Union–Find

- naïve linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
Disjoint-sets data type

**Goal.** Support three operations on a collection of disjoint sets.

- **MAKE-SET(x):** create a new set containing only element \( x \).
- **FIND(x):** return a canonical element in the set containing \( x \).
- **UNION(x, y):** replace the sets containing \( x \) and \( y \) with their union.

**Performance parameters.**

- \( m = \) number of calls to **MAKE-SET**, **FIND**, and **UNION**.
- \( n = \) number of elements = number of calls to **MAKE-SET**.

**Dynamic connectivity.** Given an initially empty graph \( G \), support three operations.

- **ADD-NODE(u):** add node \( u \).
- **ADD-EDGE(u, v):** add an edge between nodes \( u \) and \( v \).
- **IS-CONNECTED(u, v):** is there a path between \( u \) and \( v \)?
Disjoint-sets data type: applications

Original motivation. Compiling EQUIVALENCE, DIMENSION, and COMMON statements in Fortran.

An Improved Equivalence Algorithm

Bernard A. Galler and Michael J. Fisher
University of Michigan, Ann Arbor, Michigan

An algorithm for assigning storage on the basis of EQUIVALENCE, DIMENSION and COMMON declarations is presented. The algorithm is based on a tree structure, and has reduced computation time by 40 percent over a previously published algorithm by identifying all equivalence classes with one scan of the EQUIVALENCE declarations. The method is applicable in any problem in which it is necessary to identify equivalence classes, given the element pairs defining the equivalence relation.

Note. This 1964 paper also introduced key data structure for problem.
Applications.

- Percolation.
- Kruskal’s algorithm.
- Connected components.
- Computing LCAs in trees.
- Computing dominators in digraphs.
- Equivalence of finite state automata.
- Checking flow graphs for reducibility.
- Hoshen–Kopelman algorithm in physics.
- Hinley–Milner polymorphic type inference.
- Morphological attribute openings and closings.
- Matlab’s BW-LABEL function for image processing.
- Compiling EQUIVALENCE, DIMENSION and COMMON statements in Fortran.
- ...

Disjoint-sets data type: applications
UNION–FIND

- naïve linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
Disjoint-sets data structure

**Parent-link representation.** Represent each set as a tree of elements.
- Each element has an explicit parent pointer in the tree.
- The root serves as the canonical element (and points to itself).
- \textsc{find}(x): find the root of the tree containing \( x \).
- \textsc{union}(x, y): merge trees containing \( x \) and \( y \).

\textsc{union}(3, 5)
Disjoint-sets data structure

Parent-link representation. Represent each set as a tree of elements.

- Each element has an explicit parent pointer in the tree.
- The root serves as the canonical element (and points to itself).
- \( \text{FIND}(x) \): find the root of the tree containing \( x \).
- \( \text{UNION}(x, y) \): merge trees containing \( x \) and \( y \).

**\( \text{UNION}(3, 5) \)**

![Diagram of a disjoint-sets data structure](image)
Disjoint-sets data structure

Array representation. Represent each set as a tree of elements.

- Allocate an array `parent[]` of length `n`. \(\text{must know number of elements } n \text{ a priori}\)
- \(\text{parent}[i] = j\) means parent of element \(i\) is element \(j\).

```
p
0 1 2 3 4 5 6 7 8 9
- parent[]
- 7 5 7 [8] 8 7 5 7 8 8
```

```
parent of 3 is 8
```

```
root
```

```
Note. For brevity, we suppress arrows and self loops in figures.
```
Naïve linking

**Naïve linking.** Link root of first tree to root of second tree.

**UNION(5, 3)**
Naïve linking

Naïve linking. Link root of first tree to root of second tree.

\texttt{Union}(5, 3)
Naïve linking

Naïve linking. Link root of first tree to root of second tree.

```
MAKE-SET(x)

parent[x] ← x.
```

```
UNION(x, y)

r ← FIND(x).
s ← FIND(y).
parent[r] ← s.
```

```
FIND(x)

WHILE (x ≠ parent[x])
    x ← parent[x].
RETURN x.
```
Naïve linking: analysis

**Theorem.** Using naïve linking, a **UNION** or **FIND** operation can take $\Theta(n)$ time in the worst case, where $n$ is the number of elements.

**Pf.**

- In the worst case, **FIND** takes time proportional to the **height** of the tree.
- Height of the tree is $n – 1$ after the sequence of union operations: **UNION(1, 2), UNION(2, 3), …, UNION(n – 1, n).**
Union–Find

- naïve linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
Link-by-size

**Link-by-size.** Maintain a tree size (number of nodes) for each root node. Link root of smaller tree to root of larger tree (breaking ties arbitrarily).

**`UNION(5, 3)`**

```
size = 4

8
9 4 3
```

```
size = 6

7
0 5 2
1 6
```
Link-by-size. Maintain a tree size (number of nodes) for each root node. Link root of smaller tree to root of larger tree (breaking ties arbitrarily).

\textbf{Union(5, 3)}

\begin{tikzpicture}[level distance=1.5cm,sibling distance=1.5cm]
  \node (root) {7}
    child {node (a) {8}
      child {node (b) {9}}
      child {node (c) {4}}
      child {node (d) {3}}
    }
    child {node (e) {0}
      child {node (f) {1}}
      child {node (g) {6}}
      child {node (h) {5}}
      child {node (i) {2}}
    }
  \end{tikzpicture}

\textit{size = 10}
Link-by-size

Link-by-size. Maintain a tree size (number of nodes) for each root node. Link root of smaller tree to root of larger tree (breaking ties arbitrarily).

**MAKE-SET**(x)

\[ \text{parent}[x] \leftarrow x. \]
\[ \text{size}[x] \leftarrow 1. \]

**FIND**(x)

\[ \textbf{WHILE} \ (x \neq \text{parent}[x]) \]
\[ \quad x \leftarrow \text{parent}[x]. \]
\[ \textbf{RETURN} \ x. \]

**UNION**(x, y)

\[ r \leftarrow \text{FIND}(x). \]
\[ s \leftarrow \text{FIND}(y). \]
\[ \textbf{IF} \ (r = s) \ \textbf{RETURN}. \]
\[ \textbf{ELSE IF} \ (\text{size}[r] > \text{size}[s]) \]
\[ \quad \text{parent}[s] \leftarrow r. \]
\[ \quad \text{size}[r] \leftarrow \text{size}[r] + \text{size}[s]. \]
\[ \textbf{ELSE} \]
\[ \quad \text{parent}[r] \leftarrow s. \]
\[ \quad \text{size}[s] \leftarrow \text{size}[r] + \text{size}[s]. \]
Link-by-size: analysis

**Property.** Using link-by-size, for every root node $r$: $size[r] \geq 2^{height(r)}$.

**Pf.** [by induction on number of links]
- Base case: singleton tree has size 1 and height 0.
- Inductive hypothesis: assume true after first $i$ links.
- Tree rooted at $r$ changes only when a smaller (or equal) size tree rooted at $s$ is linked into $r$.
- Case 1. [$height(r) > height(s)$] $size'[r] > size[r]$

\[
\geq 2^{height(r)} \quad \text{inductive hypothesis}
\]

\[
= 2^{height'(r)}.
\]
**Property.** Using link-by-size, for every root node $r$: $\text{size}[r] \geq 2^{\text{height}(r)}$.

**Pf.** [ by induction on number of links ]

- **Base case:** singleton tree has size 1 and height 0.
- **Inductive hypothesis:** assume true after first $i$ links.
- **Tree rooted at $r$ changes only when a smaller (or equal) size tree rooted at $s$ is linked into $r$.**
- **Case 2.** [$\text{height}(r) \leq \text{height}(s)$] $\text{size}'[r] = \text{size}[r] + \text{size}[s]$
  
  \[ \geq 2 \cdot \text{size}[s] \quad \text{link-by-size} \]
  
  \[ \geq 2 \cdot 2^{\text{height}(s)} \quad \text{inductive hypothesis} \]
  
  \[ = 2^{\text{height}(s) + 1} \]
  
  \[ = 2^{\text{height}(r)}. \]
**Theorem.** Using link-by-size, any UNION or FIND operation takes $O(\log n)$ time in the worst case, where $n$ is the number of elements.

**Pf.**

- The running time of each operation is bounded by the tree height.
- By the previous property, the height is $\leq \lfloor \log n \rfloor$. □

\[ \log n = \log_2 n \]

**Note.** The UNION operation takes $O(1)$ time except for its two calls to FIND.
A tight upper bound

**Theorem.** Using link-by-size, a tree with $n$ nodes can have height $= \lg n$.

**Pf.**

- Arrange $2^k - 1$ calls to UNION to form a binomial tree of order $k$.
- An order-$k$ binomial tree has $2^k$ nodes and height $k$. □
Section 5.1.4

Union-Find

- naïve linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
Link-by-rank. Maintain an integer rank for each node, initially 0. Link root of smaller rank to root of larger rank; if tie, increase rank of new root by 1.

**Union(7, 3)**

Note. For now, rank = height.
Link-by-rank. Maintain an integer rank for each node, initially 0. Link root of smaller rank to root of larger rank; if tie, increase rank of new root by 1.

Note. For now, rank = height.
Link-by-rank. Maintain an integer rank for each node, initially 0. Link root of smaller rank to root of larger rank; if tie, increase rank of new root by 1.

\[
\text{MAKE-SET}(x) \\
\text{parent}[x] \leftarrow x. \\
\text{rank}[x] \leftarrow 0. \\
\]

\[
\text{FIND}(x) \\
\text{WHILE} (x \neq \text{parent}[x]) \\
\quad x \leftarrow \text{parent}[x]. \\
\text{RETURN} x. \\
\]

\[
\text{UNION}(x, y) \\
\text{r} \leftarrow \text{FIND}(x). \\
\text{s} \leftarrow \text{FIND}(y). \\
\text{IF} \ (r = s) \ \text{RETURN}. \\
\text{ELSE IF} \ (\text{rank}[r] > \text{rank}[s]) \\
\quad \text{parent}[s] \leftarrow r. \\
\text{ELSE IF} \ (\text{rank}[r] < \text{rank}[s]) \\
\quad \text{parent}[r] \leftarrow s. \\
\text{ELSE} \\
\quad \text{parent}[r] \leftarrow s. \\
\quad \text{rank}[s] \leftarrow \text{rank}[s] + 1. \\
\]
**Link-by-rank: properties**

**Property 1.** If $x$ is not a root node, then $\text{rank}[x] < \text{rank}[\text{parent}[x]]$.

**Proof.** A node of rank $k$ is created only by linking two roots of rank $k - 1$. □

**Property 2.** If $x$ is not a root node, then $\text{rank}[x]$ will never change again.

**Proof.** Rank changes only for roots; a nonroot never becomes a root. □

**Property 3.** If $\text{parent}[x]$ changes, then $\text{rank}[\text{parent}[x]]$ strictly increases.

**Proof.** The parent can change only for a root, so before linking $\text{parent}[x] = x$. After $x$ is linked-by-rank to new root $r$ we have $\text{rank}[r] > \text{rank}[x]$. □
Link-by-rank: properties

**PROPERTY 4.** Any root node of rank \( k \) has \( \geq 2^k \) nodes in its tree.

**Pf.** [ by induction on \( k \) ]
- Base case: true for \( k = 0 \).
- Inductive hypothesis: assume true for \( k - 1 \).
- A node of rank \( k \) is created only by linking two roots of rank \( k - 1 \).
- By inductive hypothesis, each subtree has \( \geq 2^{k-1} \) nodes
  \( \Rightarrow \) resulting tree has \( \geq 2^k \) nodes.

**PROPERTY 5.** The highest rank of a node is \( \leq \lfloor \lg n \rfloor \).

**Pf.** Immediate from **PROPERTY 1** and **PROPERTY 4**.
**Property 6.** For any integer $k \geq 0$, there are $\leq n/2^k$ nodes with rank $k$.

**Pf.**

- Any root node of rank $k$ has $\geq 2^k$ descendants. [Property 4]
- Any nonroot node of rank $k$ has $\geq 2^k$ descendants because:
  - it had this property just before it became a nonroot [Property 4]
  - its rank doesn’t change once it became a nonroot [Property 2]
  - its set of descendants doesn’t change once it became a nonroot
- Different nodes of rank $k$ can’t have common descendants. [Property 1]
Link-by-rank: analysis

**Theorem.** Using link-by-rank, any **UNION** or **FIND** operation takes $O(\log n)$ time in the worst case, where $n$ is the number of elements.

**Pf.**

- The running time of **UNION** and **FIND** is bounded by the tree height.
- By **PROPERTY 5**, the height is $\leq \lfloor \lg n \rfloor$. □
**Union–Find**

- naïve linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$. 
Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$. 
Path compression

Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$. 
Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$. 
Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$. 
Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$. 
Path compression

Path compression. When finding the root $r$ of the tree containing $x$, change the parent pointer of all nodes along the path to point directly to $r$.

\[
\text{FIND}(x) \\
\textbf{IF} (x \neq \text{parent}[x]) \\
\text{parent}[x] \leftarrow \text{FIND}(\text{parent}[x]) \\
\text{RETURN} \text{parent}[x].
\]

Note. Path compression does not change the rank of a node; so $\text{height}(x) \leq \text{rank}[x]$ but they are not necessarily equal.
Path compression

**Fact.** Path compression with naïve linking can require $\Omega(n)$ time to perform a single UNION or FIND operation, where $n$ is the number of elements.

**Pf.** The height of the tree is $n – 1$ after the sequence of union operations:

$\text{UNION}(1, 2), \text{UNION}(2, 3), \ldots, \text{UNION}(n – 1, n)$. □

naïve linking: link root of first tree to root of second tree

**Theorem.** [Tarjan–van Leeuwen 1984] Starting from an empty data structure, path compression with naïve linking performs any intermixed sequence of $m \geq n$ MAKE-SET, UNION, and FIND operations on a set of $n$ elements in $O(m \log n)$ time.

**Pf.** Nontrivial (but omitted).
**Section 5.1.4**

**Union-Find**

- naive linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
**PROPERTY.** The tree roots, node ranks, and elements within a tree are the same with or without path compression.

**Pf.** Path compression does not create new roots, change ranks, or move elements from one tree to another. □
Link-by-rank with path compression: properties

**Property.** The tree roots, node ranks, and elements within a tree are the same with or without path compression.

**Corollary.** Property 2, 4–6 hold for link-by-rank with path compression.

**Property 1.** If $x$ is not a root node, then $\text{rank}[x] < \text{rank}[\text{parent}[x]]$.

**Property 2.** If $x$ is not a root node, then $\text{rank}[x]$ will never change again.

**Property 3.** If $\text{parent}[x]$ changes, then $\text{rank}[\text{parent}[x]]$ strictly increases.

**Property 4.** Any root node of rank $k$ has $\geq 2^k$ nodes in its tree.

**Property 5.** The highest rank of a node is $\leq \lfloor \lg n \rfloor$.

**Property 6.** For any integer $k \geq 0$, there are $\leq n / 2^k$ nodes with rank $k$.

**Bottom line.** Property 1–6 hold for link-by-rank with path compression. (but we need to recheck Property 1 and Property 3)
Link-by-rank with path compression: properties

**Property 3.** If $parent[x]$ changes, then $rank[parent[x]]$ strictly increases.

**Pf.** Path compression can make $x$ point to only an ancestor of $parent[x]$.

**Property 1.** If $x$ is not a root node, then $rank[x] < rank[parent[x]]$.

**Pf.** Path compression doesn’t change any ranks, but it can change parents. If $parent[x]$ doesn’t change during a path compression, the inequality continues to hold; if $parent[x]$ changes, then $rank[parent[x]]$ strictly increases.
Iterated logarithm function

**Def.** The *iterated logarithm* function is defined by:

$$\lg^* n = \begin{cases} 
0 & \text{if } n \leq 1 \\
1 + \lg^*(\lg n) & \text{otherwise}
\end{cases}$$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\lg^* n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>[3, 4]</td>
<td>2</td>
</tr>
<tr>
<td>[5, 16]</td>
<td>3</td>
</tr>
<tr>
<td>[17, 65536]</td>
<td>4</td>
</tr>
<tr>
<td>[65537, 2^{65536}]</td>
<td>5</td>
</tr>
</tbody>
</table>

**Note.** We have $\lg^* n \leq 5$ unless $n$ exceeds the # atoms in the universe.
Analysis

Divide nonzero ranks into the following groups:

• \{ 1 \}
• \{ 2 \}
• \{ 3, 4 \}
• \{ 5, 6, ..., 16 \}
• \{ 17, 18, ..., 2^{16} \}
• \{ 65537, 65538, ..., 2^{65536} \}
• ...

Property 7. Every nonzero rank falls within one of the first $\lg^* n$ groups.

Pf. The rank is between 0 and $[\lg n]$. [PROPERTY 5]
Creative accounting

Credits. A node receives credits as soon as it ceases to be a root. If its rank is in the interval \( \{ k+1, k+2, \ldots, 2^k \} \), we give it \( 2^k \) credits.

Proposition. Number of credits disbursed to all nodes is \( \leq n \lg^* n \).

Pf.

• By Property 6, the number of nodes with rank \( \geq k + 1 \) is at most

\[
\frac{n}{2^{k+1}} + \frac{n}{2^{k+2}} + \cdots \leq \frac{n}{2^k}
\]

• Thus, nodes in group \( k \) need at most \( n \) credits in total.

• There are \( \leq \lg^* n \) groups. [Property 7]
Running time of FIND

Running time of FIND. Bounded by number of parent pointers followed.

- Recall: the rank strictly increases as you go up a tree. [PROPERTY 1]
- Case 0: $parent[x]$ is a root ⇒ only happens for one link per FIND.
- Case 1: $rank[parent[x]]$ is in a higher group than $rank[x]$.
- Case 2: $rank[parent[x]]$ is in the same group as $rank[x]$.

Case 1. At most $\lg^* n$ nodes on path can be in a higher group. [PROPERTY 7]

Case 2. These nodes are charged 1 credit to follow parent pointer.

- Each time $x$ pays 1 credit, $rank[parent[x]]$ strictly increases. [PROPERTY 1]
- Therefore, if $rank[x]$ is in the group $\{k + 1, \ldots, 2^k\}$, the rank of its parent will be in a higher group before $x$ pays $2^k$ credits.
- Once $rank[parent[x]]$ is in a higher group than $rank[x]$, it remains so because:
  - $rank[x]$ does not change once it ceases to be a root. [PROPERTY 2]
  - $rank[parent[x]]$ does not decrease. [PROPERTY 3]
  - thus, $x$ has enough credits to pay until it becomes a Case 1 node.
Link-by-rank with path compression

**Theorem.** Starting from an empty data structure, link-by-rank with path compression performs any intermixed sequence of $m \geq n$ MAKE-SET, UNION, and FIND operations on a set of $n$ elements in $O(m \log^* n)$ time.
UNION–FIND

- naïve linking
- link-by-size
- link-by-rank
- path compression
- link-by-rank with path compression
- context
Link-by-size with path compression

**Theorem.** [Fischer 1972] Starting from an empty data structure, link-by-size with path compression performs any intermixed sequence of $m \geq n$ MAKE-SET, UNION, and FIND operations on a set of $n$ elements in $O(m \log \log n)$ time.
Link-by-size with path compression

**Theorem.** [Hopcroft–Ullman 1973] Starting from an empty data structure, link-by-size with path compression performs any intermixed sequence of $m \geq n$ MAKE-SET, UNION, and FIND operations on a set of $n$ elements in $O(m \log^* n)$ time.
Link-by-size with path compression

**Theorem.** [Tarjan 1975] Starting from an empty data structure, link-by-size with path compression performs any intermixed sequence of \( m \geq n \) **MAKE-SET**, **UNION**, and **FIND** operations on a set of \( n \) elements in \( O(m \alpha(m, n)) \) time, where \( \alpha(m, n) \) is a functional inverse of the Ackermann function.

---

**Efficiency of a Good But Not Linear Set Union Algorithm**

ROBERT ENDRE TARJAN

University of California, Berkeley, California

**Abstract.** Two types of instructions for manipulating a family of disjoint sets which partition a universe of \( n \) elements are considered: **FIND**(\( x \)) computes the name of the (unique) set containing element \( x \); **UNION**(\( A, B, C \)) combines sets \( A \) and \( B \) into a new set named \( C \). A known algorithm for implementing sequences of these instructions is examined. It is shown that, if \( t(m, n) \) is the maximum time required by a sequence of \( m \geq n \) **FIND**s and \( n - 1 \) intermixed **UNION**s, then \( k_1m\alpha(m, n) \leq t(m, n) \leq k_2m\alpha(m, n) \) for some positive constants \( k_1 \) and \( k_2 \), where \( \alpha(m, n) \) is related to a functional inverse of Ackermann’s function and is very slow-growing.
Ackermann function

**Ackermann function.** [Ackermann 1928] A computable function that is not primitive recursive.

\[
A(m, n) = \begin{cases} 
  n + 1 & \text{if } m = 0 \\
  A(m - 1, 1) & \text{if } m > 0 \text{ and } n = 0 \\
  A(m - 1, A(m, n - 1)) & \text{if } m > 0 \text{ and } n > 0
\end{cases}
\]

**Note.** There are many inequivalent definitions.
**Ackermann function.** [Ackermann 1928] A computable function that is not primitive recursive.

\[
A(m, n) = \begin{cases} 
  n + 1 & \text{if } m = 0 \\
  A(m - 1, 1) & \text{if } m > 0 \text{ and } n = 0 \\
  A(m - 1, A(m, n - 1)) & \text{if } m > 0 \text{ and } n > 0 
\end{cases}
\]

**Inverse Ackermann function.**

\[
\alpha(m, n) = \min\{i \geq 1 : A(i, \lfloor m/n \rfloor) \geq \log_2 n\}
\]

"I am not smart enough to understand this easily."

— Raymond Seidel
Inverse Ackermann function

**Definition.**

\[
\alpha_k(n) = \begin{cases} 
\lfloor n/2 \rfloor & \text{if } k = 1 \\
0 & \text{if } n = 1 \text{ and } k \geq 2 \\
1 + \alpha_k(\alpha_{k-1}(n)) & \text{otherwise}
\end{cases}
\]

**Ex.**

- \(\alpha_1(n) = \lfloor n / 2 \rfloor\).
- \(\alpha_2(n) = \lfloor \lg n \rfloor = \# \text{ of times we divide } n \text{ by } 2, \text{ until we reach } 1\).
- \(\alpha_3(n) = \lg^* n = \# \text{ of times we apply the } \lg \text{ function to } n, \text{ until we reach } 1\).
- \(\alpha_4(n) = \# \text{ of times we apply the } \text{iterated } \lg \text{ function to } n, \text{ until we reach } 1\).

\[
2 \uparrow^{65536} = 2^{2^{2^{\cdot^{2}}} \cdot 2^{65536} \cdot 2} \quad \text{65536 times}
\]

|      | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | ... | 2^{16} | ... | 2^{65536} | ... | 2 \uparrow^{65536} |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| \(\alpha_1(n)\) | 1   | 1   | 2   | 2   | 3   | 3   | 4   | 4   | 5   | 5   | 5   | 6   | 6   | 7   | 7   | 8   | 8   | ... | 2^{15} | ... | 2^{65535} | ... | huge |
| \(\alpha_2(n)\) | 0   | 1   | 2   | 2   | 3   | 3   | 3   | 3   | 4   | 4   | 4   | 4   | 4   | 4   | 4   | 4   | 4   | ... | 16   | ... | 65536   | ... | 2 \uparrow^{65535} |
| \(\alpha_3(n)\) | 0   | 1   | 2   | 2   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | ... | 4    | ... | 5     | ... | 65536   |
| \(\alpha_4(n)\) | 0   | 1   | 2   | 2   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | 3   | ... | 3    | ... | 3     | ... | 4     |
Inverse Ackermann function

Definition. \[ \alpha_k(n) = \begin{cases} 
\lceil n/2 \rceil & \text{if } k = 1 \\
0 & \text{if } n = 1 \text{ and } k \geq 2 \\
1 + \alpha_k(\alpha_{k-1}(n)) & \text{otherwise}
\end{cases} \]

Property. For every \( n \geq 5 \), the sequence \( \alpha_1(n), \alpha_2(n), \alpha_3(n), \ldots \) converges to 3.

Ex. \( [n = 9876!] \) \( \alpha_1(n) \geq 10^{35163} \), \( \alpha_2(n) = 116812 \), \( \alpha_3(n) = 6 \), \( \alpha_4(n) = 4 \), \( \alpha_5(n) = 3 \).

One-parameter inverse Ackermann. \( \alpha(n) = \min \{ k : \alpha_k(n) \leq 3 \} \).

Ex. \( \alpha(9876!) = 5 \).

Two-parameter inverse Ackermann. \( \alpha(m, n) = \min \{ k : \alpha_k(n) \leq 3 + m / n \} \).
A tight lower bound

**Theorem.** [Fredman–Saks 1989] In the worst case, any **CELL-PROBE**(log n) algorithm requires \( \Omega(m \alpha(m, n)) \) time to perform an intermixed sequence of \( m \) **MAKE-SET**, **UNION**, and **FIND** operations on a set of \( n \) elements.

**Cell-probe model.** [Yao 1981] Count only number of words of memory accessed; all other operations are free.
Path compaction variants

Path splitting. Make every node on path point to its grandparent.
Path compaction variants

Path halving. Make every other node on path point to its grandparent.
Linking variants

**Link-by-size.** Number of nodes in tree.

**Link-by-rank.** Rank of tree.

**Link-by-random.** Label each element with a random real number between 0.0 and 1.0. Link root with smaller label into root with larger label.
Disjoint-sets data structures

Theorem. [Tarjan–van Leeuwen 1984] Starting from an empty data structure, link-by- { size, rank } combined with { path compression, path splitting, path halving } performs any intermixed sequence of \( m \geq n \) MAKE-SET, UNION, and FIND operations on a set of \( n \) elements in \( O(m \alpha(m, n)) \) time.

Worst-Case Analysis of Set Union Algorithms

ROBERT E. TARJAN

AT&T Bell Laboratories, Murray Hill, New Jersey

AND

JAN VAN LEEUWEN

University of Utrecht, Utrecht, The Netherlands

Abstract. This paper analyzes the asymptotic worst-case running time of a number of variants of the well-known method of path compression for maintaining a collection of disjoint sets under union. We show that two one-pass methods proposed by van Leeuwen and van der Weide are asymptotically optimal, whereas several other methods, including one proposed by Rem and advocated by Dijkstra, are slower than the best methods.