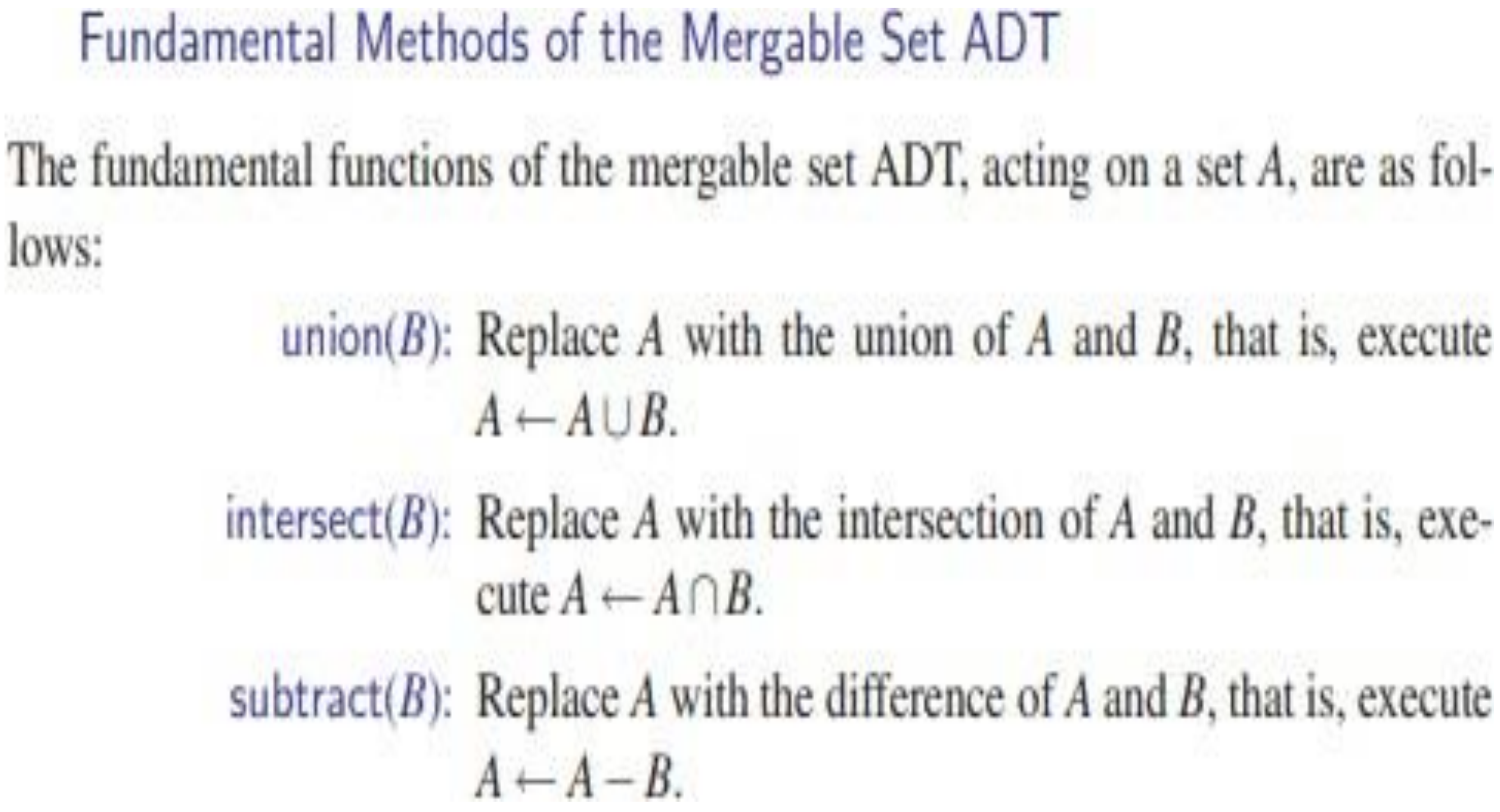
Sets, Maps and Dictionary

Ref: Data Structures and Algorithms in C++ 2e By Michael Goodrich, Roberto Tamassia and David Mount

Sets

- A set is defined as a collection that contains no duplicates
- Basic Operations we perform with sets are
  - Set Union(S1 U S2)
  - Set Intersection(S1 ∩ S2)
  - Set Difference(S1 - S2)

Sets



Fundamental Methods of the Mergable Set ADT

The fundamental functions of the mergable set ADT, acting on a set A, are as follows:

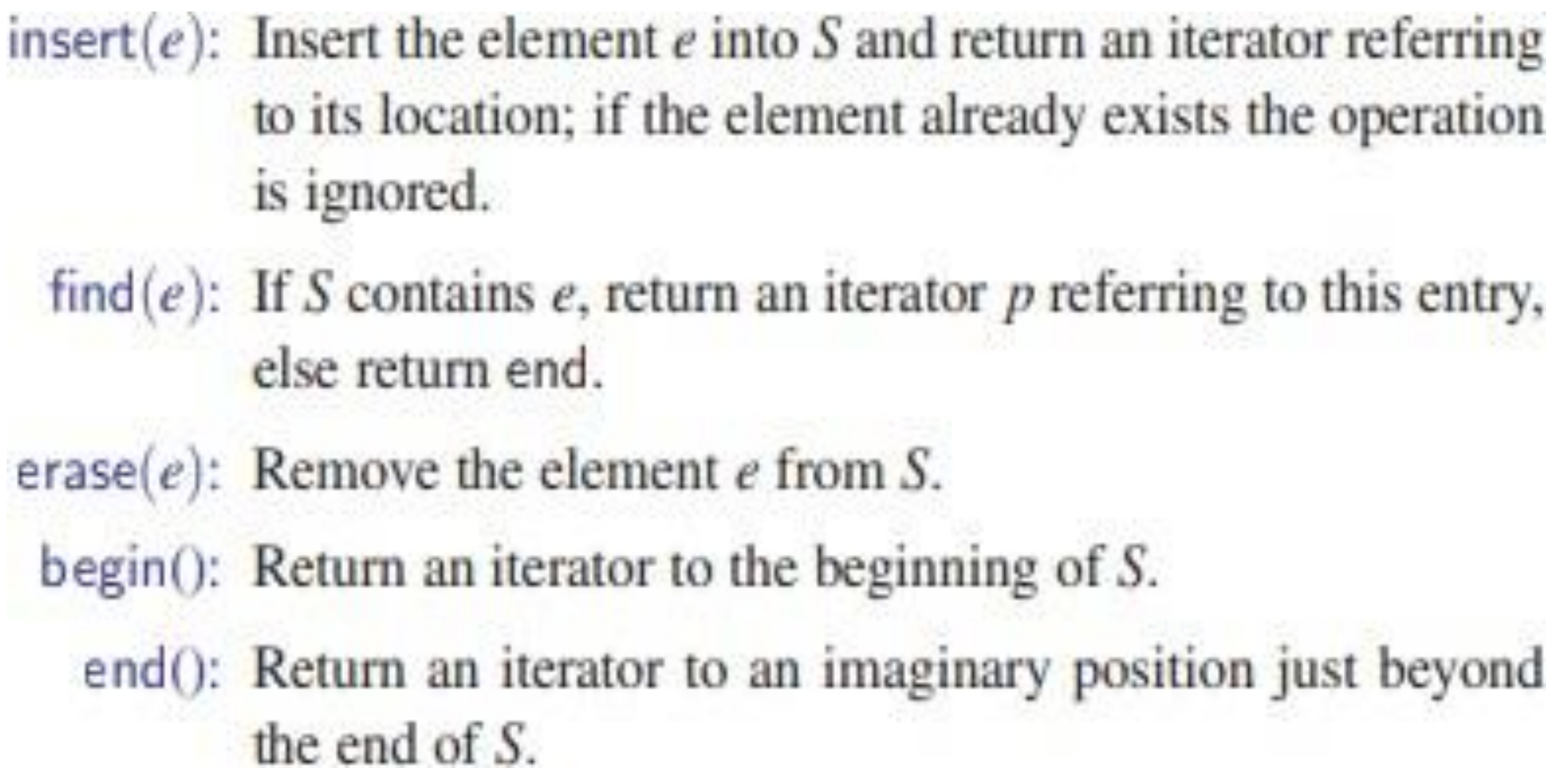
union(B): Replace A with the union of A and B, that is, execute $A \leftarrow A \cup B$.

intersect(B): Replace A with the intersection of A and B, that is, execute $A \leftarrow A \cap B$.

subtract(B): Replace A with the difference of A and B, that is, execute $A \leftarrow A - B$.

Sets ADT

Set ADT provides number of methods.



insert(e):	Insert the element e into S and return an iterator referring to its location; if the element already exists the operation is ignored.
find(e):	If S contains e, return an iterator p referring to this entry, else return end.
erase(e):	Remove the element e from S.
begin():	Return an iterator to the beginning of S.
end():	Return an iterator to an imaginary position just beyond the end of S.

basic functions associated with Set:

Set in C++ Standard Template Library (STL)

- begin() – Returns an iterator to the first element in the set.
- end() – Returns an iterator to the theoretical element that follows last element in the set.
- size() – Returns the number of elements in the set.
- max_size() – Returns the maximum number of elements that the set can hold.
- empty() – Returns whether the set is empty.

Reading assignment: Set in C++ Standard Template Library (STL)

- rbegin()– Returns a reverse iterator pointing to the last element in the container.
- rend()– Returns a reverse iterator pointing to the theoretical element right before the first element in the set container.
- crbegin()– Returns a constant iterator pointing to the last element in the container.
- crend() – Returns a constant iterator pointing to the position just before the first element in the container.

Reading assignment: Set in C++ Standard Template Library (STL)

- cbegin()– Returns a constant iterator pointing to the first element in the container.
- cend() – Returns a constant iterator pointing to the position past the last element in the container.
- size() – Returns the number of elements in the set.
- max_size() – Returns the maximum number of elements that the set can hold.
- empty() – Returns whether the set is empty.

Reading assignment: Set in C++ Standard Template Library (STL)

- insert(const g) – Adds a new element ‘g’ to the set.
- iterator insert (iterator position, const g) – Adds a new element ‘g’ at the position pointed by iterator.
- erase(iterator position) – Removes the element at the position pointed by the iterator.
- erase(const g)– Removes the value ‘g’ from the set.
- clear() – Removes all the elements from the set.

Reading assignment: Set in C++ Standard Template Library (STL)

- key_comp() / value_comp() – Returns the object that determines how the elements in the set are ordered ('<' by default).
- find(const g) – Returns an iterator to the element 'g' in the set if found, else returns the iterator to end.
- count(const g) – Returns 1 or 0 based on the element 'g' is present in the set or not.
- lower_bound(const g) – Returns an iterator to the first element that is equivalent to 'g' or definitely will not go before the element 'g' in the set.

Reading assignment: Set in C++ Standard Template Library (STL)

- upper_bound(const g) – Returns an iterator to the first element that will go after the element ‘g’ in the set.
- equal_range()– The function returns an iterator of pairs. (key_comp). The pair refers to the range that includes all the elements in the container which have a key equivalent to k.
- emplace()– This function is used to insert a new element into the set container, only if the element to be inserted is unique and does not already exists in the set.

Reading assignment: Set in C++ Standard Template Library (STL)

- emplace_hint()– Returns an iterator pointing to the position where the insertion is done. If the element passed in the parameter already exists, then it returns an iterator pointing to the position where the existing element is.
- swap()– This function is used to exchange the contents of two sets but the sets must be of same type, although sizes may differ.
- operator= – The '=' is an operator in C++ STL which copies (or moves) a set to another set and set::operator= is the corresponding operator function.
- get_allocator()– Returns the copy of the allocator object associated with the set.

Reading assignment

Refer: https://www.geeksforgeeks.org/set-in-cpp-stl/

Disjoint sets and partitions

A1 and A2 are called disjoint partitions of A iff

- A1 U A2 = A
- A1 $\cap$ A2 = Φ

- E.g. A1={1,2,3,4,5} and A2= {2,4,6}, A3= {6,7} and A= {1,2,3,4,5,6,7}
- A1 and A2 are not disjoint partitions of A
- A1 and A3 are disjoint partitions of A

Set partition using union- find operation

- Union: creates disjoint subsets
- Find: checks connectivity

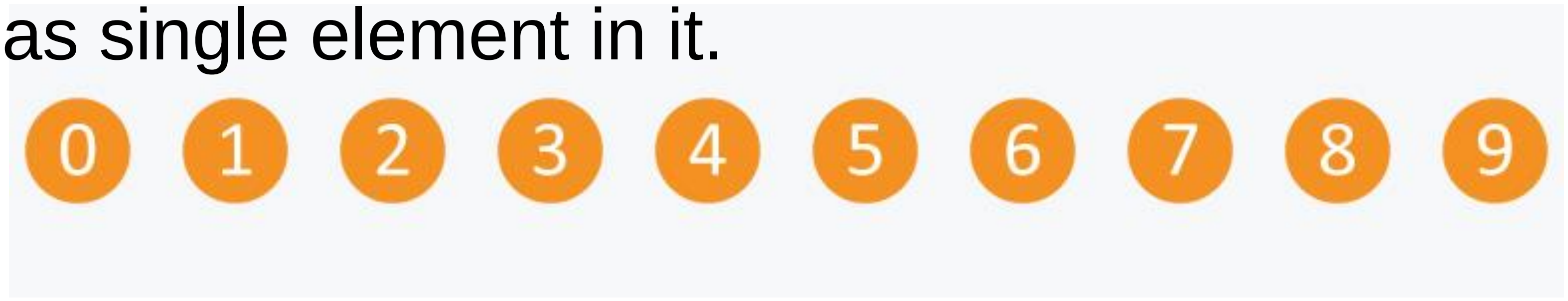
Example

Example:

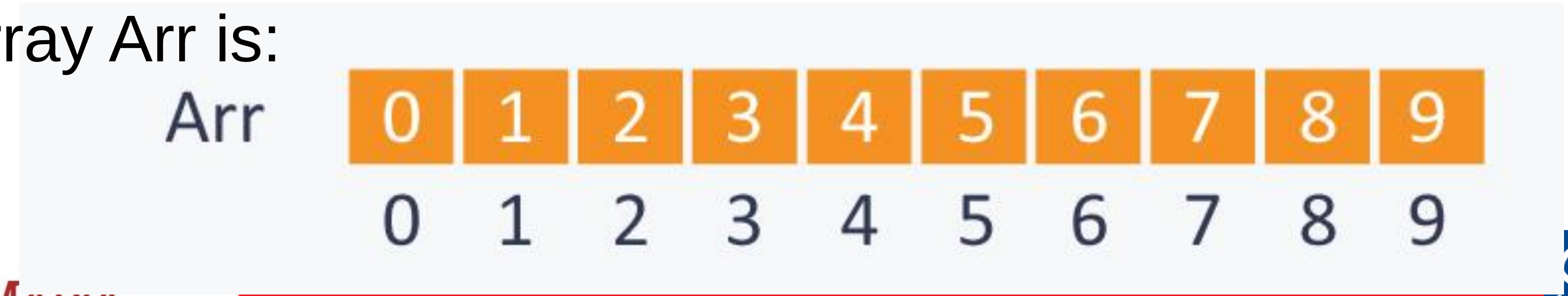
S = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}.

N = 10

Initially there are 10 subsets and each subset has single element in it.



When each subset contains only single element, the array Arr is:



Ref: https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

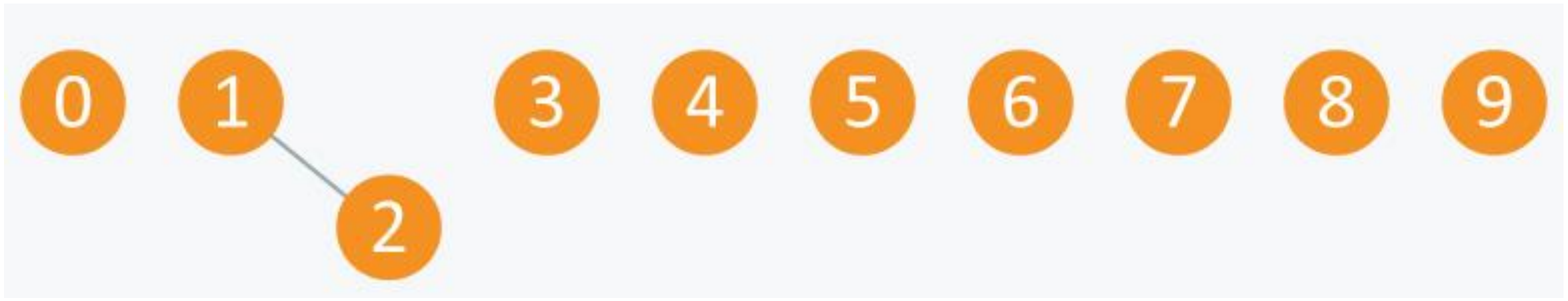
Example

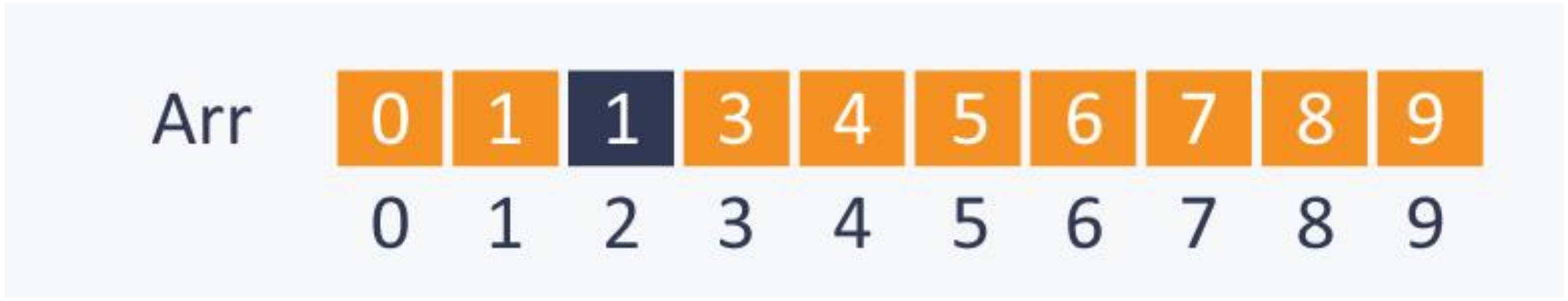
Perform the following operations on the set:

1) Union(2, 1)
2) Union(4, 3)
3) Union(8, 4)
4) Union(9, 3)
5) Union(6, 5)
6) Union(5, 2)

Find(6,1), find(8,9) find(7,1)

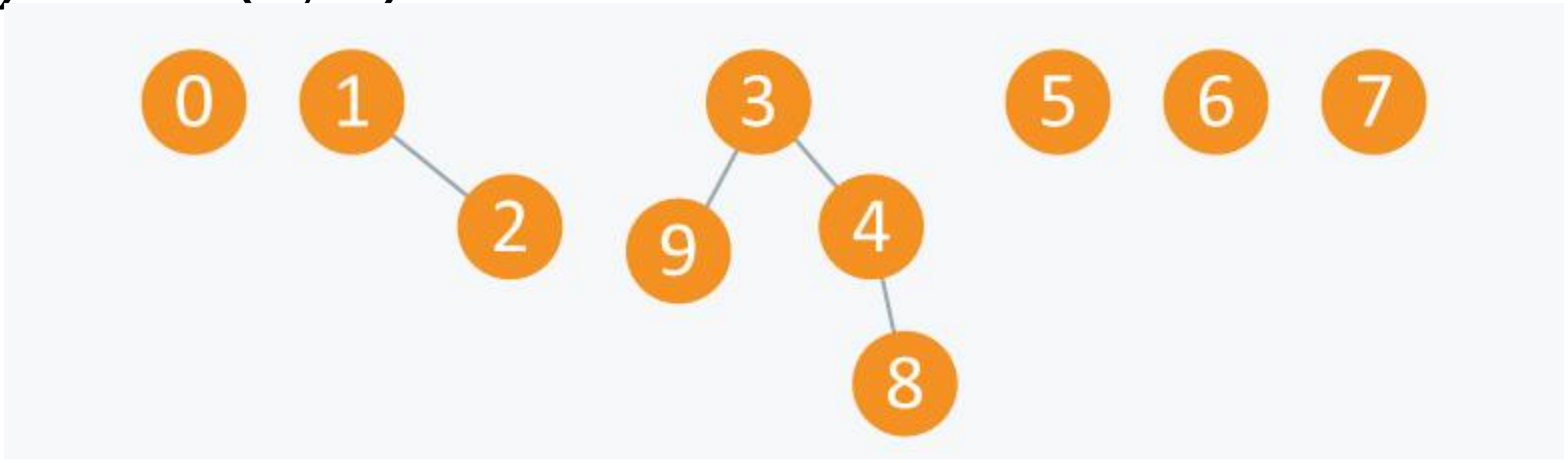
1) Union(2, 1)

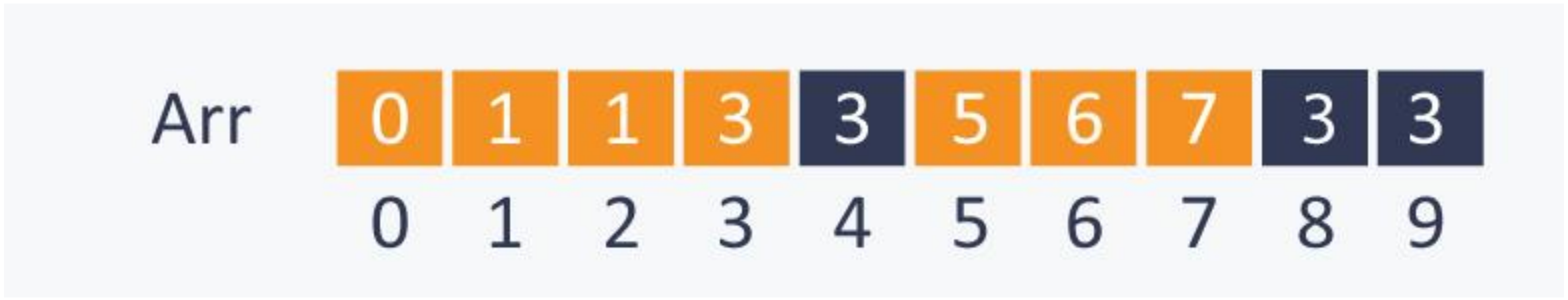




Ref: https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

2) Union(4, 3)
3) Union(8, 4)
4) Union(9, 3)

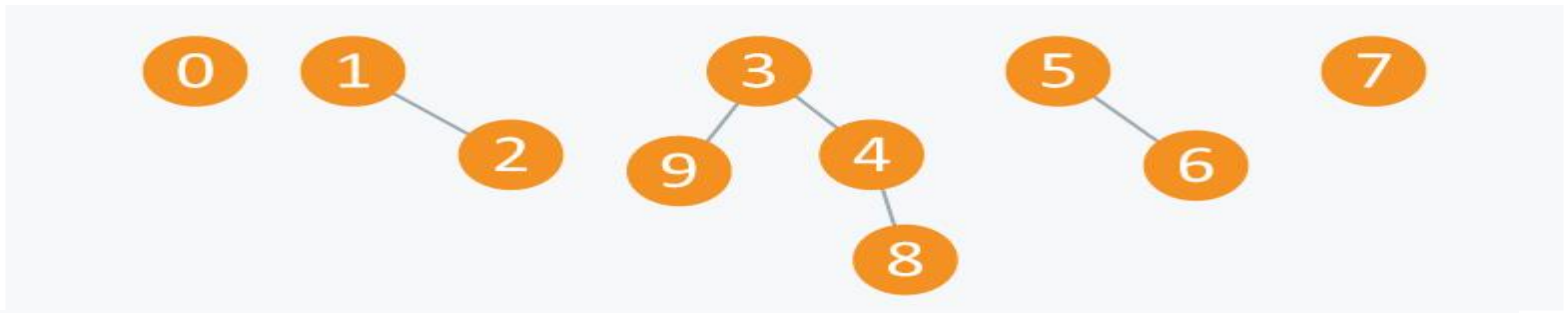


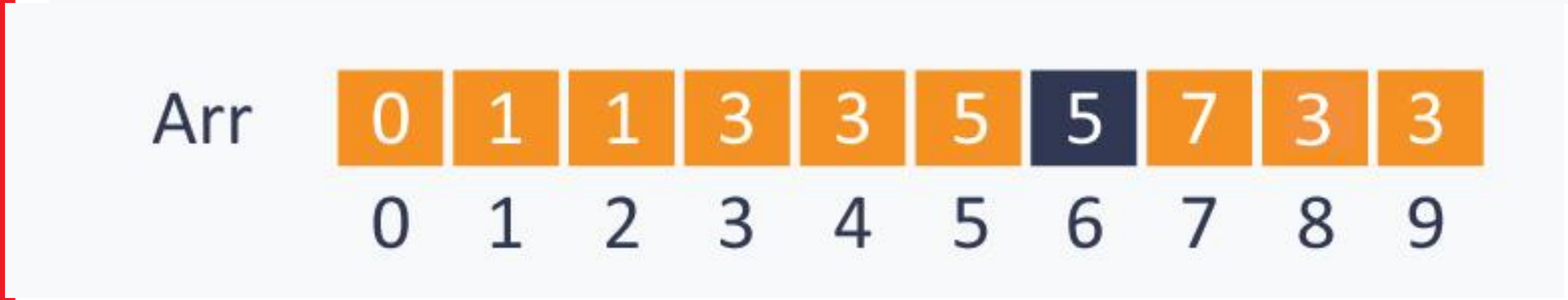


Arr	0	1	1	3	3	5	6	7	3	3
	0	1	2	3	4	5	6	7	8	9

Ref: https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

5) Union(6, 5)





Arr	0	1	1	3	3	5	5	7	3	3
	0	1	2	3	4	5	6	7	8	9

➔5 subsets.

A1= {3, 4, 8, 9},

A2= {1, 2},

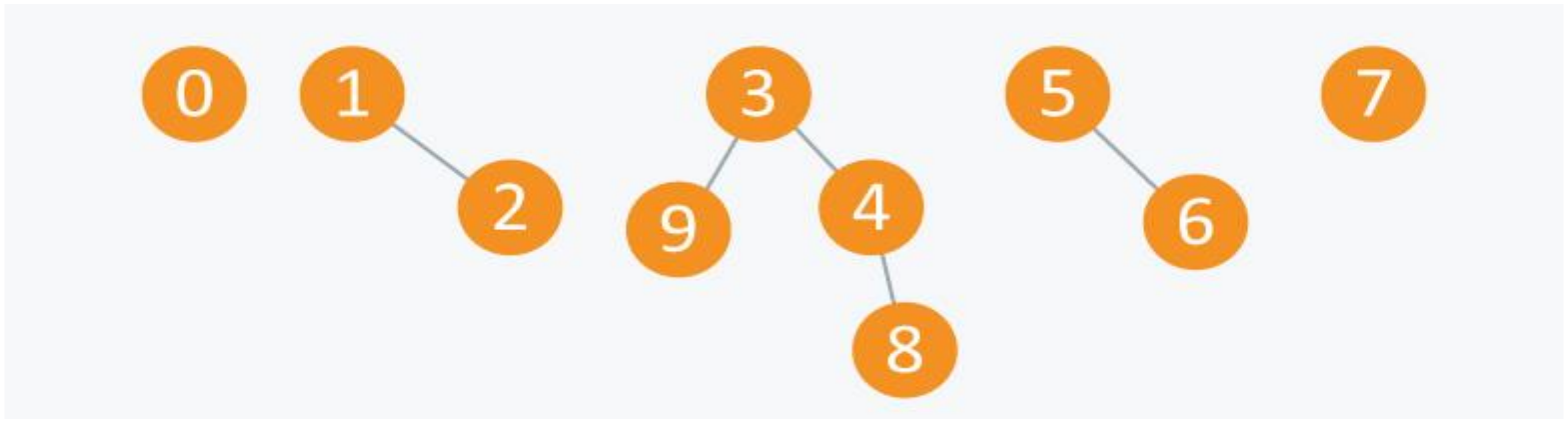
A3= {5, 6}

A4= {0}

A5 = {7}.

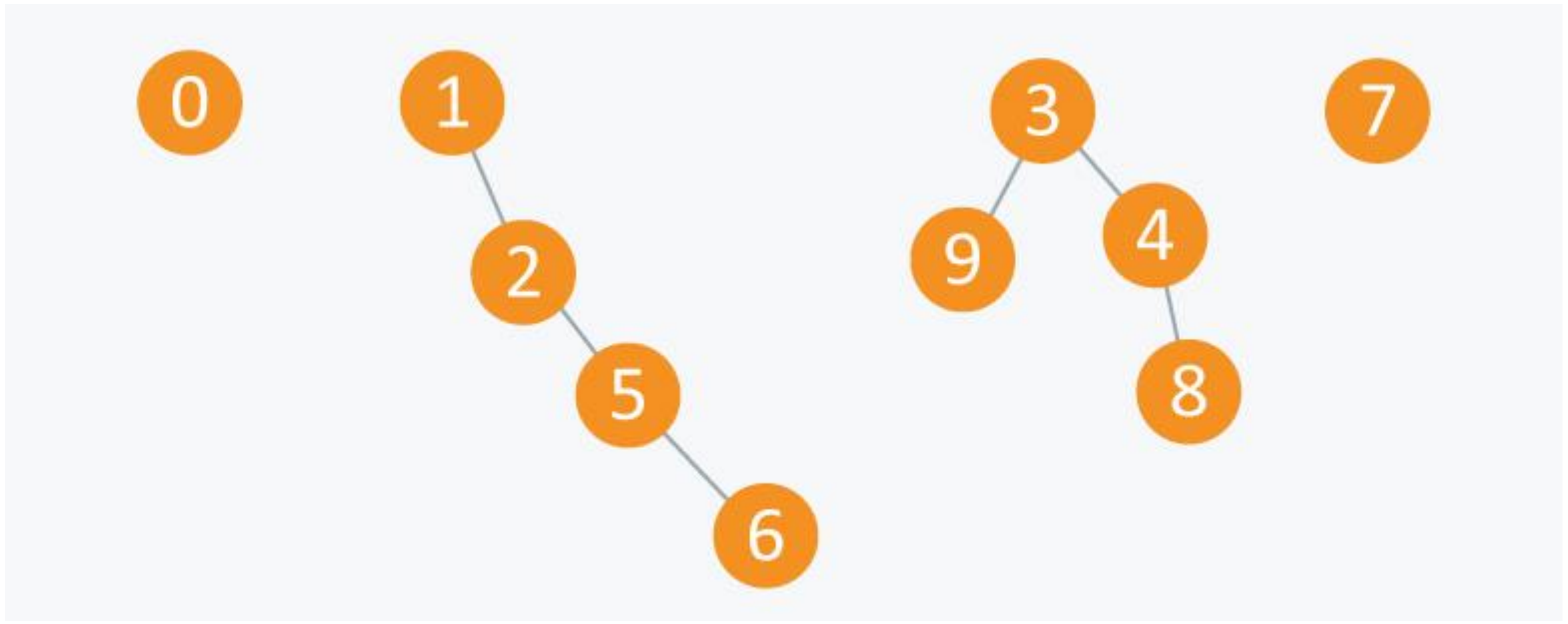
All these subsets are said to be Connected Components.

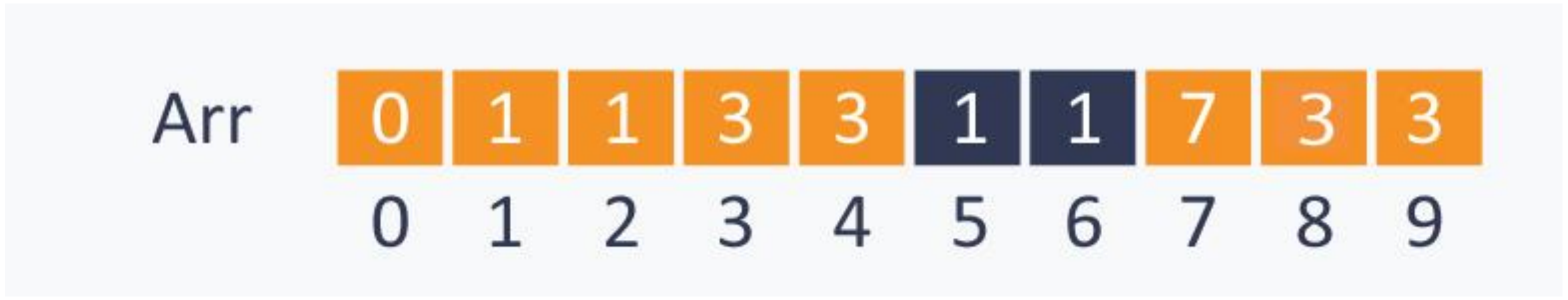
- Find (0, 7) = False as 0 and 7 are disconnected
- Find (8, 9) = True as 8 and 9 are connected directly or indirectly



Ref: https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

6) Union(5, 2)





Arr										
	0	1	1	3	3	1	1	7	3	3
	0	1	2	3	4	5	6	7	8	9

Ref: https://www.hackerearth.com/practice/notes/disjoint-set-union-union-find/

Applications of set partitioning

- Elections
- Divide and conquer
- Classification
- Pattern matching
- Mutually exclusive processes in OS
- Combinatorial explosion problem where repetition is not allowed

Applications of set partitioning

- commonly used in a variety of computer science applications, including algorithms, data analysis, and databases.
- The main advantage of using a set data structure is that it allows you to perform operations on a collection of elements in an efficient and organized way.

Maps

Also known as:

- table, search table, associative array, or associative container

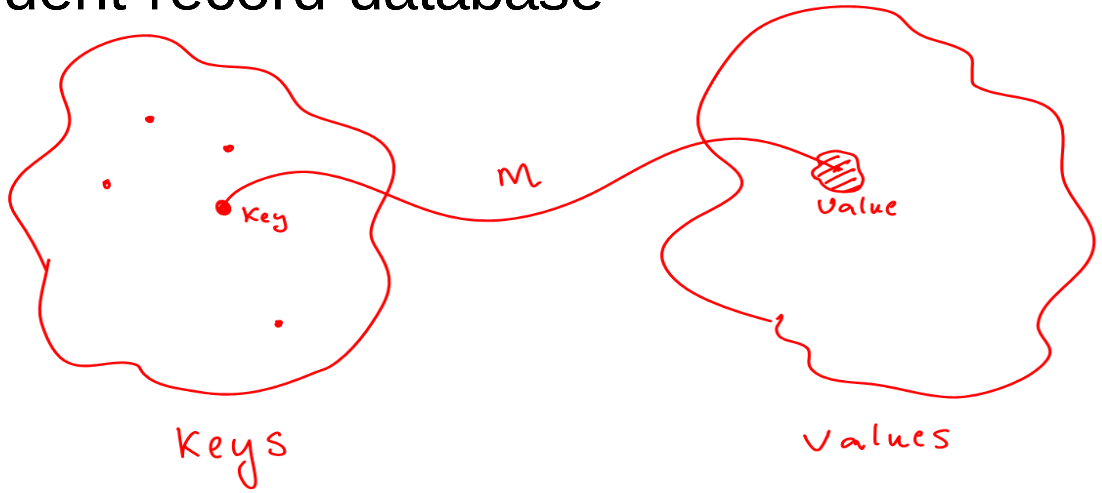
A data structure optimized for a very specific kind of search / access

- with a *bag* we access by asking "is X present"
- with a *list* we access by asking "give me item number X"
- with a *queue* we access by asking "give me the item that has been in the collection the longest."

In a *Map* we access by asking "give me the *value* associated with this *key."*

Maps

- A Map models a searchable collection of key-value entries
- The main operations of a map are for searching, inserting, and deleting items
- Multiple entries with the same key are **not** allowed
- Applications:
  - address book
  - student-record database



Maps

- A ***map*** allows to store elements so they can be located quickly using keys.
- ***key*** as a unique identifier
- A map stores key-value pairs (*k*,*v*), called ***entries***,
- each key is unique, so the association of keys to values defines a mapping.
- E.g. In a map storing student records (such as the student's name, address, and course grades), the key might be the student's ID number.
- Sometimes referred to as ***associative stores*** or ***associative containers***, as the key associated with an object determines its "location" in the data structure

Maps

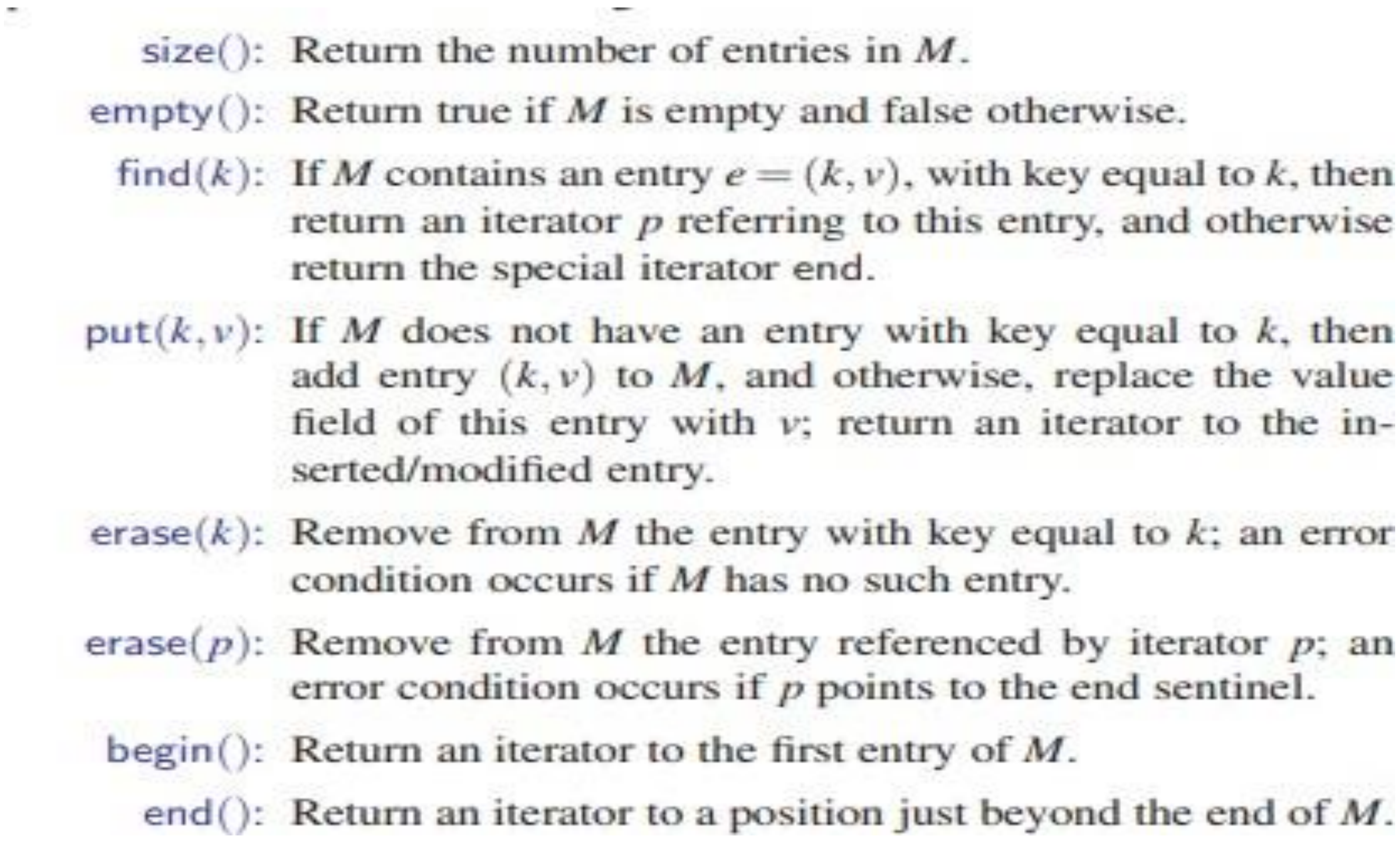
- Used where each key is to be viewed as a kind of unique ***index*** address for its value, that is,
- E.g. if we wish to store student records, we would probably want to use student ID objects as keys (and disallow two students having the same student ID).
- In other words, the key associated with an object can be viewed as an “address” for that object.

- ideal to use in look-up type situations where there is an identifying value and an actual value that is represented by the identifying value.
- examples :
  - Student ID numbers and last names.
  - House numbers on a street and the number of pets in each house

Map ADT

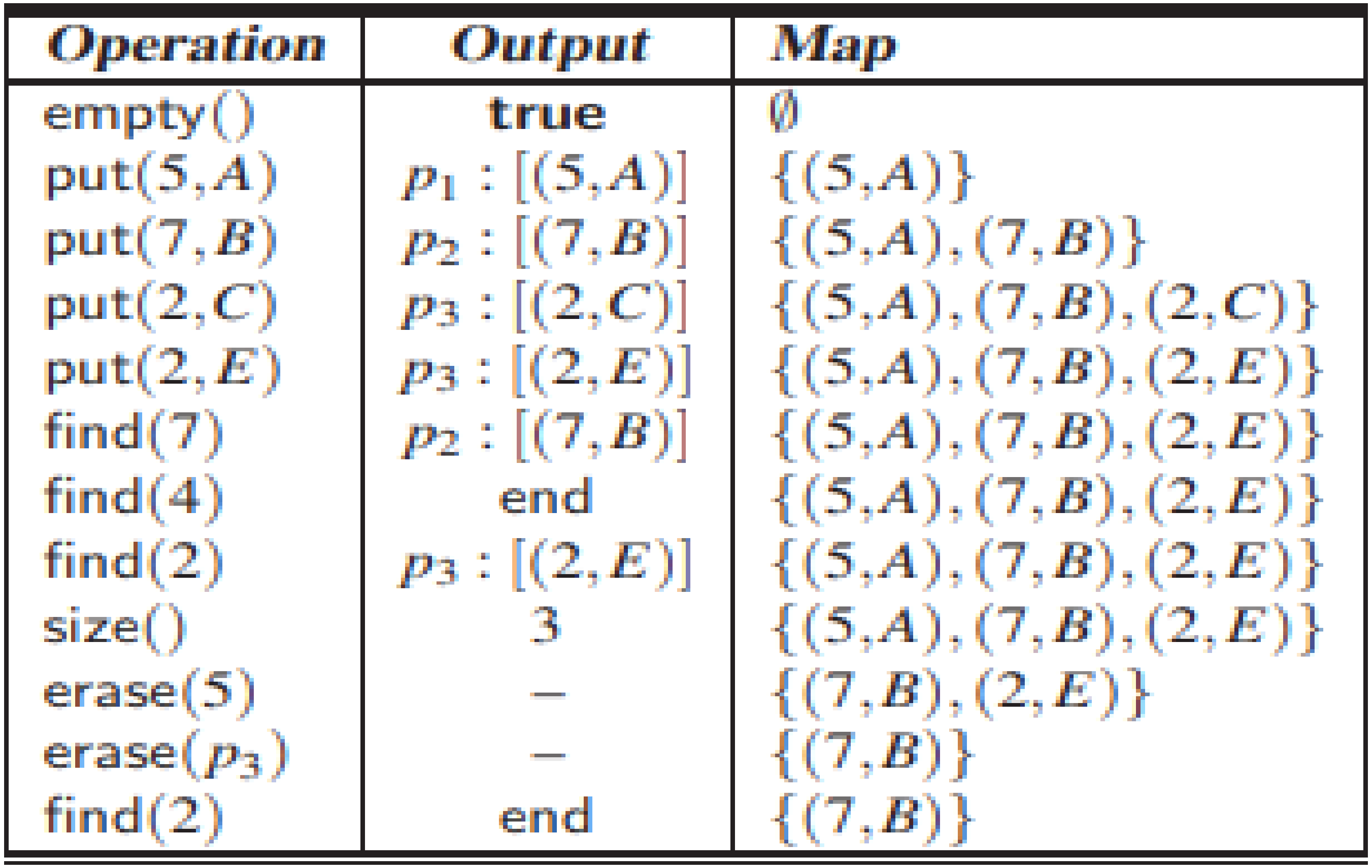
- Value definition: Map is a collection of key value entries, with each value associated with a distinct key.
- Assumption: map provides a special pointer object, which permits us to reference entries of the map, called ***position***.
- ***Iterator*** references entries and navigate around the map.
- Given a map iterator *p*, the associated entry may be accessed by dereferencing the iterator, **p*.
- The individual key and value can be accessed using *p*->key() and *p*->value(), respectively.
- Can be implemented using associative arrays

Map ADT



size():	Return the number of entries in M.
empty():	Return true if M is empty and false otherwise.
find(k):	If M contains an entry $e = (k, v)$, with key equal to k, then return an iterator p referring to this entry, and otherwise return the special iterator end.
put(k, v):	If M does not have an entry with key equal to k, then add entry (k, v) to M, and otherwise, replace the value field of this entry with v; return an iterator to the inserted/modified entry.
erase(k):	Remove from M the entry with key equal to k; an error condition occurs if M has no such entry.
erase(p):	Remove from M the entry referenced by iterator p; an error condition occurs if p points to the end sentinel.
begin():	Return an iterator to the first entry of M.
end():	Return an iterator to a position just beyond the end of M.

Example



Operation	Output	Map
empty()	true	$\emptyset$
put(5,A)	p_1 : [(5,A)]	{(5,A)}
put(7,B)	p_2 : [(7,B)]	{(5,A),(7,B)}
put(2,C)	p_3 : [(2,C)]	{(5,A),(7,B),(2,C)}
put(2,E)	p_3 : [(2,E)]	{(5,A),(7,B),(2,E)}
find(7)	p_2 : [(7,B)]	{(5,A),(7,B),(2,E)}
find(4)	end	{(5,A),(7,B),(2,E)}
find(2)	p_3 : [(2,E)]	{(5,A),(7,B),(2,E)}
size()	3	{(5,A),(7,B),(2,E)}
erase(5)	–	{(7,B),(2,E)}
erase(p_3)	–	{(7,B)}
find(2)	end	{(7,B)}

Reading Assignment

https://www.geeksforgeeks.org/map-associative-containers-the-c-standard-template-library-stl/

Some basic functions associated with Map:
begin() – Returns an iterator to the first element in the map
end() – Returns an iterator to the theoretical element that follows last element in the map
size() – Returns the number of elements in the map
max_size() – Returns the maximum number of elements that the map can hold
empty() – Returns whether the map is empty

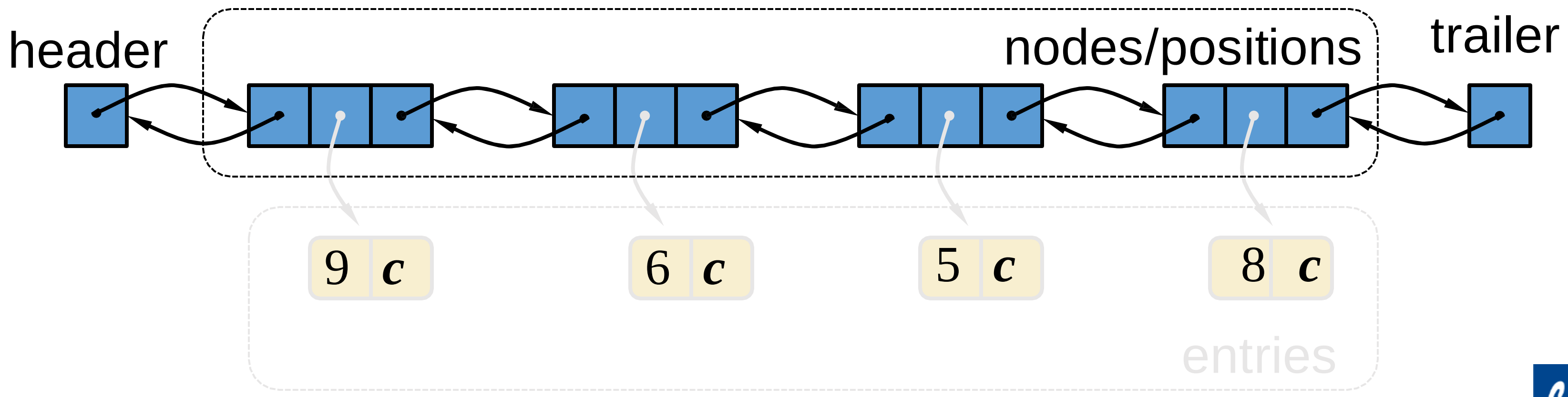
pair insert(keyvalue, mapvalue) – Adds a new element to the map
erase(iterator position) – Removes the element at the position pointed by the iterator
erase(const g)– Removes the key value ‘g’ from the map
clear() – Removes all the elements from the map

Map implementation

- Arrays
- A simple linked list of pairs
  - Slow (O(n)),
  - insufficient for general use.
- A hash table.
  - This is generally very fast (roughly O(1)),
  - Requires a good hash function for the key type.
- A binary search tree.
  - Fast (O(lg n)).
  - Unlike in a hash table, the keys will be ordered.
- A skip list.

A Simple List-Based Map

- We can efficiently implement a map using an unsorted list
  - We store the items of the map in a list S (based on a doubly-linked list), in arbitrary order



Hash-Based Map implementation

- Hash Map uses a hash table as its internal storage container.
- Keys stored based on hash codes and size of hash tables internal array

Tree-Based Map implementation

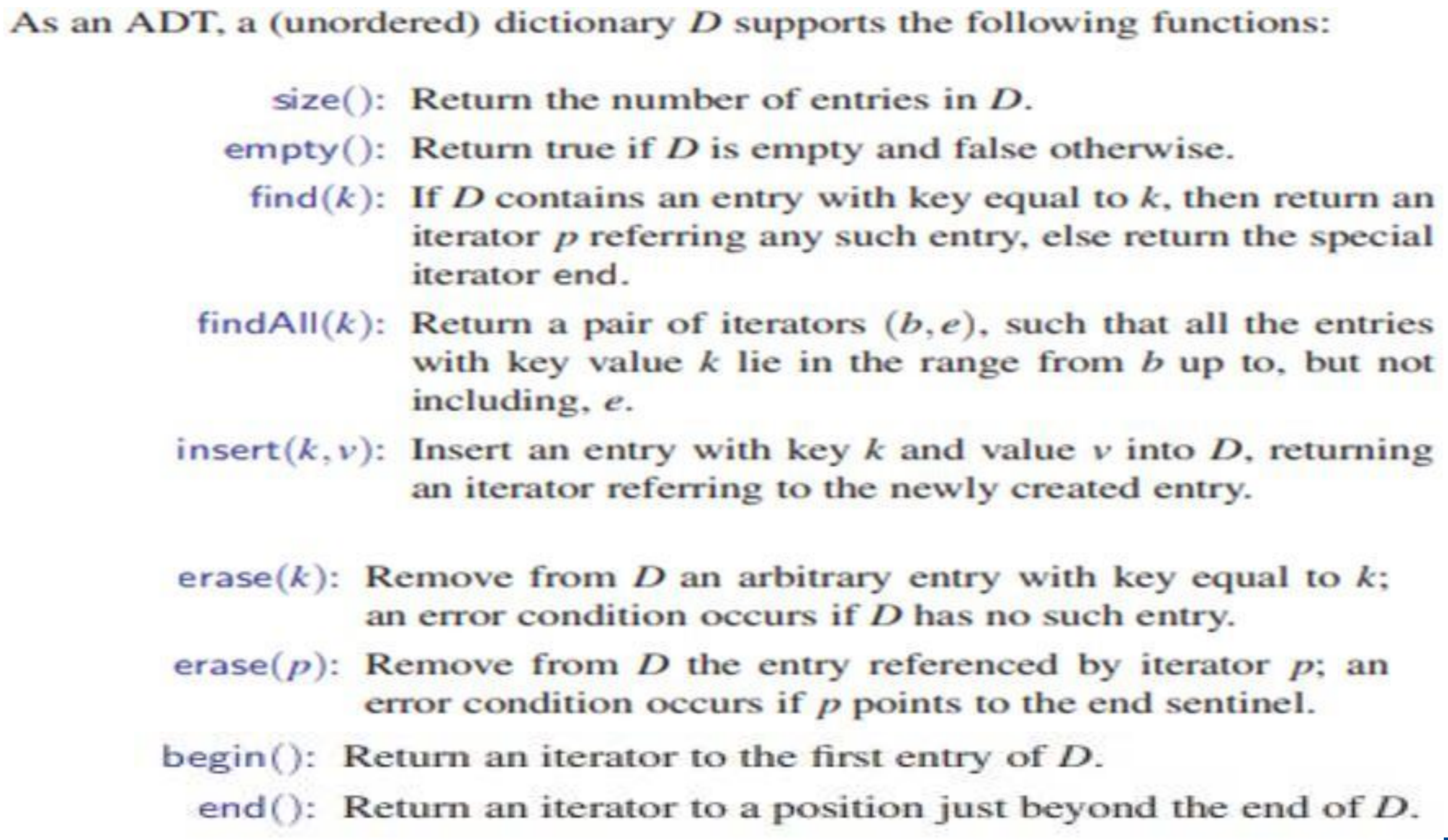
- Uses Height Balanced Binary Search Trees
- In java a Red - Black tree is used to implement a Map
- Somewhat slower than the HashMap

Dictionary

- A dictionary allows for keys and values to be of any object type.
- Unlike Map, a dictionary allows for multiple entries to have the same key
- For example
  - an English dictionary, which allows for multiple definitions for the same word.
  - we might want to store records for computer science authors indexed by their first and last names.
  - a multi-user computer game involving players visiting various rooms in a large castle might need a mapping from rooms to players. It is natural in this application to allow users to be in the same room simultaneously, however, to engage in battles.

Reading assignment:
Similarities and differences in set, map and dictionary

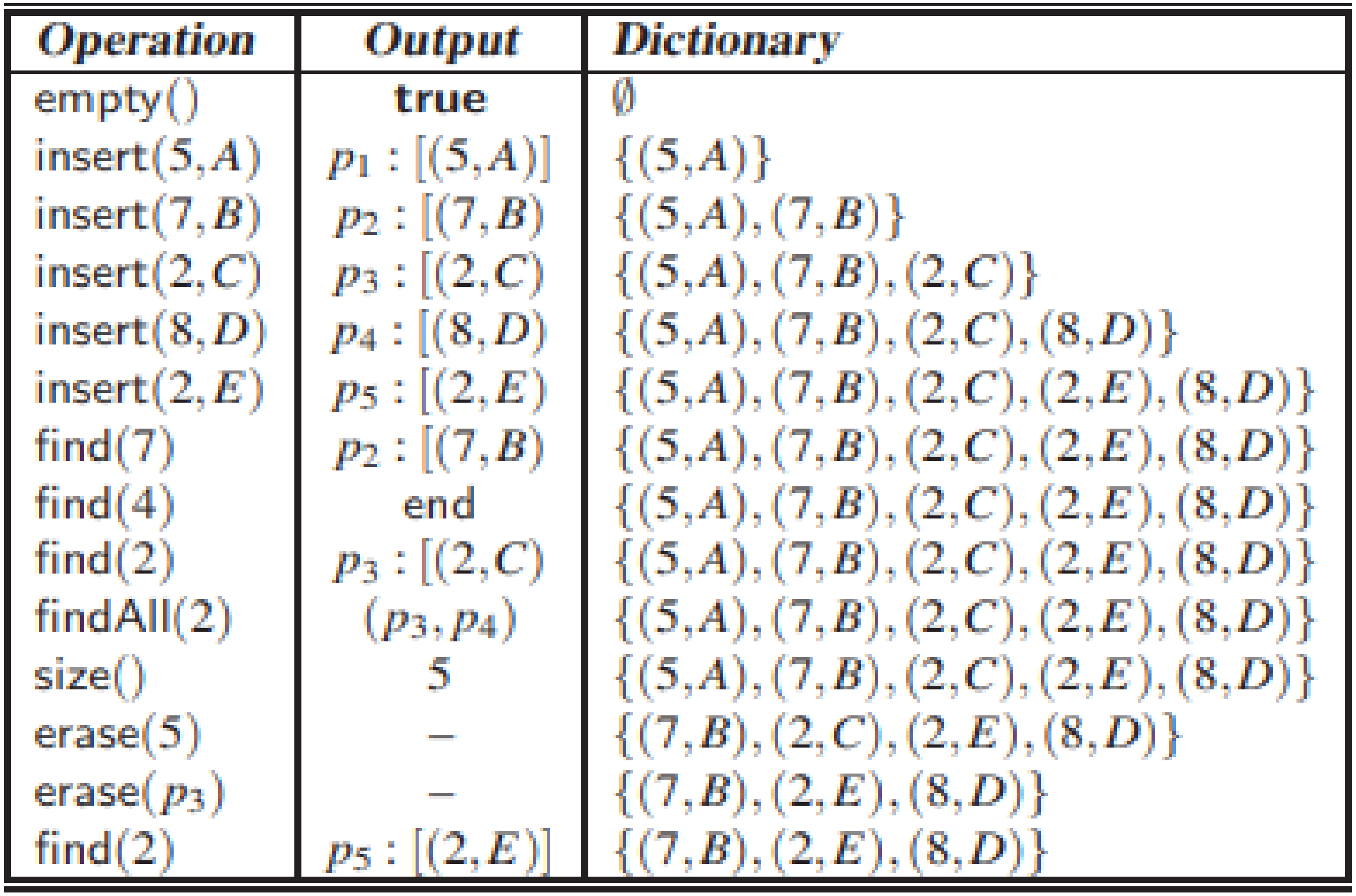
The Dictionary ADT



As an ADT, a (unordered) dictionary D supports the following functions:

size():	Return the number of entries in D.
empty():	Return true if D is empty and false otherwise.
find(k):	If D contains an entry with key equal to k, then return an iterator p referring any such entry, else return the special iterator end.
findAll(k):	Return a pair of iterators (b, e), such that all the entries with key value k lie in the range from b up to, but not including, e.
insert(k, v):	Insert an entry with key k and value v into D, returning an iterator referring to the newly created entry.
erase(k):	Remove from D an arbitrary entry with key equal to k; an error condition occurs if D has no such entry.
erase(p):	Remove from D the entry referenced by iterator p; an error condition occurs if p points to the end sentinel.
begin():	Return an iterator to the first entry of D.
end():	Return an iterator to a position just beyond the end of D.

Example



Operation	*Output*	*Dictionary*
empty()	true	$\emptyset$
insert$(5,A)$	$p_1 : [(5,A)]$	$\{(5,A)\}$
insert$(7,B)$	$p_2 : [(7,B)]$	$\{(5,A),(7,B)\}$
insert$(2,C)$	$p_3 : [(2,C)]$	$\{(5,A),(7,B),(2,C)\}$
insert$(8,D)$	$p_4 : [(8,D)]$	$\{(5,A),(7,B),(2,C),(8,D)\}$
insert$(2,E)$	$p_5 : [(2,E)]$	$\{(5,A),(7,B),(2,C),(2,E),(8,D)\}$
find(7)	$p_2 : [(7,B)]$	$\{(5,A),(7,B),(2,C),(2,E),(8,D)\}$
find(4)	end	$\{(5,A),(7,B),(2,C),(2,E),(8,D)\}$
find(2)	$p_3 : [(2,C)]$	$\{(5,A),(7,B),(2,C),(2,E),(8,D)\}$
findAll(2)	(p_3, p_4)	$\{(5,A),(7,B),(2,C),(2,E),(8,D)\}$
size()	5	$\{(5,A),(7,B),(2,C),(2,E),(8,D)\}$
erase(5)	–	$\{(7,B),(2,C),(2,E),(8,D)\}$
erase(p_3)	–	$\{(7,B),(2,E),(8,D)\}$
find(2)	$p_5 : [(2,E)]$	$\{(7,B),(2,E),(8,D)\}$

Dictionary Implementations

- **Unordered list:** In an unordered list, L, implementing a dictionary, we can maintain the location variable of each entry e to point to e’s position in the underlying linked list for L.
- **Hash table with separate chaining:** Consider a hash table, with bucket array A and hash function h, that uses separate chaining for handling collisions. We use the location variable of each entry e to point to e’s position in the list L implementing the list A[h(k)].
- **Ordered search table:** In an ordered table, T, implementing a dictionary, we should maintain the location variable of each entry e to be e’s index in T.

Example

Problem statement:

Given names and phone numbers, assemble a phone book that maps friends' names to their respective phone numbers. You will then be given an unknown number of names to query your phone book for. For each query, print the associated entry from your phone book on a new line in the form name=phoneNumber; if an entry for is not found, print Not found instead.

Example.. contd

Input Format
The first line contains an integer, , denoting the number of entries in the phone book.
Each of the subsequent lines describes an entry in the form of space-separated values on a single line. The first value is a friend's name, and the second value is an -digit phone number.
After the lines of phone book entries, there are *an unknown number of lines of queries*. Each line (query) contains a name to look up, and you must continue reading lines until there is no more input.
Output Format
On a new line for each query, print Not found if the name has no corresponding entry in the phone book; otherwise, print the full and in the format name=phoneNumber.

Solution?

- Sample dictionary entries?
- Sample queries?
- Output?

Example 2

You have a weather forecast data having temperature details of few cities for few days for the year 2018

Build data structure to answer the following queries

- What is Temperature in Delhi on 9-11-2018
- What is max temperature recorded in Chennai in 2018
- Displaying the number of entries of 2018
- Deleting the entry of Mumbai on 9-11-2018
- Deleting all entries of Mumbai

City	Date	Temperature
Delhi	9-11-2018	45
Bangalore	9-11-2018	24
Ranchi	9-12-2018	28
Chennai	9-01-2018	38