CS112 Lab 06, Feb 25, 28 2010

http://cs-people.bu.edu/deht/cs112_spring11/lab06/

Diane H. Theriault

deht@cs.bu.edu

http://cs-people.bu.edu/deht/

Application: Rank Aggregation

- Get search results from different search engines and combine the rankings

- Determining similarity between two things based on search results

- http://www.eecs.harvard.edu/~michaelm/CS222/rank.pdf
  (just the introduction will give you the flavor of the problem)

Kendall-Tau Distance

- Measure of the difference between two rankings
- Count the number of instances where two items appear in different orders in two lists
- http://en.wikipedia.org/wiki/Kendall_tau_distance

Counting the Number of Inversions

- Given an array of integers, count the number of pairs that are out of order.

- Ex: 1 2 3 4 5
  3 4 1 2 5
- Number of inversions: 4

Practical Lab

- Using Java.util.Random
  http://download.oracle.com/javase/6/docs/api/java/util/Random.html)

```
Random generator = new Random();
generator.setSeed(0);
generator.nextInt();
```

- Seeding the random number generator is very important for controlling repeatability.

Practical Lab

- Write an $O(n^2)$ algorithm to count the number of inversions in a randomly generated array

- Hint: Enumerate all pairs, then count how many are backwards

Practical Lab (Bonus)

- Write an O(n log n) algorithm to count the number of inversions.
- You can do this by hacking Mergesort and counting the number of times you have to move elements.
- http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient#Algorithms