Yet More Array Practice

Lecture 24, Mon Mar 15 2010

borrowing from slides by Kurt Eiselt

http://www.cs.ubc.ca/~tmm/courses/111-10
Reading: CORRECTION!

- This week: Sorting and Searching
  - 14.1 and 14.3 in 3rd edition
  - (19.1 and 19.3 in 2nd edition)
  - so weekly question is indeed required!

- course web page has also been updated
Midterm 2

- Midterm 2: Mon Mar 22, 6:30pm
  - FSC 1005 again
  - hour-long exam, reserve 6:30-8 time slot
    - for buffer in case of fire alarms etc
- coverage: through arrays (Chap 8)
  - includes/builds on material covered in previous midterm
- study tips - same as before!
  - write and test programs, not just read book
    - try programming exercises from book
Reading through Midterm, 3rd edition

- 1.1-1.8
- 2.1-2.10
- 3.1-3.8
- 4
- 5.1-5.4
- 6.1-6.5
- 7.1, 7.5-7.7
- 14.1, 14.3

see course page for 2nd edition list
Recap: Arrays of Arrays = 2D Arrays

- 2D array often easier to think about
- Internally, 2D arrays implemented as arrays of arrays in Java
  - they’re equivalent
Recap: 2D Array Access Patterns

scores

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>95</td>
<td>82</td>
<td>13</td>
<td>96</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>68</td>
<td>63</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>71</td>
<td>84</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>99</td>
<td>70</td>
<td>32</td>
<td>12</td>
</tr>
</tbody>
</table>

- Print average score for each student
  - for each row of scores
    - add up scores
    - divide by number of quizzes
      - length of row

```java
for (int row = 0; row < scores.length; row++) {
    double average = 0;
    for (int col = 0; col < scores[row].length; col++) {
        average = average + scores[row][col];
    }
    average = average / scores[row].length;
}
```

- Print average score for each quiz
  - for each column of scores
    - add up scores
    - divide by number of students
      - length of column

```java
for (int col = 0; col < scores[0].length; col++) {
    double average = 0;
    for (int row = 0; row < scores.length; row++) {
        average = average + scores[row][col];
    }
    average = average / scores.length;
}
```
Recap: Per-Student Averages

public class ArrayEx4
{
    public static void main(String[] args)
    {
        double[][] scores = {{95, 82, 13, 96},
                             {51, 68, 63, 57},
                             {73, 71, 84, 78},
                             {50, 50, 50, 50},
                             {99, 70, 32, 12}};
        double average;

        // here's where we control looping row by row (student by student)
        for (int row = 0; row < scores.length; row++)
        {
            average = 0;
            // and here's where we control looping through the columns
            // (i.e., quiz scores) within each row
            for (int col = 0; col < scores[row].length; col++)
            {
                average = average + scores[row][col];
            }
            average = average / scores[row].length;
            System.out.println("average of row " + row + " is " + average);
        }
    }
}
Recap: Selection Sort

- Start at beginning
- Consider unsorted array elements: beyond current spot
  - Find smallest element
  - Swap with current spot
  - Move down by one

The smallest value so far is 12

Its index is 4
Multidimensional Arrays

- now that we know 2D, we can do nD!
  - any number of dimensions: 3D, 4D…
    - up to 127D, actually
- example: student quiz scores over multiple terms
  - row: students
  - col: quiz scores
  - stack: term

- let's try it!

<table>
<thead>
<tr>
<th>04-05 Term1</th>
<th>04-05 Term2</th>
<th>05-06 Term1</th>
<th>05-06 Term2</th>
</tr>
</thead>
<tbody>
<tr>
<td>95  82  13  96</td>
<td>51  68  63  57</td>
<td>73  71  84  78</td>
<td>50  50  50  50</td>
</tr>
<tr>
<td>99  70  32  12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Now for the good stuff

Computer science folks don't spend all their time writing programs. They're also concerned with the efficiency of those programs and their underlying algorithms. Efficiency can be expressed in terms of either time or memory needed to complete the task. In the case of sorting algorithms, we're typically interested in how much time it takes to sort.

So let's try to get some sense of the time requirements of selection sort. We don't use a stopwatch...instead, we use mathematics. The fundamental operation in sorting is the comparison to see if one value is less than the other, and the time required to sort corresponds to the number of comparisons that must be made to complete the sorting.
We can go back to the selection sort example and count the comparisons. The first pass through the array of 5 elements started with 16 being compared to 3, then 3 was compared to 19, 8, and 12. There were 4 comparisons. The value 3 was moved into the location at index 0.
Estimating time required to sort

We can go back to the selection sort example and count the comparisons. The first pass through the array of 5 elements started with 16 being compared to 3, then 3 was compared to 19, 8, and 12. There were 4 comparisons. The value 3 was moved into the location at index 0. Then the second pass through the array began, starting with index 1. 16 was compared to 19, then 16 was compared to 8, which became the new minimum and was compared to 12. So among 4 elements in the array, there were 3 comparisons.
Estimating time required to sort

It takes 4 passes through the array to get it completely sorted. There are 4 comparisons on the first pass, 3 comparisons on the second pass, 2 comparisons on the third pass, and 1 comparison on the last pass. That is, it takes $4 + 3 + 2 + 1 = 10$ comparisons to sort an array of five values.

If you do this same computation on an array with six values, you'll find it takes $5 + 4 + 3 + 2 + 1 = 15$ comparisons to sort the array. Do you see a pattern?

With a little math, you can figure out that the number of comparisons required to perform selection sort on an array of $N$ values is given by the expression: $N(N-1)/2$ or $(N^2-N)/2$
Estimating time required to sort

Either way, it should be easy to see that as $N$, the number of values in the array gets very big, the number of comparisons needed to sort the array grows in proportion to $N^2$, with the other terms becoming insignificant by comparison.

So sorting an array of 1,000 values would require approximately 1,000,000 comparisons. Similarly, sorting an array of 1,000,000 values would take approximately 1,000,000,000,000 comparisons.

As the number of values to be sorted grows, the number of comparisons required to sort them grows much faster. Fortunately, there are other sorting algorithms that are much less time-consuming, but we won't be talking about them in this class. In the meantime, here are some real numbers to help you think about just how long it might take to sort some really big arrays...
## Estimating time required to sort

Let's assume that your computer could make 1 billion (1,000,000,000) comparisons per second. That's a lot of comparisons in a second. And let's say your computer was using selection sort to sort the names of the people in the following hypothetical telephone books. Here's some mathematical food for thought.

<table>
<thead>
<tr>
<th>phone book</th>
<th>number of people (N)</th>
<th>(N^2)</th>
<th>number of seconds needed to sort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Estimating time required to sort

Let's assume that your computer could make 1 billion (1,000,000,000) comparisons per second. That's a lot of comparisons in a second. And let's say your computer was using selection sort to sort the names of the people in the following hypothetical telephone books. Here's some mathematical food for thought.

<table>
<thead>
<tr>
<th>phone book</th>
<th>number of people (N)</th>
<th>(N^2)</th>
<th>number of seconds needed to sort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancouver</td>
<td>544,320</td>
<td>296,284,262,400</td>
<td>296 or 5 minutes</td>
</tr>
</tbody>
</table>
Estimating time required to sort

Let's assume that your computer could make 1 billion (1,000,000,000) comparisons per second. That's a lot of comparisons in a second. And let's say your computer was using selection sort to sort the names of the people in the following hypothetical telephone books. Here's some mathematical food for thought.

<table>
<thead>
<tr>
<th>Phone book</th>
<th>Number of people (N)</th>
<th>N²</th>
<th>Number of seconds needed to sort</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancouver</td>
<td>544,320</td>
<td>296,284,262,400</td>
<td>296</td>
<td>or 5 minutes</td>
</tr>
<tr>
<td>Canada</td>
<td>30,000,000</td>
<td>900,000,000,000,000</td>
<td>900,000</td>
<td>or 10.4 days</td>
</tr>
</tbody>
</table>
Estimating time required to sort

Let's assume that your computer could make 1 billion (1,000,000,000) comparisons per second. That's a lot of comparisons in a second. And let's say your computer was using selection sort to sort the names of the people in the following hypothetical telephone books. Here's some mathematical food for thought.

<table>
<thead>
<tr>
<th>phone book</th>
<th>number of people (N)</th>
<th>(N^2)</th>
<th>number of seconds needed to sort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancouver</td>
<td>544,320</td>
<td>296,284,262,400</td>
<td>296 or 5 minutes</td>
</tr>
<tr>
<td>Canada</td>
<td>30,000,000</td>
<td>900,000,000,000,000</td>
<td>900,000 or 10.4 days</td>
</tr>
<tr>
<td>People's Republic of China</td>
<td>1,000,000,000</td>
<td>1,000,000,000,000,000,000</td>
<td>1,000,000,000 or 31.7 years</td>
</tr>
</tbody>
</table>
Estimating time required to sort

Let's assume that your computer could make 1 billion (1,000,000,000) comparisons per second. That's a lot of comparisons in a second. And let's say your computer was using selection sort to sort the names of the people in the following hypothetical telephone books. Here's some mathematical food for thought.

<table>
<thead>
<tr>
<th>phone book</th>
<th>number of people (N)</th>
<th>$N^2$</th>
<th>number of seconds needed to sort</th>
<th>time needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vancouver</td>
<td>544,320</td>
<td>296,284,262,400</td>
<td>296</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Canada</td>
<td>30,000,000</td>
<td>900,000,000,000,000,000</td>
<td>900,000</td>
<td>10.4 days</td>
</tr>
<tr>
<td>People's Republic of China</td>
<td>1,000,000,000</td>
<td>1,000,000,000,000,000,000</td>
<td>1,000,000,000</td>
<td>31.7 years</td>
</tr>
<tr>
<td>World</td>
<td>6,000,000,000</td>
<td>36,000,000,000,000,000,000</td>
<td>36,000,000,000</td>
<td>1142 years</td>
</tr>
</tbody>
</table>
Favorite Colors

- record everybody's favorite color
- how can we do "averages" per row?