Announcements

• Markov due Sept 21 (part1) and Sept 22 (part2)
• APT-TWO due Sept 22

• Recitation this week on Markov
Using the Debugger in Eclipse

• Debugger allows you to pause your program during execution and look around.
  – First set breakpoints (right click on line of code, toggle breakpoint)
  – Under Run menu, select “Debug”
    • Stops when it reaches a breakpoint
    • You can examine the values of variables
Hashing

• Storage – hash table or hash buckets (implementation could be array, map, tree, or other structure)
• Each data item has a key
• Hash function that maps key to address in the hash table \( H(key) = \text{address} \)
• Best if the “key” is near the address
• “collision” if two items map to the same address
Example Hashing

- Duke's ACM Chapter wants to be able to quickly find out info about its members. Also add, delete and update members. Doesn't need members sorted.
  - 267-89-5432  John Smith
  - 703-25-6142  Jack Adams
  - 319-86-2100  Betty Harris
  - 476-82-5142  Rose Black

- Possible Hash Function: \( H(\text{ssn}) = \text{last 3 digits} \)
- Hash Table size is 1000, range from 0 to 999
- Will there be a collision with data above?
Bucket Hashing

Buckets 0-999 (not all buckets shown)

- Buckets hold a lot of values
- Apply hash function to data. What happens?

\[
\begin{align*}
H(267-89-5432) &= 432 \\
\text{John Smith} \\
H(703-25-6142) &= 142 \\
\text{Jack Adams} \\
H(319-86-2100) &= 100 \\
\text{Betty Harris} \\
H(476-82-5142) &= 142 \\
\text{Rose Black}
\end{align*}
\]
Buckets 0-999 (not all buckets shown)

* Apply hash function to data. What happens?

\[ H(267-89-5432) = 432 \]
John Smith

\[ H(703-25-6142) = 142 \]
Jack Adams

\[ H(319-86-2100) = 100 \]
Betty Harris

\[ H(476-82-5142) = 142 \]
Rose Black

0
100
142
143
100
432
742
999

CompSci 100, Fall 2010
### Buckets 0-999 (not all buckets shown)

- Apply hash function to data. What happens?

<table>
<thead>
<tr>
<th>Bucket</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Betty Harris</td>
</tr>
<tr>
<td>142</td>
<td>Jack Adams</td>
</tr>
<tr>
<td>143</td>
<td></td>
</tr>
<tr>
<td>432</td>
<td>John Smith</td>
</tr>
<tr>
<td>742</td>
<td></td>
</tr>
<tr>
<td>999</td>
<td></td>
</tr>
</tbody>
</table>

- \( H(267-89-5432) = 432 \)
  - John Smith

- \( H(703-25-6142) = 142 \)
  - Jack Adams

- \( H(319-86-2100) = 100 \)
  - Betty Harris

- \( H(476-82-5142) = 142 \)
  - Rose Black

Where does Rose Black go?
Buckets 0-999 (not all buckets shown)

• Apply hash function to data.
  What happens?

  \[
  H(267-89-5432) = 432 \\
  H(703-25-6142) = 142 \\
  H(319-86-2100) = 100 \\
  H(476-82-5142) = 142
  \]

  John Smith  \\
  Jack Adams  \\
  Betty Harris  \\
  Rose Black

• If buckets are one array, where does Rose Black go?

  \[
  H(267-89-5432) = 432 \\
  H(703-25-6142) = 142 \\
  H(319-86-2100) = 100 \\
  H(476-82-5142) = 142
  \]

  John Smith  \\
  Betty Harris  \\
  Jack Adams  \\
  Rose Black
If instead of Buckets, one String array 0-999, Where would Rose Black go?

- Collision, must be resolved, one way – next possible bucket (slot)
- Other Collision resolution methods

<table>
<thead>
<tr>
<th>Hash Value</th>
<th>Name</th>
<th>Hash Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H(267-89-5432) = 432</td>
<td>John Smith</td>
<td>H(703-25-6142) = 142</td>
</tr>
<tr>
<td>H(319-86-2100 )= 100</td>
<td>Betty Harris</td>
<td>H(476-82-5142) = 142</td>
</tr>
<tr>
<td>H(476-82-5142) = 142</td>
<td>Rose Black</td>
<td>H(476-82-5142) = 142</td>
</tr>
</tbody>
</table>

CompSci 100, Fall 2010
Hash Functions

• Want hash function with the fewest collisions, data distributed
• Random number Generator with seed
  – H(key) is random(key)\*N
• Shift folding with 100 buckets
  – H(123-45-6789) is (123 + 45 + 6789) mod 100
• Use ascii code
  – H(BANKS) = 66 + 65 + 78 + 75 + 83 = 367
  – Where ascii(B) = 66, etc.
• Java – hashCode()
public int calc(int n){
    return n*calc(n-1);
}

• What is the Internet?
  – A network of networks. Or ...

• What is recursive DNS?
  – What IP is fxyztl.com?

• What is PageRank?
  – Where is it used?
Other Topics

- Structure driving ‘self-referential’ code
  - motivation
- Basic understanding of recursion
  - Principles
  - Examples
- Blob and grid examples
Quota Exceeded: coping with storage

• You’re running out of disk space
  – Buy more
  – Compress files
  – Delete files

• How do you find your “big” files?
  – What’s big?
  – How do you do this?
BlobCount or edge detection or ...

- How do we find images? Components? Paths?
  - Create information from data
Tools: Solving Computational Problems

• Algorithmic techniques and paradigms
  – Brute-force/exhaustive, greedy algorithms, dynamic programming, divide-and-conquer, ...
  – Transcend a particular language
  – Designing *algorithms*, may change when turned into code

• Programming techniques and paradigms
  – Recursion, memo-izing, compute-once/lookup, tables, ...
  – Transcend a particular language
  – Help in making code work
    • Cope with correctness and maintenance
    • Cope with performance problems
Tools: Solving Computational Problems

• Java techniques
  – `java.util.*`, Comparator, Priority Queue, Map, Set, ...
  – These aren’t really Java-specific, but realized in Java
  – Map, Comparator, Set: C++, Python, ....
  – We learn idioms in a language and talk about abstractions

• Analysis of algorithms and code
  – Mathematical analysis, empirical analysis
  – We need a language and techniques for discussion
  – Theory and practice, real problems and in-the-limit issues

• “In theory there is no difference between theory and practice, but in practice there is.”
  (attributed to many)
Recursive structure matches code

```java
public static final long THRESHOLD = 1000000L; // one million bytes

public static void findBig(File dir, String tab) {
    File[] dirContents = dir.listFiles();
    System.out.println(tab + "**:" + dir.getName());
    for (File f : dirContents) {
        if (f.isDirectory()) {
            findBig(f, tab + "\t" + dir.getName());
        } else {
            if (f.length() > THRESHOLD) {
                System.out.println(tab + f.getName() + f.length());
            }
        }
    }
}
```

Does findBig call itself?
Solving Problems Recursively

• Recursion: indispensable in programmer’s toolkit
  – Allows many complex problems to be solved simply
  – Elegance and understanding in code often leads to better programs: easier to modify, extend, verify (and sometimes more efficient!!)
  – Sometimes recursion isn’t appropriate, when it’s bad it can be very bad---every tool requires knowledge and experience in how to use it

• The basic idea is to get help solving a problem from coworkers (clones) who work and act like you do
  – Ask clone to solve a simpler but similar problem
  – Use clone’s result to put together your answer

• Need both: call on the clone and use the result
Print words read, but print backwards
• Could store words and print in reverse order, but ...
  – Probably the best approach, recursion works too
    public void printReversed(Scanner s)
    {
      if (s.hasNext())
      {
        String word = s.next();  // store word
        printReversed(s);         // print rest
        System.out.println(word); // print the word
      }
    }

• The function printReversed reads a word, only
  prints word after the clones finish printing
  – Each clone has own version of the code, own word
    variable
  – Who keeps track of the clones?
  – How many words are created when reading $N$ words?
    • Can we do better?
Exponentiation

• Computing $x^n$ means multiplying $n$ numbers
  – Does it require $n$ multiplies?
  – What’s the simplest value of $n$ when computing $x^n$?
  – To only multiply once, what can you ask a clone?

```java
public static double power(double x, int n){
    if (n == 0){
        return 1.0;
    }
    return x * power(x, n-1);
}
```

• Number of multiplications?
  – Note base case: no recursion, no clones
  – Note recursive call: moves toward base case (unless ...)

CompSci 100, Fall 2010
Faster exponentiation

• Recursive calls made to compute $2^{1024}$?
  – How many multiplies on each call? Is this better?

```
public static double power(double x, int n) {
    if (n == 0) return 1.0;
    double semi = power(x, n/2);
    if (n % 2 == 0) return semi*semi;
    return x * semi * semi;
}
```

• What about an iterative version of this function?
  – Why might we want such a version?
Back to Recursion

• Recursive functions have two key attributes
  – There is a base case, aka exit case: no recursion!
    • See print reversed, exponentiation
  – All other cases make a recursive call, with some measure (e.g., parameter value) that decreases towards the base case
    • Ensure that sequence of calls eventually reaches the base case
    • “Measure” can be tricky, but usually it’s straightforward

• Example: finding large files in a directory (on a hard disk)
  – Why is this inherently recursive?
  – How is this different from exponentation?
Thinking recursively: recursive Max

public static double recMax(double[] a, int index){
    if (index == a.length-1){ // last element, done
        return a[index];
    }
    double maxAfter = recMax(a,index+1);
    return Math.max(a[index],maxAfter);
}

• What is base case (conceptually)?
  – Do we need variable maxAfter?
• Use recMax to implement arrayMax as follows:
  – Introduce auxiliary variable from arrayMax(a) to ...
  – return recMax(a,0);  

Recursive methods sometimes use extra parameters; helper methods set this up
Recognizing recursion:

```java
public void change(String[] a, int first, int last){
    if (first < last) {
        String temp = a[first];  // swap first/last
        a[first] = a[last];
        a[last] = temp;
        change(a, first+1, last-1);
    }
}
```

// original call (why?): change(a, 0, a.length-1);

- What is base case? (no recursive calls)
- What happens before recursive call made?
- How is recursive call closer to the base case?

Recursive methods sometimes use extra parameters; helper methods set this up
public static int value(int[] a, int index){
    if (index < a.length) {
        return a[index] + value(a, index+1);
    }
    return 0;
}

// original call: int v = value(a, 0);

• What is base case, what value is returned?
• How is progress towards base case realized?
• How is recursive value used to return a value?
• What if a is array of doubles, does anything change?

Recursive methods sometimes use extra parameters; helper methods set this up
From programming techniques to Java

• Is recursion a language independent concept?
  – Do all languages support recursion?
  – What are the issues in terms of computer/compiler/runtime support?

• We use a language and its libraries, study them?
  – Should we know how to implement ArrayList
    • What are the building blocks, what are our tools
  – Should we know how to implement different sorts
    • Should we know how to call existing sorts
Fran Allen

- IBM Fellow, Turing Award
  - Optimizing compilers

- Taught high school for two years, then Master’s degree and IBM
  - Teachers excited me to learn

I’ve always felt that theory without practice is maybe nice and maybe pretty, but it’s not going to influence computing as much as the practice side. But the practice has to be backed up with the ability to talk about it, reason about it, and formulate it so that it can be reproduced.
Blob Counting, Flood Fill

• Flood a region with color
  – Erase region, make transparent, ..
  – How do find the region?

• Finding regions, blobs, edges, ..
  – See blob counting code
  – What is a blob?

• Recursion helps, but necessary?
  – Performance, clarity, ...
  – Ease of development
Details and Idioms in blob code

- **Method** `blobFill` **has four parameters**
  - `(row,column)` of where search starts
  - Character being searched for (initially * or blob)
  - Character to fill with on success (e.g., count ‘2’ or ‘4’)
    - Mark for visualization
    - Mark to ensure we don’t search again!
- **If** `(row,column)` **is part of blob, count it and ask neighbors for their counts**
  - They’re part of blob (if never visited before)

- **Return total of yourself and neighbors**
  - Key to recursion: do one thing and ask for help
Blob questions

• What changes if diagonal cells are adjacent?
  – Conceptually and in code

• How do we find blob sizes in a range?
  – Not bigger than X, but between X and Y

• How would we number blobs by size rather than by when they’re found?
  – Do we have the tools to do this in existing code?

• Can we avoid recursion and do this iteratively?