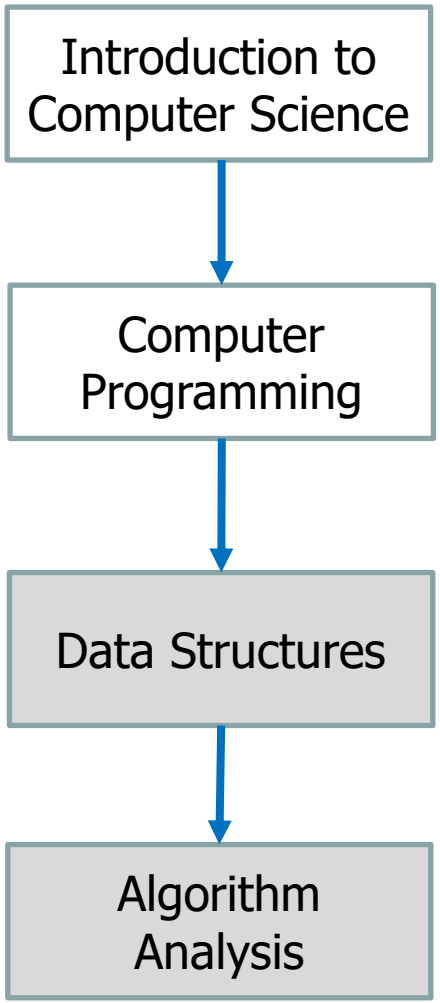
CS2001--Data Structures
Fall 2021

(06-09-2021)

Muhammad Husnain
National University of Computer and Emerging Sciences

General Overview



What is Hardware, Software, Programming, Operating System etc.

How to write software with the help of procedural and object-oriented programming.

How to organize information/data, usually in memory, so that it can be used efficiently.

How to analyze the performance of algorithms to evaluate their suitability for solving complex problems?

Some Rules

- Raise your hand before asking any question (for physical classroom)
  - Wait for the permission
- Never ever miss a class
  - No retakes (except for Final Exam*)
- Never ever "sleep" in the class
  - You might miss a class participation ☹
- Never use mobile phone in the class
  - Violators will be penalized (Drinks and/or Snacks) ☺

- Above all, whatever you do, please do not disturb others

* Conditional: as per university policy

Grading Criteria (Tentative)

- Assignments/ Quizzes 20%
- Midterm Examination 1 15%
- Midterm Examination 2 15%
- Project 05%
- Final Examination 45%

Dishonesty and Plagiarism

- Any kind of cheating will be considered serious offense
- All parties involve in cheating will get minus (-) 50% marks
  - Exams, quizzes, assignments, projects
- Habitual cases will be awarded "F"

Assignments and Quizes

- A number of assignments and quizzes will be taken
- Quizzes will always be pre-announced

What is a Data Structure?

- In a general sense, any representation that is used for storing information is a data structure
  - Example: integers, structures, classes, arrays, etc.

- More typically, a data structure provides a way for organizing a collection of data items
  - Storing, accessing and modifying data items

- Most data structures have associated algorithms to perform operations on data
  - Search, insert, remove etc.

Costs and Benefits

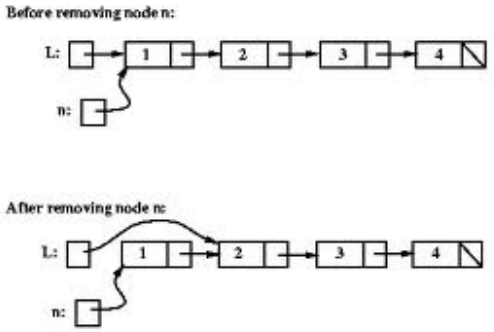
- Each data structure requires
  - Space for storing data items
  - Time to perform each basic operation
  - Programming efforts

- Cost of a data structure
  - Time and space resources it consumes

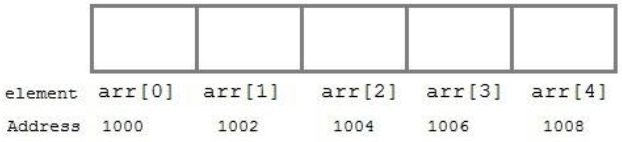
- Choice of data structure depends on many factors
  - Type of data
  - Frequency with which various data operations are applied

- Hard to define data structure that performs better in all situations
  - Time and space tradeoff

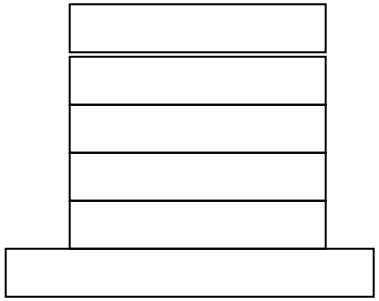
Some Example Data Structures



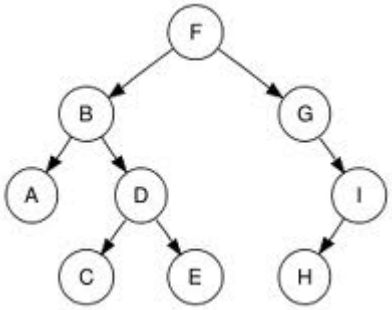
Linked List



Arrays



Stack



Tree

Data structure = representation and operations associated with a data type

Data Structure Applications

- Arrays
  - Consecutive memory locations
  - lists (one dimensional arrays)
  - matrices (two dimensional arrays)
  - database applications
  - to implement other data structures, such as heaps, queues, stacks, etc.
- Stacks
  - expression evaluation and syntax parsing
- Queues
  - Task scheduling
  - Helpful in BFS in trees
  - First in first out management

Data Structure Applications

- Trees
  - efficient searching of data (Binary search tree)
  - manipulate hierarchical data
- Linked lists
  - can be used to implement several other common abstract data structures, such as
    - stacks
    - queues
    - trees etc

Classification of Data Structures

- **Linear**
  - Data values are arranged in linear/sequential fashion
  - E.g., Arrays, Linked lists, stacks and queues
- **Non-Linear**
  - Data values are not arranged in sequential order
  - E.g., Graphs, and Trees etc.
- **Homogenous**
  - Same data values are stored or grouped together.
  - Such as Arrays
- **Non-Homogenous**
  - Data values of different types are grouped together
  - E.g., structures and classes in C++ can group multiple types of data together

Common Operations on Data Structures

Some commonly used operations performed on data structures are:

- **Inserting:** adding new data items to the structure.
- **Deleting:** removing data items from a data structure.
- **Searching:** finding specific item in a data structure.
- **Traversing:** accessing each record in the structure exactly once for processing. Also know as visiting.
- **Sorting:** arranging data item in a data structure into specific order.
- **Merging (not so frequent):** combining contents of two similar data structures into one.

What is an Algorithm?

- In mathematics and computer science, an **algorithm** is a step-by-step procedure for solving a problem
  - expressed as a finite list of well-defined instructions
  - requiring a finite amount of time
  - on a finite amount of data

- Algorithms are used for
  - calculation
  - data processing, and
  - Solving complex problem

- In simple words
  - an algorithm is a step-by-step procedure for calculations and operations

Course Objective

- Introduce the basic concepts of data structures, and use them efficiently in algorithms for solving various problems using C/C++
- What should you expect in this course?
  - A lot of thinking
  - Extensive programming
- What should you learn by the end of this course
  - Ability to understand common programming problems and design and implement efficient data structures to solve them

Why should we study this course?

- Well, because it is the core CS course

- Any other reason to study this course?
  - We want to make a successful career after graduation
  - The most common programming interview questions are
    - Linked lists
    - Strings
    - Binary Trees
    - Arrays
    - Queues
- Any other?
  - Obliviously it will boost your programming logic.

Source: http://maxnoy.com/interviews.html

Course Content

- Simple and abstract data types
- Arrays
- Searching and sorting techniques
- Linked Lists
- Stacks
- Queues

First Session

- Trees
- Heap
- Hashing
- Graphs

Final

Motivational Example

- A cellular company provides contacts and related services to its 40 million users
- Due to new security enforcements, the company wants to prevent issuing of multiple contacts to the same user

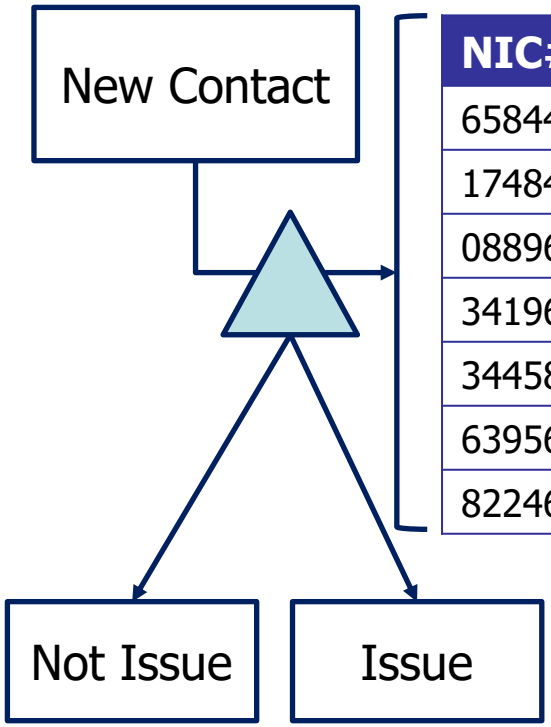
- Method of Detecting Multiple Contacts
  - Before issuing a new contact to a user:
  - First search the id of the user in the existing contacts database
  - In case of a failure → issue a new contact to the user
  - In case of a success → do not issue a new contact to the user

Example: Linear Array Data Structure (1)

NIC#	Name	Address
6584495-9	Muhammad Faheem	House No 3 Gulshan Bahar Sec 16
1748425-5	Naeem Alam	A-11 Shams Plaza Block-B N.Nazimabad
0889679-1	Arslan Akhtar	H No 152 Bostang Colony
3419668-1	Zain Ahmed	Sharfabad Street Gulshan Karachi
3445864-3	Sumair Farooq	Post Office Tayyar, Multan
6395653-4	Ali Affan	H.No. 425, Sector F-11/4, Islamabad
8224641-1	Syed Faraz	Sharfabad Street Gulshan, Faisalabad

- Linear Array (with 40 million entries)
  - Three arrays (NIC, Name, Address)
  - One array of structures *(struct)*
  - One array of class's objects

Example: Linear Array Data Structure (2)



NIC#	Name	Address
6584495-9	Muhammad Faheem	House No 3 Gulshan Bahar Sec 16
1748425-5	Naeem Alam	A-11 Shams Plaza Block-B N.Nazimabad
0889679-1	Arslan Akhtar	H No 152 Bostang Colony
3419668-1	Zain Ahmed	Sharfabad Street Gulshan Karachi
3445864-3	Sumair Farooq	Post Office Tayyar, Multan
6395653-4	Ali Affan	H.No. 425, Sector F-11/4, Islamabad
8224641-1	Syed Faraz	Sharfabad Street Gulshan, Faisalabad

- Any disadvantage of the linear array data structure?
- How to improve?

Example: Improved Data Structure

- Create a dictionary like data structure
  - Group similar records together
  - Similarity in terms of first digit of NIC number
  - Add a dictionary entry for each distinct digit (0 – 9)

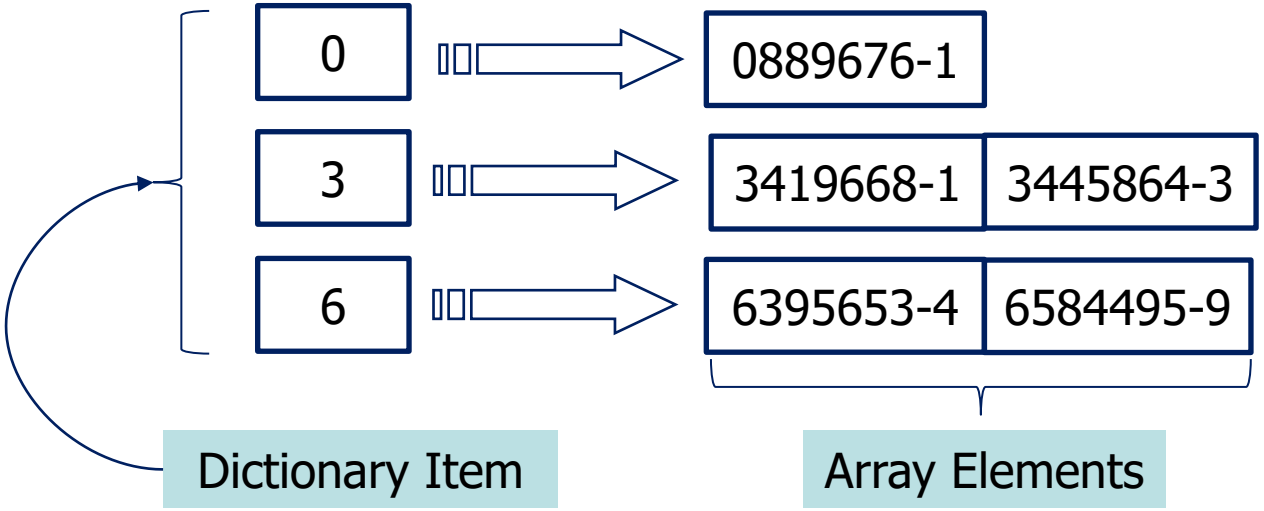
- Example: 3419668-1, 3445864-3, 1748425-5.
  - 3 and 1 are dictionary entries

- Existing contacts are searched in two steps
  1. Search the dictionary entry (i.e., group searching)
  2. Search contact within the group (i.e., with the same NIC digit)

- How much improvement w.r.t. linear array?

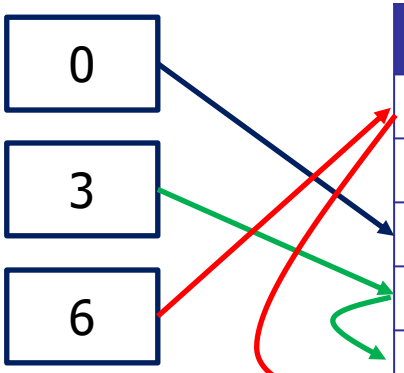
Example: Dictionary Data Structure (1)

NIC#	Name	Address
6584495-9	Muhammad Faheem	House No 3 Gulshan Bahar Sec 16
1748425-5	Naeem Alam	A-11 Shams Plaza Block-B N.Nazimabad
0889679-1	Arslan Akhtar	H No 152 Bostang Colony
3419668-1	Zain Ahmed	Sharfabad Street Gulshan Karachi
3445864-3	Sumair Farooq	Post Office Tayyar, Multan
6395653-4	Ali Affan	H.No. 425, Sector F-11/4, Islamabad
8224641-1	Syed Faraz	Sharfabad Street Gulshan, Faisalabad



Example: Dictionary Data Structure (2)

- Another possibility
  - Maintain pointers with structures (or records)
  - Non-NULL pointer indicates presence of next record



NIC#	Name	Address
6584495-9	Muhammad Faheem	House No 3 Gulshan Bahar Sec 16
1748425-5	Naeem Alam	A-11 Shams Plaza Block-B N.Nazimabad
0889679-1	Arslan Akhtar	H No 152 Bostang Colony
3419668-1	Zain Ahmed	Sharfabad Street Gulshan Karachi
3445864-3	Sumair Farooq	Post Office Tayyar, Multan
6395653-4	Ali Affan	H.No. 425, Sector F-11/4, Islamabad
8224641-1	Syed Faraz	Sharfabad Street Gulshan, Faisalabad

Any Question So Far?