If it is not allowed it checks if the solution is the current node and if it's not it returns null. If it is allowed to go deeper the function is recursively called on the nodes created by the getMoves() method and the allowed depth is decremented by one. If the deepening function called on these moves is not null it means that a solution has been found.

It is a modified version of DFS algorithm where the maximum depth increases until a solution is found.

#### 2.4 A\*

A\* algorithm is a modified version of Dijkstra's algorithm, also adapted for a tree in this situation, that uses heuristics. It considers nodes with a lower cost to be more promising. The cost of a node is sum of the Manhattan Distance from each block to its corresponding position in the solution and the depth of the node.

It functions the same way as BFS except it uses a PriorityQueue instead of a Queue. Could be improved by making it a graph search the same way as BFS.

**Additional note about all methods**: they could be drastically improved if getMoves() doesn't return the opposite of the last move made (A\* being able to solve with size 11 and 3 blocks under 10 seconds; it takes A\* over 15 minutes to solve for size 9 and 3 blocks without this), but the current implementation is the simplest. I expand on this in Extras 4.1.1.

## **2.5 Outputs** (U = up; D = down; L = left; R = right)

Original problem output:

DFS output not shown as it is 10339 moves long.

More output example (DFS solution not shown):

```
Size is 3X3
                                                  Size is 5X5
Number of Blocks is 3
                                                  Number of Blocks is 3
                                                 BFS failed because 15 minute exceeded
BFS time = 668 ms
Nodes expanded: 259297
                                                 DFS time = 339 ms
                                                 Nodes expanded: 230928
DFS time = 0 ms
Nodes expanded: 1664
                                                 IDS time = 4599 \text{ ms}
                                                 Nodes expanded: 47447008
IDS time = 69 ms
Nodes expanded: 398687
                                                 Solution path is : U L L L D L U R D R U U L D L
Solution path is : L L U R D R U U L D D R
                                                  A* time = 37 ms
A* time = 16 ms
                                                 Nodes expanded: 63625
Nodes expanded: 1750
Solution path is : L L U R D R U U L D D R
                                                  Solution path is : U L L L D L U R D R U U L D L
```

```
Size is 8X8

Number of Blocks is 3

BEST time = 15373 ms

Nodes expanded : 6081356

BEST time = 325845 ms

Nodes expanded : 3901577720

Solution path is : U L L L L L D L U R D R U U L D L

A* time = 6842 ms

Nodes expanded : 4187489

Solution path is : L U L L L L D L U R D R U U L D L
```

```
Size is 5X5
Number of Blocks is 4

DFS time = 14439 ms
Nodes expanded : 5092132
TDS failed because 15 minute exceeded
A* failed because 15 minute exceeded

Size is 6X6
Number of Blocks is 4

DFS time = 16698 ms
Nodes expanded : 8839779

Size is 7X7
Number of Blocks is 4

DFS failed because 15 minute exceeded
```

Full output of size 4 and 3 blocks at the end.

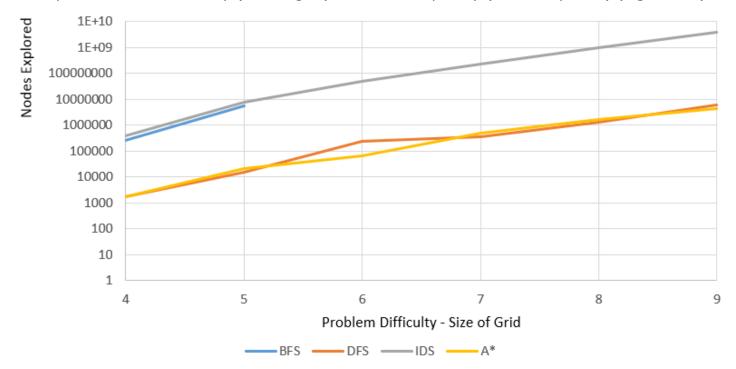
# 3. Scalability Study

## 3.1 Description

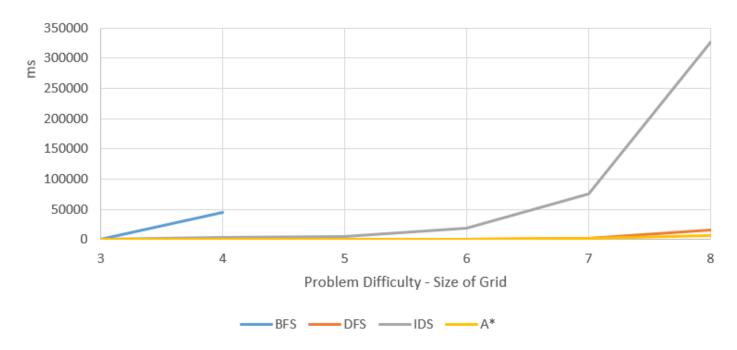
My implementation general and works for any size > 2 and numberOfBlocks < size. I do not have a test harness class as I test all my methods in Main and they write the output to a file which I save.

I control the problem difficulty by altering the size of the grid in the graphs below.

Graph 1: Problem Difficulty (size of grid) vs Time Complexity (nodes explored) (log10 scale)



Graph 2: Problem Difficulty (size of grid) vs Time Complexity (time taken; in ms)



From this data and these graphs, I can interpret and conclude the following:

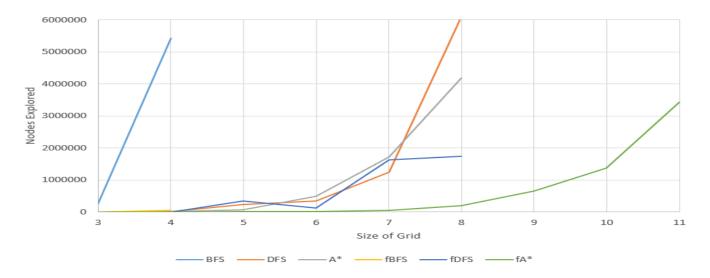
- BFS: It's the slowest and worst algorithm when the size is larger than 5, as it does not even finish the search in under 15 minutes. When the size is smaller than 5 it performs slightly better than IDS and worse than A\* and DFS. It always finds the optimal solution compared to DFS.
- DFS: It's a better algorithm than BFS and IDS on average but the downside is that it's completely random. It usually fails around size 7 9 but because of it's random nature it is possible to fail at sizes smaller than 7 or larger than 8. The solution it finds will never be the optimal compared to the other 3 algorithms.
- Iterative Deepening Search: It performs better than BFS and worse than A\* and DFS. It suffers from the same problem as BFS, namely it explores too many nodes. As the first graph shows, IDS and BFS explore as many nodes when the size is 4 as does DFS and A\* when the size is 7. The solution it finds is always optimal.
- A\*: Best algorithm in terms of nodes explored and time taken. Always gives the best solution and is more reliable compared to DFS.

## 4. Extras and Limitations

#### 4.1 Extras

- 4.1.1 getMoves() improvement: getFastMoves() returns the nodes created after taking each possible move without the node created by going back to parent state.
- 4.1.2 Improving the efficiency by making the tree searches (BFS, DFS and A\*) into graph searches by using HashSet and overriding the hashCode method in Block and Node and using getFastMoves() instead getMoves().(methods are called fbfs, fdfs, fastar-fastar has no HashSet)

**Note**: A\* is faster if I use getFastMoves() and no HashSet. With or without HashSet it explores the same number of nodes but substantially faster without – over 5 minutes for 7x7 with HashSet and under 1 minute for 11x11 without.



**Note on the graph**: I consider fDFS, fBFS and fA\* to fail if they take more than 5 minutes, the others still more than 15 minutes.

**Interpretation of the graph**: As expected A\* graph search is the optimal method. fBFS times out at size 5. fDFS has random results as it still shuffles the ArrayList. fDFS and A\* both fail at size 9.

#### 4.1.3 Adding immovable blocks:

- added getMovesObstacles() returns the ArrayList of nodes created after taking each possible move.
- added a new Node Constructor that getMovesObstacles() calls when it creates the nodes that it adds to the ArrayList
- methods are called ibfs, idfs, iastar and their implementation is the same as fbfs, fdfs, fastar except they call getMovesObstacles() instead of getFastMoves().

#### 4.2 Limitations and weakness of work

A better heuristic method to calculate the cost from the node to the solution could improve  $A^*$ .

Writing IDS to not be recursive and use a Stack could help with time and would prevent possible stack overflow error.

Using a faster programming language could result in better performance space and time wise but the implementation might be more difficult to achieve.

When gathering the data for the report I could have tested DFS more times and made an average for better results.

I assume that the methods only receive valid input. I could check for valid input when calling the methods for making possible the adaptation of the classes in different programs.

The program is heavily limited in speed by the CPU and heap memory. An implementation that uses less data structures and less object-oriented programming could achieve better space efficiency and possibly better time as well.

My implementation could also be improved by adding a TestHarness class that tests the algorithms for various sizes and logs the errors and outputs, eliminating the need of having the main method cluttered with comments.

## 5. Reference

(En.wikipedia.org, 2017). https://en.wikipedia.org/wiki/Iterative\_deepening\_depth-first\_search

# 6. Code

# **Block Class**

```
public class Block {
    //no getters and setters is a deliberate design choice
    public char name;
    public int x, y;

public Block(char name, int x, int y) {
        this.name = name;
        this.x = x;
        this.y = y;
    }

public int hashCode() {
        Character c = name;
        int sum = c.hashCode();
        sum = sum*17+ x;
        sum = sum*17+ y;
        return sum;
    }
}
```

# **Node Class**

```
ArrayList < Block > blocks;
```

```
public ArrayList<Node> getFastMoves() {
```

```
ArrayList<Node> moves = new ArrayList<Node>();
public ArrayList<Node> getMovesObstacle() {
```

# Main class

```
import java.io.BufferedWriter;
import java.io.FileWriter;
import java.util.*;
```

```
iterativeDeepening(s);
```

```
static void iterativeDeepening(Node start) throws Exception {
```

```
static void fbfs(Node start) throws Exception {
    startTime = System.currentTimeMillis();
```

```
static void printSol(Stack<Character> solution) throws Exception {
    fout.write("Solution path is : ");
    while (!solution.isEmpty())
        fout.write(solution.pop() + " ");
    fout.write('\n');
}

//puts the solution in a stack
//as it is written end to beginning
//and stack is lifo
static Stack<Character> foundSol(Node n) {

    Stack<Character> solution = new Stack<>();
    while (n.getParent() != null) {
        solution.add(n.getLastMove());
        n = n.getParent();
    }
    return solution;
}
```

# 7. Output for 4X4 and 3 blocks

IDS time = 3527 ms Nodes expanded : 7849127

Solution path is : U L L D L U R D R U U L D I

A\* time = 32 ms

Nodes expanded: 20421

Solution path is : U L L D L U R D R U U L D L