



CSC384

*Introduction to Artificial Intelligence*

SINAN LI

2024



---

# CONTENTS

## I Notes

5

### 1 | Chapter 1 Introduction to Artificial Intelligence

- 1.1 Artificial Intelligence (AI) 7
  - 1.1.1 Intelligence 7
  - 1.1.2 The Turing Test 7
  - 1.1.3 Computational Intelligence 8
- 1.2 Rationality 8
- 1.3 Subareas of AI 9
  - 1.3.1 Further Courses in AI 9
- 1.4 A Brief History of AI 9

### 2 | Chapter 2 Search

- 2.1 Search Problems 11
  - 2.1.1 Formalizing a Problem as a Search Problem 11
  - 2.1.2 Graphical Representation 13
  - 2.1.3 Algorithms for Search 14
- 2.2 Uninformed Search Algorithms 15
  - 2.2.1 Uninformed Search Strategies 15
  - 2.2.2 Breadth-First Search (BFS) 16
  - 2.2.3 Depth-First Search (DFS) 17
  - 2.2.4 Depth-Limited Search (DLS) 18
  - 2.2.5 Iterative Deepening Search (IDS) 19

### 3 | Chapter 3 Game Tree Search

### 4 | Chapter 4 Constraint Satisfaction Problems

**Chapter 5**  
Representing and Reasoning under Uncertainty

**6** **Chapter 6**  
Symbolic Knowledge Representation and Reasoning

**II Appendices 31**

**Bibliography 33**

# Part I

## Notes



# INTRODUCTION TO ARTIFICIAL INTELLIGENCE

# 1

## 1.1 Artificial Intelligence (AI)

Artificial Intelligence is a branch of computer science utilizing *computational* ideas and examining how we can achieve *intelligent* behavior through computation.

### 1.1.1 Intelligence

*The ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests)*

— Merriam-Webster Dictionary

Human features that are considered intelligent includes the ability to learn, understand, reason, plan, communicate, and perceive. For example, a human can learn to play a game, understand the rules of the game, reason about the best moves to make, plan a sequence of moves, communicate with other players, and perceive the state of the game.

### 1.1.2 The Turing Test

The Turing Test is a test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. The test was introduced by Alan Turing in his 1950 paper, *Computing Machinery and Intelligence*, while working at the University of Manchester [1]. Turing proposed that a human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses. The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another. The conversation would be limited to a text-only channel such as a computer keyboard and screen so that the result would not be dependent on the machine's ability to render words as speech. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test. The test does not check the ability to give the correct answer to questions; it checks how closely the answer resembles typical human answers. The conversation is limited to a single topic chosen by the examiner.

- Turing provided some very persuasive arguments that a system passing the Turing test is *intelligent*.
- We can only really say it *behaves* like a human

- **No guarantee** that it *thinks* like a human
- The Turing test does not provide much traction on the question of **how to build** an intelligent system.

### Why not just simulate human's brain?

- Brains are very good at making rational decisions, but **not perfect**.
- Brains **aren't as modular** as software, so hard to reverse engineer!
- Computers and Humans have quite **different abilities**.
  - *Memory* and *simulation* are key to decision making.
  - *Perceptual tasks* (vision, sound, etc.) are effectively accomplished by architectures related to the way the brain works (deep neural networks).

## 1.1.3 Computational Intelligence

Artificial Intelligence tries to understand and model *intelligence* as a **computational process**. Thus we try to construct systems whose **computation** achieves or approximates the desired notion of intelligence.

Other areas interested in the study of intelligence lie in other areas or study, e.g., cognitive science which focuses on human intelligence. Such areas are very related, but their central focus tends to be different.

## 1.2 Rationality

Formally, we can define an **agent** as something that **perceives** and **acts** in an *environment*. An agent can be a *human*, a *robot*, or a *software agent*.

A **rational agent** is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome. Rationality is distinct from omniscience (knowing everything) and omnipotence (being able to do anything). Rationality maximizes **expected utility**, which is the sum of the utility of each possible outcome of an action weighted by its probability of occurring.

Rationality is measured by the *outcome*, not the *action* itself. It is a precise *mathematical* notion of what it means to do *the right thing* in any particular circumstance. Provides

- A **precise mechanism** for analyzing and understanding the properties of this ideal behaviour we are trying to achieve.
- A **precise benchmark** against which we can measure the behaviour the systems we build.

### Trying/Expectation

Rational action is not always equal to rational decision.

- 1 We often don't have **full control** or **knowledge** of the world we are interacting with.
- 2 We usually don't know **precisely** what the **effects** of our actions will be.

In some contexts we can *simplify* the computational task by assuming that we do have full knowledge/-control.



A common misconception is to equate AI with Machine Learning. But AI is much more than that. This course will not focus on Machine Learning, but rather on the other subareas of AI. What we cover is not an exhaustive list of all subareas of AI, but rather a starting point for further exploration.

- **Perception:** vision, speech understanding, etc.
- **Machine Learning, Neural Networks**
- **Robotics**
- **Natural Language Processing**
- **Reasoning and Decision Making**
  - **Symbolic Knowledge Representation**
  - **Reasoning** (logical, probabilistic)
  - **Decision Making** (search, planning, decision theory)

### 1.3.1 Further Courses in AI

- Perception: vision, speech understanding, etc.
  - CSC487H1 “Computational Vision”
  - CSC420H1 “Introduction to Image Understanding”
- Machine Learning, Neural networks
  - CSC311H “Introduction to Machine Learning”
  - CSC412H “Probabilistic Learning and Reasoning”
  - CSC413H1 “Neural Networks and Deep Learning”
- Robotics
  - Engineering courses
- Natural language processing
  - CSC401H1 “Natural Language Computing”
  - CSC485H1 “Computational Linguistics”
- Reasoning and decision making
  - CSC486H1 “Knowledge Representation and Reasoning”

- 1940-1950: Early days
  - 1943: McCulloch & Pitts: Boolean circuit model of brain
  - 1950: Turing’s “Computing Machinery and Intelligence”
- 1950—70: Excitement: Look, Ma, no hands!
  - 1950s: Early AI programs, including Samuel’s checkers program, Newell & Simon’s Logic Theorist, Gelernter’s Geometry Engine

- 1956: Dartmouth meeting: “Artificial Intelligence” adopted
- 1965: Robinson’s complete algorithm for logical reasoning
- 1970—Early 2000: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms
  - 1988—93: Expert systems industry busts: “AI Winter”
  - 1997: IBM’s Deep Blue beats chess grandmaster Garry Kasparov
- Early 2000— present: Statistical approaches
  - Resurgence of probability, focus on uncertainty
  - Agents and learning systems. . . “AI Spring”
  - 2007: DARPA Urban Challenge – CMU autonomous vehicle drives 55 miles in an urban environment while adhering to traffic hazards and traffic laws.
  - 2016: AlphaGo beats 9-Dan pro Go player Lee Sedol
  - 2017: AlphaGo Zero – learns by playing with itself
  - 2022: Large Language Models (LLM) Chat Generative Pre-trained Transformer, which has been fueled by advances in Neural Net architecture and access to big data.

There are many **unsolved** problems yet. . . including lots of **legal/ethical** ones.

# SEARCH

## 2.1 Search Problems

Searching is one of the most *fundamental techniques* in AI, underlying sub-module in many AI systems. Ai can *solve* many problems that humans are not good at, and achieve *super-human performance* in many domains (e.g. chess, go, etc.).

- **Benefits**
  - Useful as a general algorithmic technique for solving problems, both in AI and in other areas.
  - Outperform humans in some areas (e.g. games).
  - Practical:
    - Many problems don't have specific algorithms for solving them.
    - Useful in approximation (e.g., local search in optimization problems).
  - Some critical aspects of intelligent behaviour, e.g., planning, can be cast as search.
- **Limitations**
  - Only shows how to solve the problem once we have it correctly formulated.

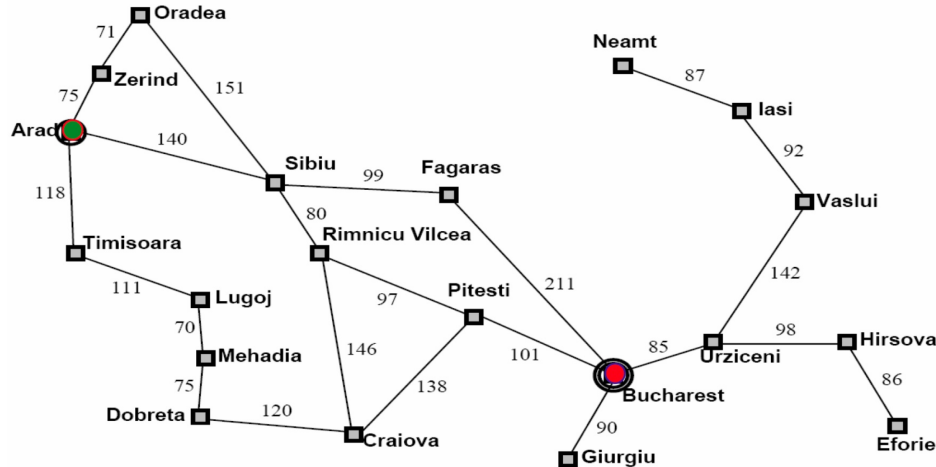
### 2.1.1 Formalizing a Problem as a Search Problem

- Necessary components
  - 1 **State Space:** A **state** is a representation of a *configuration* of the problem domain. The state space is the *set of all states* include in our model of the problem.
  - 2 **Initial State:** The starting configuration.
  - 3 **Goal State:** The configuration one wants to achieve.
  - 4 **Actions** (or State Space Transitions): Allowed changes to move from one state to another.
- Optional Ingredients
  - *Costs:* Representing the cost of moving from state to state.

- **Heuristics:** Help guide the search process.

Once a search problem is formalized, there are a number of algorithms one can use to solve it. A **solution** is a **sequence of actions** or moves that can transform the **initial state** into a **goal state**.

**Example (Romania Travel).** We want to travel in Romania from Arad to Bucharest as fast as possible.



Each state would be a city.

- **State Space:** The set of all cities on the map.
- **Initial State:** Arad.
- **Goal State:** Bucharest.
- **Actions:** Driving between neighbouring cities.



**Example (Water Jugs).** We have a 3-liter jug and a 4-liter jug. We can fill either jug to the top from a tap, or we can empty either jug onto the ground. We can also pour the contents of one jug into the other until the receiving jug is full or the pouring jug is empty.

Suppose initially the 4-liter jug is full, we want to have exactly 2 liters in the 3-liter jug.

We can use a pair of numbers to represent the state of the system: the amount of water in the 3-liter jug and the amount of water in the 4-liter jug.

- **State Space:** The set of all pairs of numbers  $(a, b)$  where  $a$  is the amount of water in the 3-liter jug and  $b$  is the amount of water in the 4-liter jug.
- **Initial State:**  $(0, 4)$ .
- **Goal State:**  $(2, 0)$ ,  $(2, 1)$ ,  $(2, 2)$ ,  $(2, 3)$ ,  $(2, 4)$ .
- **Actions:**
  - Fill the 3-liter jug from the tap.
  - Fill the 4-liter jug from the tap.
  - Empty the 3-liter jug onto the ground.
  - Empty the 4-liter jug onto the ground.
  - Pour the contents of the 3-liter jug into the 4-liter jug.
  - Pour the contents of the 4-liter jug into the 3-liter jug.

## Remark

When formalizing a search problem, always consider these questions:

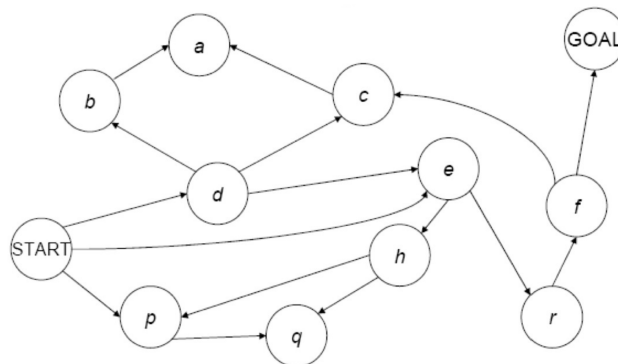
- ❶ Can we reach all states from any given start state?
- ❷ Will all actions result in a change of state?  
**No!** Imagine you have  $(3, 4)$  and you try to fill the 3-liter jug from the tap. You will still have  $(3, 4)$ .



In more complex situations,

- Actions may lead to **multiple states**.  
For example, flipping a coin may lead to heads or tails.
- We may not be **sure of a given state**  
For example, when prize is behind door 1, 2, or 3.
- Such situations require techniques for reasoning under uncertainty: assign probabilities to given outcomes.

## 2.1.2 Graphical Representation



Assuming a finite search space, the

- **vertices** represent states in the search space; and the
- **edges** represent transitions resulting from actions (or successor functions).

## Search Tree

### Definition 2.1.1 Search Tree

A **search tree** is a *directed graph* where

- Each node represents a state.
- Each edge represents an action.
- The root node represents the initial state.
- The leaf nodes represent goal states.

A search tree reflects the behaviour of an algorithm as it walks through a search problem. It has two important properties:

- **Solution depth**, denoted  $d$ , the depth of the shallowest goal node in the tree.
- **Maximum branching factor**, denoted  $b$ , the maximum number of children of any node in the tree.

### Remark

Note that the **same** state may appear **multiple times** in a search tree.

### Remark

It is important to distinguish between **states** from **nodes**.

- A **state** represents a possible configuration of the world.
- A **node** is a data structure constituting part of a search tree. It includes
  - a **state** and
  - the **parent node**,
  - the **action** that led to this node, and
  - the **cost** of the path from the initial node to this node.
- Intuitively speaking, each node corresponds with a path from the initial state to the node's state.
- Two **different nodes** are allowed to contain the **same world state**.

## 2.1.3 Algorithms for Search

- **Input**
  - **Initial node**
  - **Successor Function**  $S(x)$   
returns the set of nodes that can be reached from node  $c$  via a single action.
  - **Goal Test Function**  $G(x)$   
returns true if node  $c$  satisfies the goal condition.
  - **Action Cost Function**  $C(x, a, y)$   
returns the cost of moving from node  $x$  to node  $y$  using action  $a$ .  
Note that  $C(x, a, y) = \infty$  if  $y$  is not reachable from  $x$  via  $a$ .
- **Output**
  - A **sequence of actions** that transforms the initial node satisfying the goal test.
  - The sequence might be, **optimal in cost** for some algorithms, **optimal in length** for some algorithms, come with **no optimality** guarantees from other algorithms.
- **Procedure**
  - Put nodes have not yet expanded in a list called the **Frontier** (or **Open**).
  - Initially, only the **initial node** is in the **Frontier**.
  - At each iteration, pull a node from the **Frontier**, apply  $S(x)$ , and insert the children back into the **Frontier**.
  - Repeat until pulling a goal node.

Note that the search terminates only when a goal node is expanded into the **Frontier**.

---

**Algorithm 1** Tree Search Algorithm

---

```
1: function TREE-SEARCH(Frontier, Successors, Goal?)
2:   if Frontier is empty then
3:     return failure
4:   Curr  $\leftarrow$  select state from frontier
5:   if Goal?(Curr) then
6:     return Curr
7:   Frontier'  $\leftarrow$  (Frontier - { Curr })  $\cup$  Successors(Curr)
8:   return TREESEARCH(Frontier', Successors, Goal?)
```

---

### Critical Properties of Search

- **Completeness**

- A search algorithm is **complete** if it always finds a solution if one exists.

- **Optimality**

- A search algorithm is **optimal** if it always finds a solution with the lowest cost.

- **Time Complexity**

- The **time complexity** of a search algorithm is the number of nodes it expands or generates.

- **Space Complexity**

- The **space complexity** of a search algorithm is the maximum number of nodes it stores in memory.

## 2.2 Uninformed Search Algorithms

---

### 2.2.1 Uninformed Search Strategies

The **order** of nodes in the **Frontier** determines the behaviour of the search algorithm. It determines whether or not the algorithm is complete, optimal, and how much time and space it requires. Search algorithms differ in their selection rule, but they all keep the **Frontier** as an ordered set. The question of which node to select next from the **Frontier** is now equivalent to the question of how to order the **Frontier**.

There are strategies that adopt a *fixed-rule* for selecting the next node to be expanded. These strategies are called **uninformed search strategies** because they do **not** use any *domain specific information* about the problem other than the definition of the problem itself. These strategies include Breadth-First Search, Depth-First Search, Uniform-Cost Search, Depth-Limited Search, and Iterative Deepening Search.

In those cases, it is assumed that the graph we are searching is *explicitly* represented as an adjacency list (or adjacency matrix). However, this will **not** work when there are an exponential number of nodes and edges. In AI applications, there are typically an exponential number of nodes, making it impossible to explicitly represent them all. AI search algorithms work with *implicitly* defined state spaces, where actions are compacted as *successor state functions*, and nodes must contain enough information to allow the successor state function to be applied.

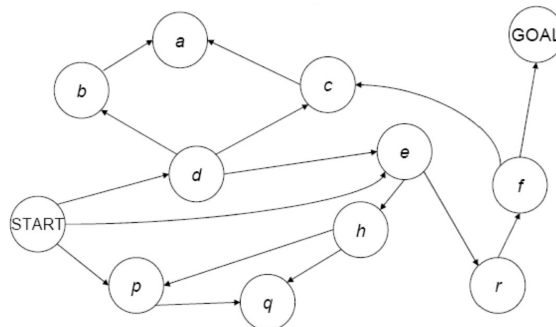
## 2.2.2 Breadth-First Search (BFS)

### Definition 2.2.1 Breadth-First Search

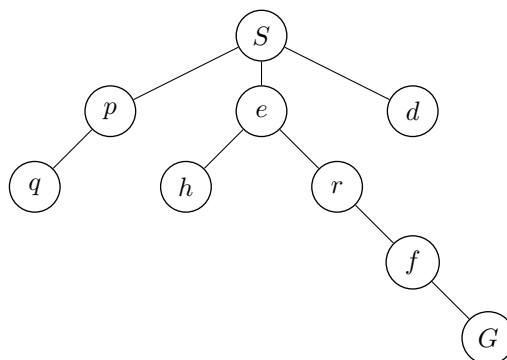
**Breadth-First Search** is a search strategy that expands the shallowest unexpanded node first.

BFS explores the search tree *level by level*. It place the children of the current node at the **end** of the Frontier. This means the Frontier is a *queue*, and we always extract the *first* element from the Frontier.

**Example.** We run BFS on the following graph,



which yields the search tree



- **Completeness**

- BFS is **complete** if the **branching factor** is *finite*.
- The length of the path (from the initial node to the node removed from the Frontier) is *non-decreasing*, as we replace each expanded node with path length  $k$  with a node with path length of  $k + 1$ .

- **Optimality**

- BFS gives the **shortest length solution**.  
All nodes with shorter paths are expanded prior to any node with longer path. We examine all paths of length  $< k$  before all paths of length  $k$ . Thus, if there is a solution with length  $k$ , we will find it before longer solutions.
- BFS is **not optimal** in terms of *cost*.  
BFS does not take into account the cost of the path. It is possible that a longer path has a lower cost than a shorter path.



## Properties of BFS

Let  $b$  be the maximum branching factor and  $d$  be the depth of the shallowest goal node.

- **Time Complexity**

$$1 + b + b^2 + \dots + b^d + b(b^d - 1) \in O(b^{d+1})$$

The last term is  $b(b^d - 1)$  because we need to remove the goal node, which does not need any successors.

- **Space Complexity**

$$O(b^{d+1})$$

In the worst case, only the last node of depth  $d$  satisfies the goal. So all nodes at depth  $d$  except the last one will be expanded by the search and each such expansion will add up to  $b$  new nodes to the **Frontier**.

Depth	Paths	Time	Memory
1	1	0.01 millisec	100 bytes
6	$10^6$	10 sec.	100 MB
8	$10^8$	17 min.	10 GB
9	$10^9$	3 hrs.	100 GB

Space complexity is a bigger problem than time complexity. Typically, BFS runs out of memory before it runs out of time.

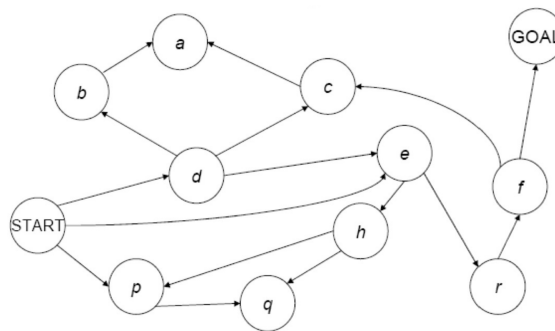
## 2.2.3 Depth-First Search (DFS)

### Definition 2.2.2 Depth-First Search

**Depth-First Search** is a search strategy that expands the deepest unexpanded node first.

DFS explores the search tree *branch by branch*. It places the children of the current node at the **front** of the **Frontier**. This means the **Frontier** is a **stack**, and we always extract the *first* element from the **Frontier**.

**Example.** We run DFS on the following graph,



The **Frontier** changes as follows:

Frontier	Node Removed	Successors Added
$S$	$S$	$d, e, p$
$d, e, p$	$d$	$b, c, e$
$b, c, e, e, p$	$b$	$a$
$a, c, e, e, p$	$a$	
$c, e, e, p$	$c$	$a$
$a, e, e, p$	$a$	
$e, e, p$	$e$	$h, r$

...



- **Completeness**

- DFS is **not complete**. **Infinite** paths cause incompleteness.
- This is because DFS does not keep track of visited nodes, so it can get stuck in a loop. An infinite search space can also cause DFS to run forever.
- We can prune paths with **cycles** to get completeness, if state space is finite.

- **Optimality**

- DFS is **not optimal** in terms of *length* or *cost*.
- DFS returns the *first* solution it finds, which may not be the *shortest* or *cheapest* solution.

## Properties of DFS

Let  $b$  be the maximum branching factor and  $m$  be the maximum depth of the search tree.

- **Time Complexity**

$$\mathcal{O}(b^m)$$

where  $m$  is the length of the *longest path* in the state space.

- This is **very bad** if  $m$  is much larger than  $d$ , but if there are *many solution paths*, DFS may be much faster than BFS.
- Using **Heuristics**, we can determine which successor to explore first.

- **Space Complexity**

$$\mathcal{O}(bm)$$

DFS has a **linear** space complexity, which is a significant advantage.

This is because DFS only explores a single branch of the search tree at a time. **Frontier** only contains the current node  $v$  along with the  $m$  descendants of  $v$ . Each node can have at most  $b$  unexplored siblings and there are at most  $m$  nodes on the current branch, which gives  $\mathcal{O}(bm)$ .

## 2.2.4 Depth-Limited Search (DLS)

Despite DFS being *incomplete* and *non-optimal*, it has a great advantage over BFS: *linear space complexity*. This makes it possible to use DFS in situations where BFS would run out of memory.

In order to address the incompleteness of DFS, we can impose a *depth limit*  $D$  on the search. This means that we will only explore paths of length  $D$  or less. If we reach a node at depth  $D$  that is not a goal node, we will *backtrack* and explore other paths. This is called **Depth-Limited Search** (DLS).

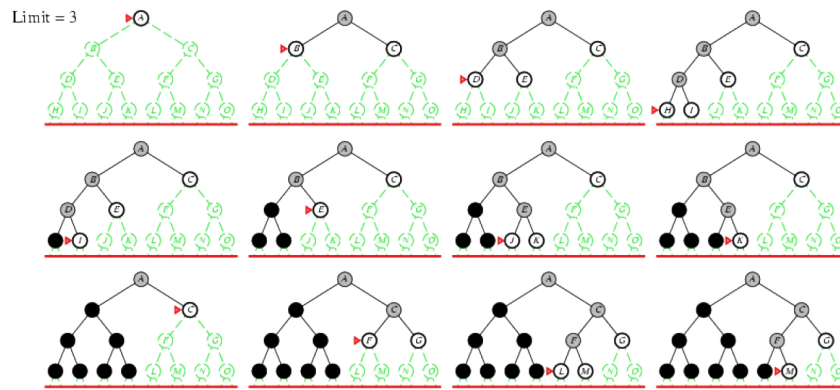
### Definition 2.2.3 Depth-Limited Search

**Depth-Limited Search** is a search strategy that expands the deepest unexpanded node first, but only up to a given depth limit  $D$ .

- **Benefit:** DLS overcomes the infinite length paths issue,
- **Limitation:** DLS only finds a solution if the *depth limit*  $D$  is greater than or equal to the depth of the shallowest goal node  $d$ .

#### Remark

The root of a search tree is at depth 0. A path consisting only of the root node has length 0. Recall that the length of a path is the number of actions (edges) in the path.



#### Algorithm 2 Depth-Limited Search Algorithm

```

1: function DLS(start, frontier, successors, goal?, maxd)
2:   frontier.insert(<start>)                                ▷ frontier must be a stack for DFS
3:   cutoff ← false
4:   while frontier is not empty do
5:     curr ← frontier.extract()                              ▷ Remove node from frontier
6:     if goal?(curr.final()) then
7:       return curr, cutoff                                  ▷ curr is solution
8:     else if LENGTH(curr) < maxd then                      ▷ Only successors if length(curr) < maxd
9:       for succ ∈ successors(curr.final()) do
10:        frontier.insert(<p, succ>)
11:     else
12:       cutoff ← true                                       ▷ Reached depth limit, some nodes not expanded
13:   return failure, cutoff

```

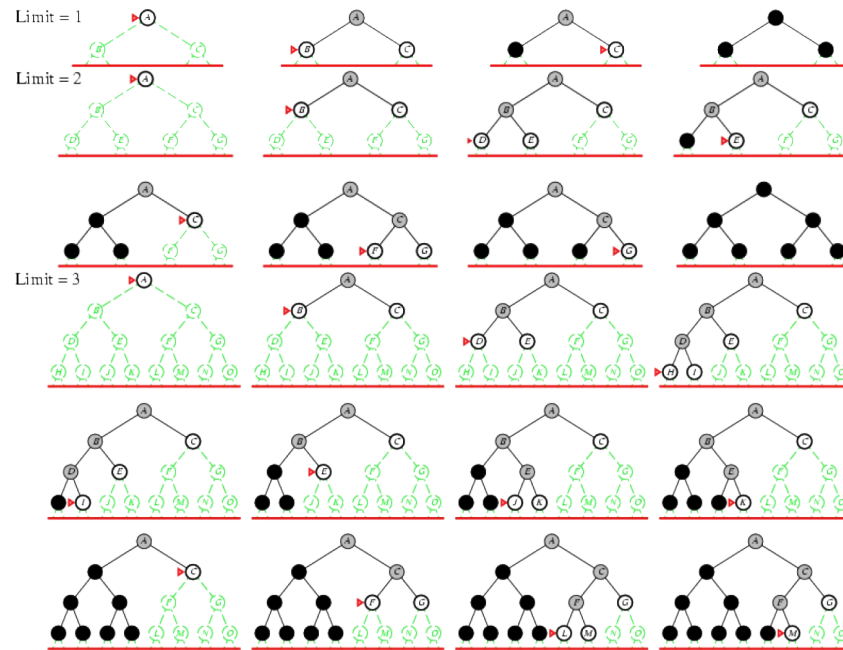
### 2.2.5 Iterative Deepening Search (IDS)

DLS is *complete* if the *depth limit*  $D$  is greater than or equal to the depth of the shallowest goal node  $d$ . However, we do not know  $d$  in advance. We can start at a depth limit  $D = 0$ , and iteratively **increase**  $D$  until we find a solution. We stop if **a solution is found**, or the search **failed without cutting off** any nodes because of the depth limit. This is called **Iterative Deepening Search** (IDS).

Note that if no nodes were cut off, the search examined all nodes in the state space and found no solution, hence there is no solution.

### Definition 2.2.4 Iterative Deepening Search

**Iterative Deepening Search** is a search strategy that repeatedly applies DLS with increasing depth limits until a solution is found.



### Algorithm 3 Iterative Deepening Search Algorithm

```

function IDS(start, frontier, successors, goal?)
    depth  $\leftarrow$  0
    while true do
        solution, cutoff  $\leftarrow$  DLS(start, successors, goal?, depth)
        if solution is not failure then
            return solution
        else if cutoff is false then                                 $\triangleright$  No nodes at deeper levels exist, no solution
            return failure
        depth  $\leftarrow$  depth + 1
    
```

- **Completeness**

- IDS is **complete**.

- **Optimality**

- IDS gives the **shortest length solution**.
- IDS is **not optimal** in terms of **cost**, for the same reason as DFS.

We can make IDS optimal by doing the following:

- Can use a **cost bound** instead of depth bound.
- Only expand nodes of cost **less** than the cost bound.
- Keep track of the **minimum cost unexpanded node** in each iteration, **increase the cost bound** to that on the next iteration.
- Can be **computationally expensive** since need as many iterations of the search as there are distinct node costs.

## Properties of IDS

- **Time Complexity**

$$(d+1)b^0 + db^1 + (d-1)b^2 + \dots + b^d \in \mathcal{O}(b^d)$$

- This is **very bad** if  $d$  is much larger than  $m$ , but if there are *many solution paths*, IDS may be much faster than BFS.
- Using **Heuristics**, we can determine which successor to explore first.

- **Space Complexity**

$$\mathcal{O}(bd)$$

The space complexity of IDS is still linear.



---

# GAME TREE SEARCH

# 3





---

# CONSTRAINT SATISFACTION PROBLEMS

# 4



---

REPRESENTING AND  
REASONING UNDER  
UNCERTAINTY

5



---

# SYMBOLIC KNOWLEDGE REPRESENTATION AND REASONING

# 6



# Part II

## Appendices





---

## BIBLIOGRAPHY

- [TUR50] A. M. TURING. “I.—COMPUTING MACHINERY AND INTELLIGENCE”. In: *Mind* LIX.236 (Oct. 1950), pages 433–460. ISSN: 0026-4423. DOI: [10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433). eprint: <https://academic.oup.com/mind/article-pdf/LIX/236/433/30123314/lix-236-433.pdf>. URL: <https://doi.org/10.1093/mind/LIX.236.433> (cited on page 7).

