

BLM267

Chapter 2: Introduction to Data Structures and Algorithms

1

Data Structures Using C, Second Edition
Reema Thareja

- **BASIC TERMINOLOGY**
- **CLASSIFICATION OF DATA STRUCTURES**
- **OPERATIONS ON DATA STRUCTURES**
- **ABSTRACT DATA TYPE**
- **ALGORITHMS**
- **DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM**
- **CONTROL STRUCTURES USED IN ALGORITHMS**
- **TIME AND SPACE COMPLEXITY**

BASIC TERMINOLOGY

- We know how to write, debug, and run simple programs in C language.
- Our aim has been to design good programs, where a good program is defined as a program that
 - runs correctly
 - is easy to read and understand
 - is easy to debug
 - is easy to modify

BASIC TERMINOLOGY

- A program should give correct results, but along with that it should also run efficiently.
- A program is said to be efficient when it executes in minimum time and with minimum memory space.
- In order to write efficient programs we need to apply certain data management concepts.
- Data structure is a crucial part of data management.
- A data structure is basically a group of data elements that are put together under one name, and which defines a particular way of storing and organizing data in a computer so that it can be used efficiently.

BASIC TERMINOLOGY

- Data structures are used in almost every program or software system.
- Some common examples of data structures are arrays, linked lists, queues, stacks, binary trees, and hash tables.
- Representing information is fundamental to computer science.
- The primary goal of a program or software is not to perform calculations or operations but to store and retrieve information as fast as possible.

BASIC TERMINOLOGY

- The application of an appropriate data structure provides the most efficient solution.
- A solution is said to be efficient if it solves the problem within the required resource constraints like the total space available to store the data and the time allowed to perform each subtask.
- And the best solution is the one that requires fewer resources than known alternatives.
- Moreover, the cost of a solution is the amount of resources it consumes.
- The cost of a solution is basically measured in terms of one key resource such as time, with the implied assumption that the solution meets the other resource constraints.

- Today computer programmers do not write programs just to solve a problem but to write an efficient program.
- For this, they first analyze the problem to determine the performance goals that must be achieved and then think of the most appropriate data structure for that job.
- However, program designers with a poor understanding of data structure concepts ignore this analysis step and apply a data structure with which they can work comfortably.
- The applied data structure may not be appropriate for the problem at hand and therefore may result in poor performance (like slow speed of operations).

BASIC TERMINOLOGY

- If a program meets its performance goals with a data structure that is simple to use, then it makes no sense to apply another complex data structure just to exhibit the programmer's skill.
- When selecting a data structure to solve a problem, the following steps must be performed.
 - Analysis of the problem to determine the basic operations that must be supported. For example, basic operation may include inserting/deleting/searching a data item from the data structure.
 - Quantify the resource constraints for each operation.
 - Select the data structure that best meets these requirements.
- This three-step approach to select an appropriate data structure for the problem at hand supports a data-centred view of the design process.

BASIC TERMINOLOGY

- In the approach, the first concern is the data and the operations that are to be performed on them.
- The second concern is the representation of the data, and the final concern is the implementation of that representation.
- There are different types of data structures that the C language supports.
- While one type of data structure may permit adding of new data items only at the beginning, the other may allow it to be added at any position.
- While one data structure may allow accessing data items sequentially, the other may allow random access of data.
- So, selection of an appropriate data structure for the problem is a crucial decision and may have a major impact on the performance of the program.

Elementary Data Structure Organization

- Data structures are building blocks of a program.
- A program built using improper data structures may not work as expected.
- So as a programmer it is mandatory to choose most appropriate data structures for a program.
- The term data means a value or set of values.
- It specifies either the value of a variable or a constant (e.g., marks of students, name of an employee, address of a customer, value of pi, etc.).

Elementary Data Structure Organization

- While a data item that does not have subordinate data items is categorized as an elementary item, the one that is composed of one or more subordinate data items is called a group item.
- For example, a student's name may be divided into three sub-items—first name, middle name, and last name—but his roll number would normally be treated as a single item.
- A record is a collection of data items. For example, the name, address, course, and marks obtained are individual data items.
- But all these data items can be grouped together to form a record.

Elementary Data Structure Organization

- A file is a collection of related records.
- For example, if there are 60 students in a class, then there are 60 records of the students.
- All these related records are stored in a file.
- Similarly, we can have a file of all the employees working in an organization, a file of all the customers of a company, a file of all the suppliers, so on and so forth.
- Moreover, each record in a file may consist of multiple data items but the value of a certain data item uniquely identifies the record in the file. Such a data item K is called a primary key, and the values K1, K2 ... in such field are called keys or key values.

Elementary Data Structure Organization

- For example, in a student's record that contains roll number, name, address, course, and marks obtained, the field roll number is a primary key.
- Rest of the fields (name, address, course, and marks) cannot serve as primary keys, since two or more students may have the same name, or may have the same address (as they might be staying at the same place), or may be enrolled in the same course, or have obtained same marks.
- This organization and hierarchy of data is taken further to form more complex types of data structures, which is discussed in Section 2.2.

CLASSIFICATION OF DATA STRUCTURES

- Data structures are generally categorized into two classes: primitive and non-primitive data structures.
- Primitive data structures are the fundamental data types which are supported by a programming language.
- Some basic data types are integer, real, character, and boolean.
- The terms 'data type', 'basic data type', and 'primitive data type' are often used interchangeably.
- Non-primitive data structures are those data structures which are created using primitive data structures.
- Examples of such data structures include linked lists, stacks, trees, and graphs.

CLASSIFICATION OF DATA STRUCTURES

Linear and Non-linear Structures

- Non-primitive data structures can further be classified into two categories: linear and non-linear data structures.
- If the elements of a data structure are stored in a linear or sequential order, then it is a linear data structure.
- Examples include arrays, linked lists, stacks, and queues.
- Linear data structures can be represented in memory in two different ways.
- One way is to have to a linear relationship between elements by means of sequential memory locations.
- The other way is to have a linear relationship between elements by means of links.

CLASSIFICATION OF DATA STRUCTURES

Linear and Non-linear Structures

- However, if the elements of a data structure are not stored in a sequential order, then it is a non-linear data structure.
- The relationship of adjacency is not maintained between elements of a non-linear data structure.
- Examples include trees and graphs. C supports a variety of data structures.
- We will now introduce all these data structures and they would be discussed in detail in subsequent chapters.

Arrays

- An array is a collection of similar data elements.
- These data elements have the same data type.
- The elements of the array are stored in consecutive memory locations and are referenced by an index (also known as the subscript).
- In C, arrays are declared using the following syntax: `type name(size);`
- For example, `int marks(10);` The above statement declares an array marks that contains 10 elements.
- In C, the array index starts from zero.
- This means that the array marks will contain 10 elements in all.
- The first element will be stored in `marks(0)`, second element in `marks(1)`, so on and so forth.
- Therefore, the last element, that is the 10th element, will be stored in `marks(9)`. In the memory, the array will be stored as shown in Fig. 2.1.

1 st element	2 nd element	3 rd element	4 th element	5 th element	6 th element	7 th element	8 th element	9 th element	10 th element
<code>marks[0]</code>	<code>marks[1]</code>	<code>marks[2]</code>	<code>marks[3]</code>	<code>marks[4]</code>	<code>marks[5]</code>	<code>marks[6]</code>	<code>marks[7]</code>	<code>marks[8]</code>	<code>marks[9]</code>

Figure 2.1 Memory representation of an array of 10 elements

Arrays

1 st element	2 nd element	3 rd element	4 th element	5 th element	6 th element	7 th element	8 th element	9 th element	10 th element
marks[0]	marks[1]	marks[2]	marks[3]	marks[4]	marks[5]	marks[6]	marks[7]	marks[8]	marks[9]

Figure 2.1 Memory representation of an array of 10 elements

- Arrays are generally used when we want to store large amount of similar type of data.
- But they have the following limitations:
 - Arrays are of fixed size.
 - Data elements are stored in contiguous memory locations which may not be always available.
 - Insertion and deletion of elements can be problematic because of shifting of elements from their positions.
- However, these limitations can be solved by using linked lists.
- We will discuss more about arrays in Chapter 3.

Linked Lists

- A linked list is a very flexible, dynamic data structure in which elements (called nodes) form a sequential list.
- In contrast to static arrays, a programmer need not worry about how many elements will be stored in the linked list.
- This feature enables the programmers to write robust programs which require less maintenance.
- In a linked list, each node is allocated space as it is added to the list.

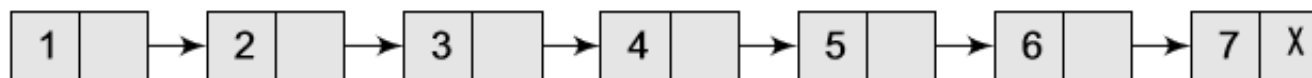


Figure 2.2 Simple linked list

Linked Lists

- Every node in the list points to the next node in the list.
 - Therefore, in a linked list, every node contains the following two types of data: The value of the node or any other data that corresponds to that node
 - A pointer or link to the next node in the list
- The last node in the list contains a NULL pointer to indicate that it is the end or tail of the list.
- Since the memory for a node is dynamically allocated when it is added to the list, the total number of nodes that may be added to a list is limited only by the amount of memory available.

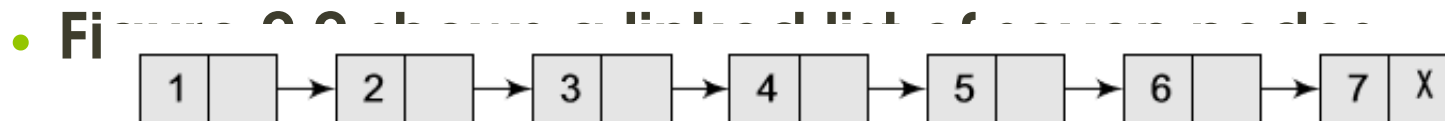


Figure 2.2 Simple linked list

CLASSIFICATION OF DATA STRUCTURES

Stacks

- A stack is a linear data structure in which insertion and deletion of elements are done at only one end, which is known as the top of the stack.
- Stack is called a last-in, first-out (LIFO) structure because the last element which is added to the stack is the first element which is deleted from the stack.
- In the computer's memory, stacks can be implemented using arrays or linked lists.
- Figure 2.3 shows the array implementation of a stack.
- Every stack has a variable top associated with it. top is used to store the address of the topmost element of the stack.
- It is this position from where the element will be added or deleted.
- There is another variable MAX, which is used to store the maximum number of elements that the stack can store.
- If $\text{top} = \text{NULL}$, then it indicates that the stack is empty and if $\text{top} = \text{MAX}-1$, then the stack is full.

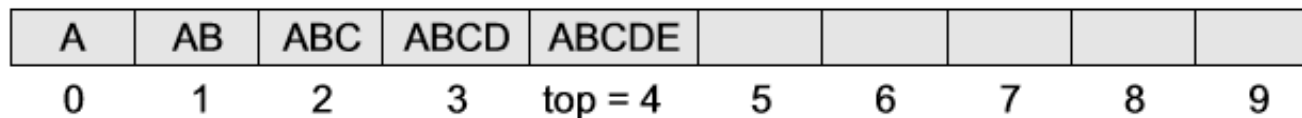


Figure 2.3 Array representation of a stack

CLASSIFICATION OF DATA STRUCTURES

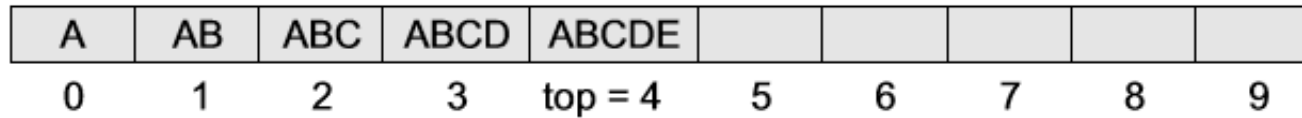


Figure 2.3 Array representation of a stack

Stacks

- In Fig. 2.3, $\text{top} = 4$, so insertions and deletions will be done at this position.
- Here, the stack can store a maximum of 10 elements where the indices range from 0–9.
- In the above stack, five more elements can still be stored.
- A stack supports three basic operations: push, pop, and peep.
- The push operation adds an element to the top of the stack.
- The pop operation removes the element from the top of the stack.
- And the peep operation returns the value of the topmost element of the stack (without deleting it).
- However, before inserting an element in the stack, we must check for overflow conditions.
- An overflow occurs when we try to insert an element into a stack that is already full.
- Similarly, before deleting an element from the stack, we must check for underflow conditions.
- An underflow condition occurs when we try to delete an element from a stack that is already empty.

CLASSIFICATION OF DATA STRUCTURES

Queues

- A queue is a first-in, first-out (FIFO) data structure in which the element that is inserted first is the first one to be taken out.
- The elements in a queue are added at one end called the rear and removed from the other end called the front.
- Like stacks, queues can be implemented by using either arrays or linked lists.
- Every queue has front and rear variables that point to the position from where deletions and insertions can be done, respectively.

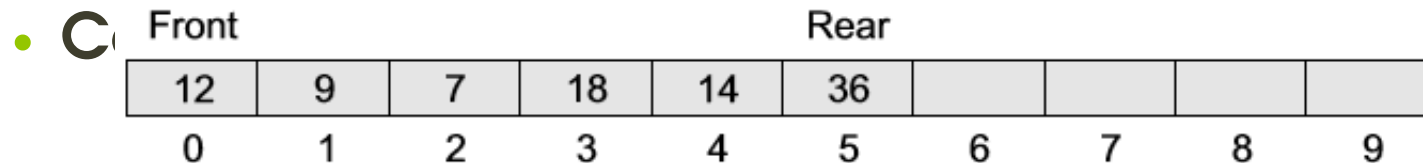


Figure 2.4 Array representation of a queue

Queues

- Here, front = 0 and rear = 5.
- If we want to add one more value to the list, say, if we want to add another element with the value 45, then the rear would be incremented by 1 and the value would be stored at the position pointed by the rear.
- The queue, after the addition, would be as shown in Fig. 2.5.
- Here, front = 0 and rear = 6.
- Every time a new element is to be added, we will repeat the same procedure.

Front			Rear						
12	9	7	18	14	36	45			
0	1	2	3	4	5	6	7	8	9

Figure 2.5 Queue after insertion of a new element

CLASSIFICATION OF DATA STRUCTURES

Queues

- Now, if we want to delete an element from the queue, then the value of front will be incremented.
- Deletions are done only from this end of the queue.
- The queue after the deletion will be as shown in Fig. 2.6.

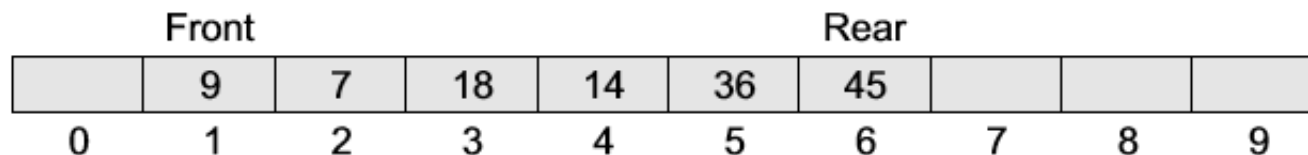


Figure 2.6 Queue after deletion of an element

CLASSIFICATION OF DATA STRUCTURES

- However, before inserting an element in the queue, we must check for overflow conditions.
- An overflow occurs when we try to insert an element into a queue that is already full.
- A queue is full when $\text{rear} = \text{MAX} - 1$, where MAX is the size of the queue, that is MAX specifies the maximum number of elements in the queue.
- Note that we have written $\text{MAX} - 1$ because the index starts from 0.
- Similarly, before deleting an element from the queue, we must check for underflow conditions.
- An underflow condition occurs when we try to delete an element from a queue that is already empty.
- If $\text{front} = \text{NULL}$ and $\text{rear} = \text{NULL}$, then there is no element in the queue.

Trees

- A tree is a non-linear data structure which consists of a collection of nodes arranged in a hierarchical order.
- One of the nodes is designated as the root node, and the remaining nodes can be partitioned into disjoint sets such that each set is a sub-tree of the root.
- The simplest form of a tree is a binary tree.
- A binary tree consists of a root node and left and right sub-trees, where both sub-trees are also binary trees.

Trees

- Each node contains a data element, a left pointer which points to the left sub-tree, and a right pointer which points to the right sub-tree.
- The root element is the topmost node which is pointed by a 'root' pointer.
- If root = NULL then the tree is empty.
- Figure 2.7 shows a binary tree, where R is the root node and T1 and T2 are the left and right subtrees of R.
- If T1 is non-empty, then T1 is said to be the left successor of R.
- Likewise, if T2 is non-empty, then it is called the right successor of R.

CLASSIFICATION OF DATA STRUCTURES

Trees

- In Fig. 2.7, node 2 is the left child and node 3 is the right child of the root node 1.
- Note that the left sub-tree of the root node consists of the nodes 2, 4, 5, 8, and 9.
- Similarly, the right sub-tree of the root node consists of the nodes 3, 6, 7, 10, 11, and 12.

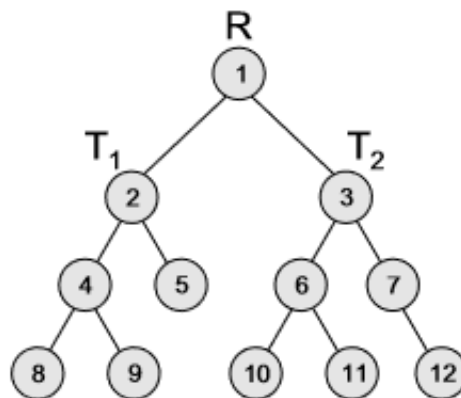


Figure 2.7 Binary tree

Graphs

- A graph is a non-linear data structure which is a collection of vertices (also called nodes) and edges that connect these vertices.
- A graph is often viewed as a generalization of the tree structure, where instead of a purely parent-to-child relationship between tree nodes, any kind of complex relationships between the nodes can exist.

CLASSIFICATION OF DATA STRUCTURES

Graphs

- In a tree structure, nodes can have any number of children but only one parent, a graph on the other hand relaxes all such kinds of restrictions.
- Figure 2.8 shows a graph with five nodes.
- A node in the graph may represent a city and the edges connecting the nodes can represent roads.
- A graph can also be used to represent a computer network where the nodes are workstations and the edges are the network connections.

CLASSIFICATION OF DATA STRUCTURES

Graphs

- Graphs have so many applications in computer science and mathematics that several algorithms have been written to perform the standard graph operations, such as searching the graph and finding the shortest path between the nodes of a graph.
- Note that unlike trees, graphs do not have any root node.
- Rather, every node in the graph can be connected with every another node in the graph.
- When two nodes are connected via an edge, the two nodes are known as neighbors.
- For example, in Fig. 2.8, node A has two neighbors: B and D.

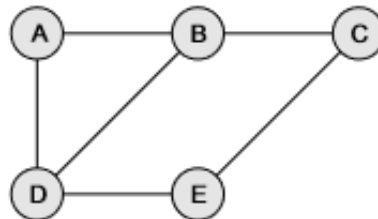


Figure 2.8 Graph

- This section discusses the different operations that can be performed on the various data structures previously mentioned.
- *Traversing*: It means to access each data item exactly once so that it can be processed. For example, to print the names of all the students in a class.
- *Searching*: It is used to find the location of one or more data items that satisfy the given constraint. Such a data item may or may not be present in the given collection of data items. For example, to find the names of all the students who secured 100 marks in mathematics.
- *Inserting*: It is used to add new data items to the given list of data items. For example, to add the details of a new student who has recently joined the course.

OPERATIONS ON DATA STRUCTURES

- **Deleting:** It means to remove (delete) a particular data item from the given collection of data items. For example, to delete the name of a student who has left the course.
- **Sorting:** Data items can be arranged in some order like ascending order or descending order depending on the type of application. For example, arranging the names of students in a class in an alphabetical order, or calculating the top three winners by arranging the participants' scores in descending order and then extracting the top three.
- **Merging:** Lists of two sorted data items can be combined to form a single list of sorted data items.
- Many a time, two or more operations are applied simultaneously in a given situation. For example, if we want to delete the details of a student whose name is X, then we first have to search the list of students to find whether the record of X exists or not and if it exists then at which location, so that the details can be deleted from that particular location

ABSTRACT DATA TYPE

- An abstract data type (ADT) is the way we look at a data structure, focusing on what it does and ignoring how it does its job.
- For example, stacks and queues are perfect examples of an ADT.
- We can implement both these ADTs using an array or a linked list.
- This demonstrates the 'abstract' nature of stacks and queues.
- To further understand the meaning of an abstract data type, we will break the term into 'data type' and 'abstract', and then discuss their meanings.
- Data type Data type of a variable is the set of values that the variable can take.

ABSTRACT DATA TYPE

- We have already read the basic data types in C include int, char, float, and double.
- When we talk about a primitive type (built-in data type), we actually consider two things: a data item with certain characteristics and the permissible operations on that data.
- For example, an int variable can contain any whole-number value from -32768 to 32767 and can be operated with the operators $+$, $-$, $*$, and $/$.
- In other words, the operations that can be performed on a data type are an inseparable part of its identity.
- Therefore, when we declare a variable of an abstract data type (e.g., stack or a queue), we also need to specify the operations that can be performed on it.

ABSTRACT DATA TYPE

- **Abstract** The word ‘abstract’ in the context of data structures means considered apart from the detailed specifications or implementation.
- In C, an abstract data type can be a structure considered without regard to its implementation.
- It can be thought of as a ‘description’ of the data in the structure with a list of operations that can be performed on the data within that structure.
- The end-user is not concerned about the details of how the methods carry out their tasks.
- They are only aware of the methods that are available to them and are only concerned about calling those methods and getting the results.

ABSTRACT DATA TYPE

- They are not concerned about how they work.
- For example, when we use a stack or a queue, the user is concerned only with the type of data and the operations that can be performed on it.
- Therefore, the fundamentals of how the data is stored should be invisible to the user.
- They should not be concerned with how the methods work or what structures are being used to store the data.
- They should just know that to work with stacks, they have `push()` and `pop()` functions available to them.
- Using these functions, they can manipulate the data (insertion or deletion) stored in the stack.

ABSTRACT DATA TYPE

- **Advantage of using ADTs** In the real world, programs evolve as a result of new requirements or constraints, so a modification to a program commonly requires a change in one or more of its data structures.
- For example, if you want to add a new field to a student's record to keep track of more information about each student, then it will be better to replace an array with a linked structure to improve the program's efficiency.
- In such a scenario, rewriting every procedure that uses the changed structure is not desirable.
- Therefore, a better alternative is to separate the use of a data structure from the details of its implementation.
- This is the principle underlying the use of abstract data types.

- The typical definition of algorithm is ‘a formally defined procedure for performing some calculation’.
- If a procedure is formally defined, then it can be implemented using a formal language, and such a language is known as a programming language.
- In general terms, an algorithm provides a blueprint to write a program to solve a particular problem.
- It is considered to be an effective procedure for solving a problem in finite number of steps.
- That is, a well-defined algorithm always provides an answer and is guaranteed to terminate.

- Algorithms are mainly used to achieve software reuse.
- Once we have an idea or a blueprint of a solution, we can implement it in any high-level language like C, C++, or Java.
- An algorithm is basically a set of instructions that solve a problem.
- It is not uncommon to have multiple algorithms to tackle the same problem, but the choice of a particular algorithm must depend on the time and space complexity of the algorithm.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

- Algorithms are used to manipulate the data contained in data structures.
- When working with data structures, algorithms are used to perform operations on the stored data.
- A complex algorithm is often divided into smaller units called modules.
- This process of dividing an algorithm into modules is called modularization.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

- The key advantages of modularization are as follows:
 - It makes the complex algorithm simpler to design and implement.
 - Each module can be designed independently.
 - While designing one module, the details of other modules can be ignored, thereby enhancing clarity in design which in turn simplifies implementation, debugging, testing, documenting, and maintenance of the overall algorithm.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

- There are two main approaches to design an algorithm—top-down approach and bottom-up approach, as shown in Fig. 2.9.

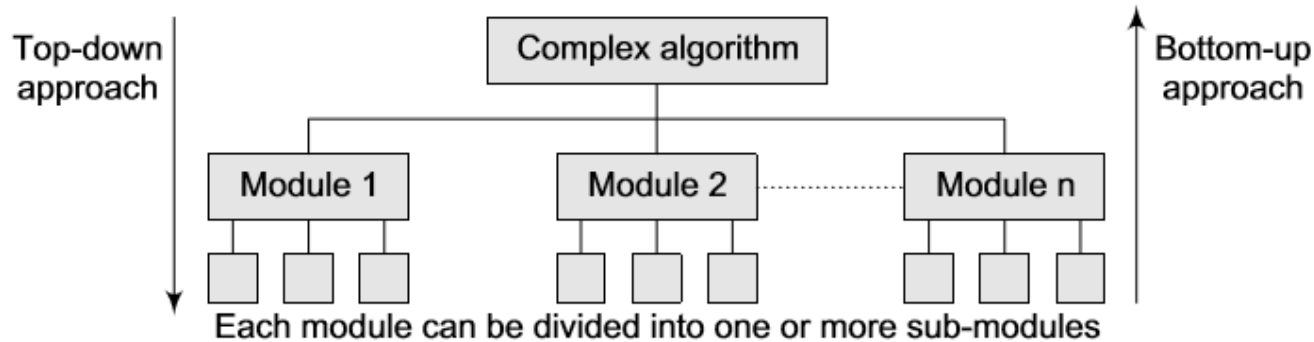


Figure 2.9 Different approaches of designing an algorithm

Top-down approach

- A top-down design approach starts by dividing the complex algorithm into one or more modules.
- These modules can further be decomposed into one or more sub-modules, and this process of decomposition is iterated until the desired level of module complexity is achieved.
- Top-down design method is a form of stepwise refinement where we begin with the topmost module and incrementally add modules that it calls.
- Therefore, in a top-down approach, we start from an abstract design and then at each step, this design is refined into more concrete levels until a level is reached that requires no further refinement.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

- There are two main approaches to design an algorithm—top-down approach and bottom-up approach, as shown in Fig. 2.9.

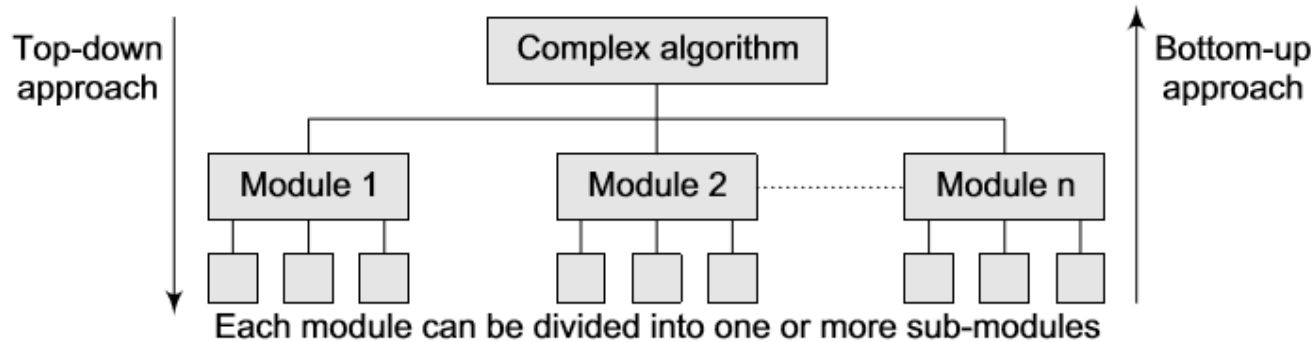


Figure 2.9 Different approaches of designing an algorithm

Bottom-up approach

- A bottom-up approach is just the reverse of top-down approach. In the bottom-up design, we start with designing the most basic or concrete modules and then proceed towards designing higher level modules.
- The higher level modules are implemented by using the operations performed by lower level modules.
- Thus, in this approach sub-modules are grouped together to form a higher level module.
- All the higher level modules are clubbed together to form even higher level modules.
- This process is repeated until the design of the complete algorithm is obtained.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

Top-down vs bottom-up approach

- Whether the top-down strategy should be followed or a bottom-up is a question that can be answered depending on the application at hand.
- While top-down approach follows a stepwise refinement by decomposing the algorithm into manageable modules, the bottom-up approach on the other hand defines a module and then groups together several modules to form a new higher level module.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

47

Top-down vs bottom-up approach

- Top-down approach is highly appreciated for ease in documenting the modules, generation of test cases, implementation of code, and debugging.
- However, it is also criticized because the sub-modules are analyzed in isolation without concentrating on their communication with other modules or on reusability of components and little attention is paid to data, thereby ignoring the concept of information hiding.

DIFFERENT APPROACHES TO DESIGNING AN ALGORITHM

Top-down vs bottom-up approach

- Although the bottom-up approach allows information hiding as it first identifies what has to be encapsulated within a module and then provides an abstract interface to define the module's boundaries as seen from the clients.
- But all this is difficult to be done in a strict bottom-up strategy.
- Some top-down activities need to be performed for this.
- All in all, design of complex algorithms must not be constrained to proceed according to a fixed pattern but should be a blend of top-down and bottom-up approaches.

- An algorithm has a finite number of steps.
- Some steps may involve decision-making and repetition.
- Broadly speaking, an algorithm may employ one of the following control structures: (a) sequence, (b) decision, and (c) repetition.
- **Sequence:** By sequence, we mean that each step of an algorithm is executed in a specified order.
- Let us write an algorithm to add two numbers.
- This algorithm performs the steps in a purely sequential order, as shown in Fig. 2.10.

```
Step 1: Input first number as A  
Step 2: Input second number as B  
Step 3: SET SUM = A+B  
Step 4: PRINT SUM  
Step 5: END
```

Figure 2.10 Algorithm to add two numbers

Decision:

- Decision statements are used when the execution of a process depends on the outcome of some condition.
- For example, if $x = y$, then print EQUAL. So the general form of IF construct can be given as: IF condition Then process A condition in this context is any statement that may evaluate to either a true value or a false value.
- In the above example, a variable x can be either equal to y or not equal to y . However, it cannot be both true and false.
- If the condition is true, then the process is executed.

Decision:

- A decision statement can also be stated in the following manner: IF condition Then process1 ELSE process2 This form is popularly known as the IF–ELSE construct.
- Here, if the condition is true, then process1 is executed, else process2 is executed.
- Figure 2.11 shows an algorithm to check if two numbers are equal

```

Step 1: Input first number as A
Step 2: Input second number as B
Step 3: IF A = B
        PRINT "EQUAL"
      ELSE
        PRINT "NOT EQUAL"
      [END OF IF]
Step 4: END
    
```

Figure 2.11 Algorithm to test for equality of two numbers

Repetition:

- Repetition, which involves executing one or more steps for a number of times, can be implemented using constructs such as while, do-while, and for loops.
- These loops execute one or more steps until some condition is true.
- Figure 2.12 shows an algorithm that prints the first 10 natural numbers.

```
Step 1: [INITIALIZE] SET I = 1, N = 10  
Step 2: Repeat Steps 3 and 4 while I<=N  
Step 3: PRINT I  
Step 4: SET I = I+1  
        [END OF LOOP]  
Step 5: END
```

Figure 2.12 Algorithm to print the first 10 natural of

TIME AND SPACE COMPLEXITY

- Analyzing an algorithm means determining the amount of resources (such as time and memory) needed to execute it.
- Algorithms are generally designed to work with an arbitrary number of inputs, so the efficiency or complexity of an algorithm is stated in terms of time and space complexity.
- The time complexity of an algorithm is basically the running time of a program as a function of the input size.
- Similarly, the space complexity of an algorithm is the amount of computer memory that is required during the program execution as a function of the input size.
- In other words, the number of machine instructions which a program executes is called its time complexity.
- This number is primarily dependent on the size of the program's input and the algorithm used.

- **Generally, the space needed by a program depends on the following two parts:**
 - **Fixed part:** It varies from problem to problem. It includes the space needed for storing instructions, constants, variables, and structured variables (like arrays and structures).
 - **Variable part:** It varies from program to program. It includes the space needed for recursion stack, and for structured variables that are allocated space dynamically during the runtime of a program.
- **However, running time requirements are more critical than memory requirements.**
- **Therefore, in this section, we will concentrate on the running time efficiency of algorithms.**

TIME AND SPACE COMPLEXITY

Worst-case, Average-case, Best-case, and Amortized Time Complexity

- **Worst-case running time** This denotes the behavior of an algorithm with respect to the worst possible case of the input instance.
- The worst-case running time of an algorithm is an upper bound on the running time for any input.
- Therefore, having the knowledge of worst-case running time gives us an assurance that the algorithm will never go beyond this time limit.
- **Average-case running time** The average-case running time of an algorithm is an estimate of the running time for an 'average' input.
- It specifies the expected behavior of the algorithm when the input is randomly drawn from a given distribution.

TIME AND SPACE COMPLEXITY

Worst-case, Average-case, Best-case, and Amortized Time Complexity

- Average-case running time assumes that all inputs of a given size are equally likely.
- Best-case running time The term 'best-case performance' is used to analyze an algorithm under optimal conditions. For example, the best case for a simple linear search on an array occurs when the desired element is the first in the list.
- However, while developing and choosing an algorithm to solve a problem, we hardly base our decision on the best-case performance.
- It is always recommended to improve the average performance and the worst-case performance of an algorithm.
- Amortized running time refers to the time required to perform a sequence of (related) operations averaged over all the operations performed.
- Amortized analysis guarantees the average performance of each operation in the worst case.

Time–Space Trade-off

- The best algorithm to solve a particular problem at hand is no doubt the one that requires less memory space and takes less time to complete its execution.
- But practically, designing such an ideal algorithm is not a trivial task.
- There can be more than one algorithm to solve a particular problem.
- One may require less memory space, while the other may require less CPU time to execute.

Time–Space Trade-off

- Thus, it is not uncommon to sacrifice one thing for the other.
- Hence, there exists a time–space trade-off among algorithms.
- So, if space is a big constraint, then one might choose a program that takes less space at the cost of more CPU time.
- On the contrary, if time is a major constraint, then one might choose a program that takes minimum time to execute at the cost of more space.

Expressing Time and Space Complexity

- The time and space complexity can be expressed using a function $f(n)$ where n is the input size for a given instance of the problem being solved.
- Expressing the complexity is required when
 - We want to predict the rate of growth of complexity as the input size of the problem increases.
 - There are multiple algorithms that find a solution to a given problem and we need to find the algorithm that is most efficient.
- The most widely used notation to express this function $f(n)$ is the Big O notation. It provides the upper bound for the complexity.

Algorithm Efficiency

- If a function is linear (without any loops or recursions), the efficiency of that algorithm or the running time of that algorithm can be given as the number of instructions it contains.
- However, if an algorithm contains loops, then the efficiency of that algorithm may vary depending on the number of loops and the running time of each loop in the algorithm.

Algorithm Efficiency

- Let us consider different cases in which loops determine the efficiency of an algorithm.
- *Linear Loops:* To calculate the efficiency of an algorithm that has a single loop, we need to first determine the number of times the statements in the loop will be executed.
- This is because the number of iterations is directly proportional to the loop factor.
- Greater the loop factor, more is the number of iterations.
- For example, consider the loop given below:
for(i=0;i<100;i++) statement block;
- Here, 100 is the loop factor.

Algorithm Efficiency

- We have already said that efficiency is directly proportional to the number of iterations.
- Hence, the general formula in the case of linear loops may be given as $f(n) = n$
- However calculating efficiency is not as simple as is shown in the above example.
- Consider the loop given below: `for(i=0;i<100;i+=2)` statement block;
- Here, the number of iterations is half the number of the loop factor. So, here the efficiency can be given as $f(n) = n/2$

Logarithmic Loops

- We have seen that in linear loops, the loop updation statement either adds or subtracts the loop-controlling variable.
- However, in logarithmic loops, the loop-controlling variable is either multiplied or divided during each iteration of the loop.
- For example, look at the loops given below:
`for(i=1;i<1000;i*=2)` `for(i=1000;i>=1;i/=2)` statement block; statement block;
- Consider the first for loop in which the loop-controlling variable *i* is multiplied by 2.
- The loop will be executed only 10 times and not 1000 times because in each iteration the value of *i* doubles.

Logarithmic Loops

- Now, consider the second loop in which the loop-controlling variable i is divided by 2.
- In this case also, the loop will be executed 10 times.
- Thus, the number of iterations is a function of the number by which the loop-controlling variable is divided or multiplied.
- In the examples discussed, it is 2.
- That is, when $n = 1000$, the number of iterations can be given by $\log 1000$ which is approximately equal to 10.
- Therefore, putting this analysis in general terms, we can conclude that the efficiency of loops in which iterations divide or multiply the loop-controlling variables can be given as $f(n) = \log n$

Nested Loops

- Loops that contain loops are known as nested loops.
- In order to analyze nested loops, we need to determine the number of iterations each loop completes.
- The total is then obtained as the product of the number of iterations in the inner loop and the number of iterations in the outer loop. In this case, we analyze the efficiency of the algorithm based on whether it is a linear logarithmic, quadratic, or dependent quadratic nested loop.

Linear logarithmic loop

- Consider the following code in which the loop-controlling variable of the inner loop is multiplied after each iteration.
- The number of iterations in the inner loop is $\log 10$.
- This inner loop is controlled by an outer loop which iterates 10 times.
- Therefore, according to the formula, the number of iterations for this code can be given as $10 \log 10$.

```
for(i=0;i<10;i++)
```

```
    for(j=1; j<10;j*=2) statement block;
```

- In more general terms, the efficiency of such loops can be given as $f(n) = n \log n$.

Quadratic loop

- In a quadratic loop, the number of iterations in the inner loop is equal to the number of iterations in the outer loop.
- Consider the following code in which the outer loop executes 10 times and for each iteration of the outer loop, the inner loop also executes 10 times.
- Therefore, the efficiency here is 100.
 for(i=0;i<10;i++)
 for(j=0; j<10;j++) statement block;
- The generalized formula for quadratic loop can be given as $f(n) = n^2$.
- Dependent quadratic loop In a dependent quadratic loop, the number of iterations in the inner loop is dependent on the outer loop.

TIME AND SPACE COMPLEXITY

Quadratic loop

- Consider the code given below:

```
for(i=0;i<10;i++)
```

```
    for(j=0; j<=i;j++) statement block;
```

- In this code, the inner loop will execute just once in the first iteration, twice in the second iteration, thrice in the third iteration, so on and so forth.
- In this way, the number of iterations can be calculated as $1 + 2 + 3 + \dots + 9 + 10 = 55$
- If we calculate the average of this loop ($55/10 = 5.5$), we will observe that it is equal to the number of iterations in the outer loop (10) plus 1 divided by 2.
- In general terms, the inner loop iterates $(n + 1)/2$ times.
- Therefore, the efficiency of such a code can be given as $f(n) = n(n + 1)/2$