Impact of Great Algorithms

Internet. Packet routing, Google, Akamai.
Biology. Human genome project, protein folding.
Computers. Circuit layout, file system, compilers.
Multimedia. CD player, DVD, MP3, JPEG, DivX, HDTV.
Transportation. Airline crew scheduling, map routing.
Physics. N-body simulation, particle collision simulation.
Information processing. Database search, data compression.
...

"For me, great algorithms are the poetry of computation. Just like verse, they can be terse, allusive, dense, and even mysterious. But once unlocked, they cast a brilliant new light on some aspect of computing." - Francis Sullivan

Overview

What is COS 226?
- Intermediate-level survey course.
- Programming and problem solving with applications.
  - Data structure: method to store information.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Data Structures and Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>data types</td>
<td>stack, queue, list, union-find, priority queue</td>
</tr>
<tr>
<td>sorting</td>
<td>quicksort, mergesort, heapsort, radix sorts</td>
</tr>
<tr>
<td>searching</td>
<td>hash table, BST, red-black tree, B-tree</td>
</tr>
<tr>
<td>graphs</td>
<td>DFS, Prim, Kruskal, Dijkstra, Ford-Fulkerson</td>
</tr>
<tr>
<td>strings</td>
<td>KMP, Rabin-Karp, TST, Huffman, LZW</td>
</tr>
<tr>
<td>geometry</td>
<td>Graham scan, k-d tree, Voronoi diagram</td>
</tr>
</tbody>
</table>

A misperception: *algiros* [painful] + *arithmos* [number].

Why Study Algorithms?

Using a computer?
- Want it to go faster? Process more data?
- Want it to do something that would otherwise be impossible?

Algorithms as a field of study.
- Old enough that basics are known.
- New enough that new discoveries arise.
- Burgeoning application areas.
- Philosophical implications.
The Usual Suspects

Lectures. Kevin Wayne (Kevin)
• MW 11-12:20, Bowen 222.

Precepts. Harlan Yu (Harlan), Keith Morley (Keith)
• T 12:30, Friend 110.
• T 3:30, Friend 111.
• Clarify programming assignments, exercises, lecture material.
• First precept meets 9/20.

Coursework and Grading

Regular programming assignments: 45%
• Due 11:59pm, starting 9/26.
• More details next lecture.

Weekly written exercises: 15%
• Due at beginning of Thursday lecture, starting 9/22.

Exams:
• Closed book with cheatsheet.
• Midterm. 15%
• Final. 25%

Staff discretion. Adjust borderline cases.

Course Materials

• Syllabus.
• Exercises.
• Lecture slides.
• Programming assignments.

Algorithms in Java, 3rd edition.
• Parts 1-4. (sorting, searching)
• Part 5. (graph algorithms)

• Strings and geometry handouts.

Union Find

Network Connectivity

<table>
<thead>
<tr>
<th>in</th>
<th>out</th>
<th>evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>3 4</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>4 9</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>8 0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2 3</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>5 6</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>(2–3–4–9)</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>5 9</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>7 3</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>4 8</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>(5–6)</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>(2–3–4–8–0)</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6 1</td>
</tr>
</tbody>
</table>

An Example Problem: Network Connectivity

Network connectivity.
- Nodes at grid points.
- Add connections between pairs of nodes.
- Is there a path from node A to node B?

Goal. Design efficient data structure for union and find.
- Number of operations M can be huge.
- Number of objects N can be huge.

Union-Find Abstraction

What are critical operations we need to support?
- Objects.

| 0 1 2 3 4 5 6 7 8 9 | grid points |
- Disjoint sets of objects.

| 0 1 2 3 4 5 6 7 8 9 | subsets of connected grid points |
- Find: are objects 2 and 9 in the same set?

| 0 1 2 3 4 5 6 7 8 9 | are two grid points connected? |
- Union: merge sets containing 3 and 8.

| 0 1 2 3 4 5 6 7 8 9 | add a connection between two grid points |
Objects

Applications involve manipulating objects of all types.
- Variable name aliases.
- Pixels in a digital photo.
- Computers in a network.
- Web pages on the Internet.
- Transistors in a computer chip.
- Metallic sites in a composite system.

When programming, convenient to name them 0 to N-1.
- Details not relevant to union-find.
- Integers allow quick-access to object-related info (array indices).

Quick-Find

Data structure.
- Integer array $id[]$ of size $N$.
- Interpretation: $p$ and $q$ are connected if they have the same id.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$id[i]$</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Find. Check if $p$ and $q$ have the same id.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$id[i]$</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

Union. To merge components containing $p$ and $q$,
change all entries with $id[p]$ to $id[q]$.

Quick-Find: Java Implementation

```java
public class QuickFind {
    private int[] id;

    public QuickFind(int N) {
        id = new int[N];
        for (int i = 0; i < N; i++)
            id[i] = i;
    }

    public boolean find(int p, int q) {
        return id[p] == id[q];
    }

    public void unite(int p, int q) {
        int pid = id[p];
        for (int i = 0; i < id.length; i++)
            if (id[i] == pid) id[i] = id[q];
    }
}
```

Quick-Find [eager approach]

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$id[i]$</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

5 and 6 are connected
2, 3, 4, and 9 are connected

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$id[i]$</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

union of 3 and 6
2, 3, 4, 5, 6, and 9 are connected
many values can change

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$id[i]$</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

union of 3 and 6
2, 3, 4, 5, 6, and 9 are connected
many values can change
**Problem Size and Computation Time**

Rough standard for 2000.
- $10^9$ operations per second.
- $10^9$ words of main memory.
- Touch all words in approximately 1 second. [unchanged since 1950!]

Ex. Huge problem for quick find.
- $10^{10}$ edges connecting $10^9$ nodes
- Quick-find might take $10^{20}$ operations. [10 ops per query]
- 3,000 years of computer time!

Paradoxically, quadratic algorithms get worse with newer equipment.
- New computer may be 10x as fast.
- But, has 10x as much memory so problem may be 10x bigger.
- With quadratic algorithm, takes 10x as long!

---

**Quick-Union**

Data structure.
- Integer array $id[]$ of size $N$.
- Interpretation: $id[x]$ is parent of $x$.
- Root of $x$ is $id[id[...id[x]...]]$.

Find. Check if $p$ and $q$ have the same root.

Union. Set the id of $q$’s root to the id of $p$’s root.

**Quick-Union: Java Implementation**

```java
public class QuickUnion {
    private int[] id;

    public QuickUnion(int N) {
        id = new int[N];
        for (int i = 0; i < N; i++) id[i] = i;
    }

    private int root(int x) {
        while (x != id[x]) x = id[x];
        return x;
    }

    public boolean find(int p, int q) {
        return root(p) == root(q);
    }

    public void unite(int p, int q) {
        int i = root(p), j = root(q);
        if (i == j) return;
        id[i] = j;
    }
}
```
Weighted Quick-Union

Weighted quick-union.
- Modify quick-union to avoid tall trees.
- Keep track of size of each component.
- Balance by linking small tree below large one.

Ex: union of 5 and 3.
- Quick union: link 9 to 6.
- Weighted quick union: link 6 to 9.

Java implementation.
- Almost identical to quick-union.
- Maintain extra array sz[] to count number of elements in the tree rooted at i.

Find. Identical to quick-union.

Union. Same as quick-union, but merge smaller tree into the larger tree, and update the sz[] array.

```
if (sz[i] < sz[j]) { id[i] = j; sz[j] += sz[i]; }
else
    { id[j] = i; sz[i] += sz[j]; }
```
Weighted Quick-Union: Analysis

Analysis.
- Find: takes time proportional to depth of p and q.
- Union: takes constant time, given roots.
- Fact: depth is at most 1 + \( \lg N \). [needs proof]

<table>
<thead>
<tr>
<th>Data Structure</th>
<th>Union</th>
<th>Find</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quick-find</td>
<td>( N )</td>
<td>1</td>
</tr>
<tr>
<td>Quick-union</td>
<td>( 1 )</td>
<td>( \lg N )</td>
</tr>
<tr>
<td>Weighted QU</td>
<td>( \lg N )</td>
<td>( \lg N )</td>
</tr>
</tbody>
</table>

Stop at guaranteed acceptable performance? No, can improve further.

Weighted Quick-Union with Path Compression

Path compression.
- Standard implementation: add second loop to root to set the id of each examined node to the root.
- Simpler one-pass variant: make each examined node point to its grandparent.

```java
public int root(int x) {
    while (x != id[x]) {
        id[x] = id[id[x]];  
        x = id[x];
    }  
    return x;
}
```

In practice. No reason not to! Keeps tree almost completely flat.

Path compression. Just after computing the root of x, set id of each examined node to root(x).
Weighted Quick-Union with Path Compression

Theorem. A sequence of $M$ union and find operations on $N$ elements takes $O(N + M \log^* N)$ time.

- Proof is very difficult.
- But the algorithm is still simple!

Remark. $\log^* N$ is a constant in this universe.

Linear algorithm?
- Cost within constant factor of reading in the data.
- Theory: WQUPC is not quite linear.
- Practice: WQUPC is linear.

### Context

Ex. Huge practical problem.
- $10^{10}$ edges connecting $10^9$ nodes.
- WQUPC reduces time from 3,000 years to 1 minute.
- Supercomputer wouldn’t help much.
- Good algorithm makes solution possible.

Bottom line. WQUPC on cell phone beats QF on supercomputer!

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quick-find</td>
<td>$M N$</td>
</tr>
<tr>
<td>Quick-union</td>
<td>$M N$</td>
</tr>
<tr>
<td>Weighted QU</td>
<td>$N + M \log N$</td>
</tr>
<tr>
<td>Path compression</td>
<td>$N + M \log N$</td>
</tr>
<tr>
<td>Weighted + path</td>
<td>$5 (M + N)$</td>
</tr>
</tbody>
</table>

Other Applications

Union-find applications.
- Hex.
- Percolation.
- Connectivity.
- Image processing.
- Least common ancestor.
- Equivalence of finite state automata.
- Hinley-Milner polymorphic type inference.
- Kruskal’s minimum spanning tree algorithm.
- Compiling equivalence statements in Fortran.
- Scheduling unit-time tasks to $P$ processors so that each job finishes between its release time and deadline.
**Hex.** [Piet Hein 1942, John Nash 1948, Parker Brothers 1962]
- Two players alternate in picking a cell in a hex grid.
- Black: make a black path from upper left to lower right.
- White: make a white path from lower left to upper right.

**Percolation**

**Q.** What is percolation threshold $p^*$ at which charge carriers can percolate from top to bottom?

**A.** $\approx 0.592746$ for square lattices. [constant only known via simulation]

**Goal.** Algorithm to detect when a player has won.

**Percolation phase-transition.**
- Two parallel conducting bars (top and bottom).
- Electricity flows from a site to one of its 4 neighbors if both are occupied by conductors.
- Suppose each site is randomly chosen to be a conductor or insulator with probability $p$.

**Lessons.**
- Simple algorithms can be very useful.
- Start with brute force approach.
  - don’t use for large problems
  - can’t use for huge problems
- Strive for worst-case performance guarantees.
- Identify fundamental abstractions: union-find.
- Apply to many domains.

**Summary**

- **Apply to many domains.**