CSCB63 – Design and Analysis of Data Structures

Akshay Arun Bapat[1]

[1]based on notes by Anya Tafliovich, Anna Bretscher and Albert Lai

Disjoint sets

Operations:

- `make-set(`x`)`: create a set that contains x
- `find-set(`x`)`: return the set that contains x
- `union(`S_1`, `S_2`)`: merge sets S_1 and S_2, or
- `union(`x_1`, `x_2`)`: merge set that contains x_1 and set that contains x_2

Where have we seen this? Kruskal's algorithm

Linked lists implementation

- each set is a linked list
- x.set is a pointer to x's owning linked list
- `find-set(`x`)` is: follow pointer, $\Theta(1)$ time
- `union(`S_1`, `S_2`)` is merging two linked lists
- choose to always move the smaller list into the larger one

What is the **amortised** complexity of `union`?

Linked lists implementation: complexity

What is the **amortised** complexity of `union`?

Consider a sequence of k operations `make-set`, `find-set`, and `union`, with n operations `make-set`.

- the longest a list can be is: n elements
- operations `make-set` and `find-set` are: $\mathcal{O}(1)$ for a total of $\mathcal{O}(k)$
- operation `union` is:
  - in the best case: smaller list has one node: 1 update
  - in the worst case: smaller list has (almost) as many nodes as larger list
  - in the worst case: the size of list roughly doubles as a result
- then how many such updates can we do?
  - each x`.set` field is updated at most: $\log n$ times
  - there are n `.set` fields
  - total number of updates at most: $n \log n$

Linked lists implementation: complexity

What is the **amortised** complexity of `union`?

Consider a sequence of k operations `make-set`, `find-set`, and `union`, with n operations `make-set`.

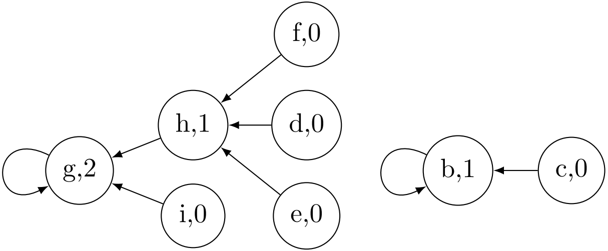
Total time at most:

$$k + n \log n \leq k + k \log n \in \mathcal{O}(k \log n)$$

Amortised time: $\mathcal{O}(\log n)$

Forest implementation

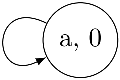
- each set is a tree
- pointers from children to parents
- root points to itself
- each node stores `rank`
    - an upper bound on the height of the tree rooted at that node



Forest implementation: `make-set`

- `make-set` creates a single-node tree

```
make-set(x):
0. root := new node(value=x, rank=0)
1. root.parent := root
2. return root
```



Forest implementation: `union`

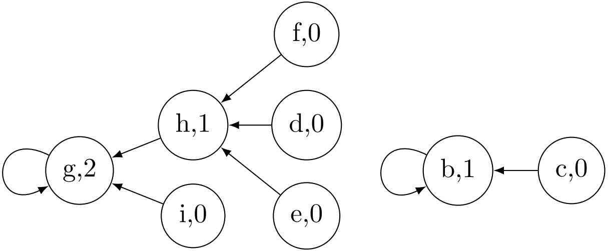
- `union` makes root of shorter tree a child of root of taller tree

```
union(node1, node2):
  link(find-set(node1), find-set(node2))

link(root1, root2):
  if root1.rank > root2.rank:
    root2.parent := root1
  else:
    root1.parent := root2
    if root1.rank = root2.rank:
      root2.rank++
```

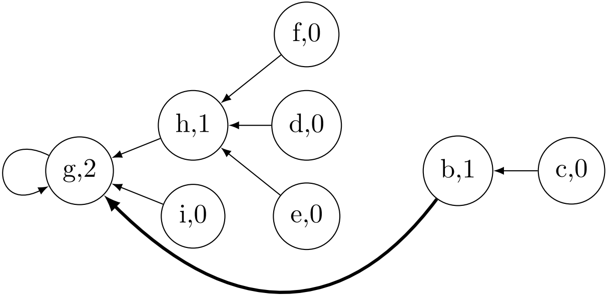
Forest implementation: `union`

- `union` makes root of shorter tree a child of root of taller tree



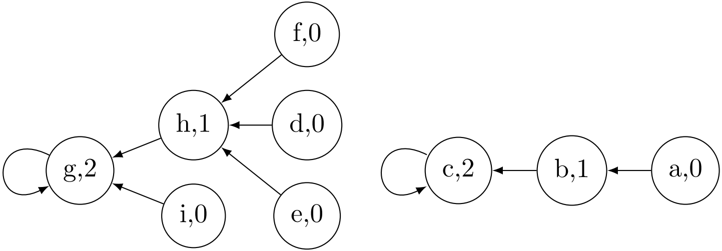
Forest implementation: `union`

- `union` makes root of shorter tree a child of root of taller tree



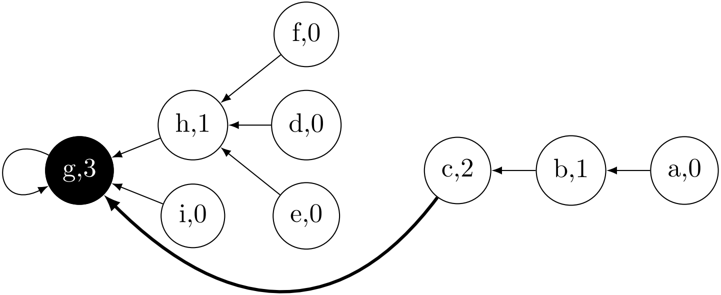
Forest implementation: `union`

- `union` makes root of shorter tree a child of root of taller tree



Forest implementation: `union`

- `union` makes root of shorter tree a child of root of taller tree

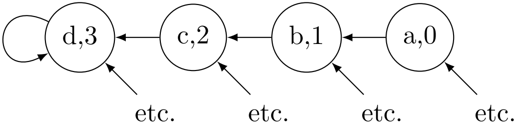


Called <u>union by rank</u>

Forest implementation: `find-set`

- `find-set` follows links to root

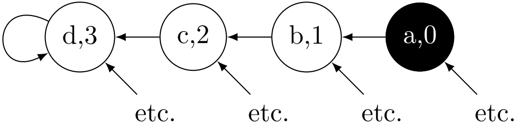
```
find-set(node):
0. if node.parent != node:
1.   return find-set(node.parent)
2. return node
```



Forest implementation: `find-set`

- `find-set` follows links to root

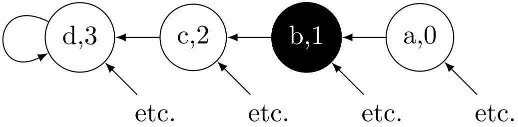
```
find-set(node):
0. if node.parent != node:
1.   return find-set(node.parent)
2. return node
```



Forest implementation: `find-set`

- `find-set` follows links to root

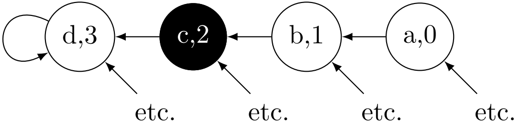
```
find-set(node):
0. if node.parent != node:
1.   return find-set(node.parent)
2. return node
```



Forest implementation: `find-set`

- `find-set` follows links to root

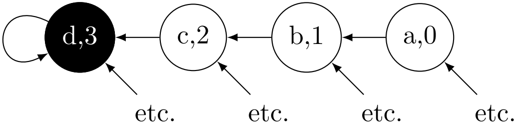
```
find-set(node):
0. if node.parent != node:
1.   return find-set(node.parent)
2. return node
```



Forest implementation: `find-set`

- `find-set` follows links to root

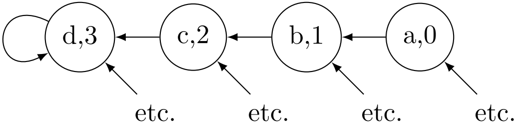
```
find-set(node):
0. if node.parent != node:
1.   return find-set(node.parent)
2. return node
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

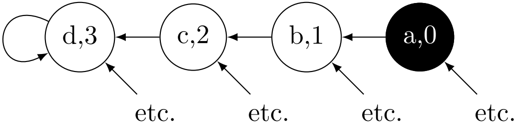
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

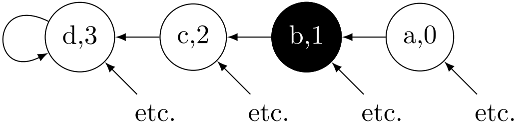
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- ranks are not updated

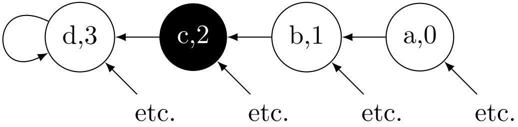
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

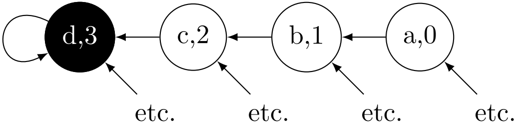
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

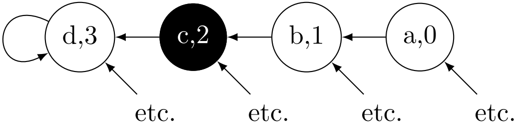
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

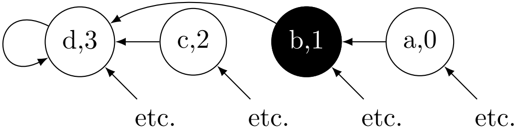
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

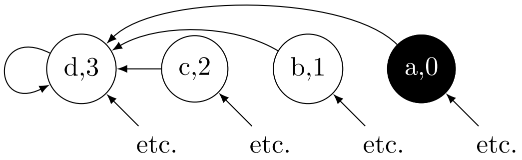
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

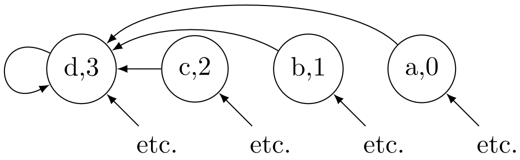
```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: `find-set`

- better: path compression
- `find-set` updates parent link directly to root
- `ranks` are not updated

```
find-set(node):
0. if node.parent != node:
1.   node.parent := find-set(node.parent)
2. return node.parent
```



Forest implementation: complexity

- The best disjoint set implementation is forests using union-by-rank and path compression.
- What is the worst-case sequence complexity?
- Can show worst-case time for a sequence of k operations with n `make-sets`, is $\mathcal{O}(k\alpha(n)) \in \mathcal{O}(k \log^* n)$
- Here $\log^* n$ is the number of times that you need to apply log to n until the answer is < 1
  - for example, if $n = 40$, then $1 < \log\log 40 < 2$ but $0 < \log\log\log 40 < 1$, so $\log^* 40 = 3$
- The function α grows very, very slowly, virtually a constant.
- Amortised time of disjoint set operations is $\mathcal{O}(\alpha(n))$.
- Full proof outside the scope of this course.
- Note this means the best implementation of Kruskal's algorithm has complexity $\mathcal{O}(|E|\log|E| + |E|\alpha(|V|)))$ time.