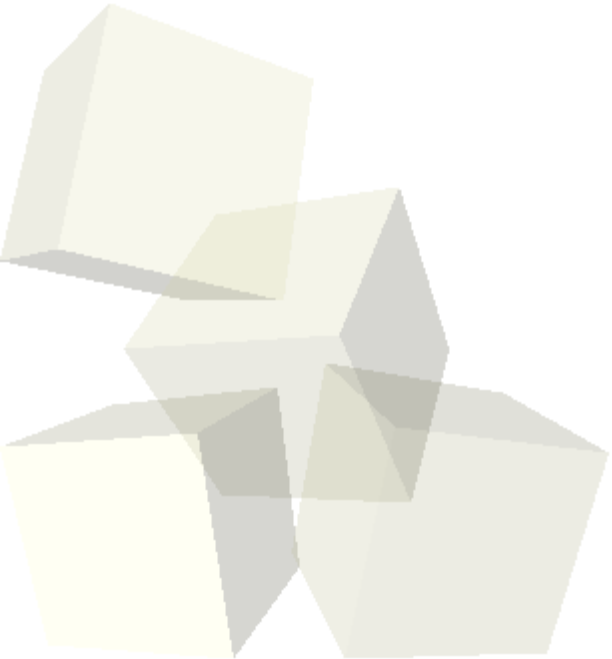
Sequence Searching

11-14-2005

Opening Discussion

- What did we talk about last class?
- How did the assignment go?



Comparing Sequences

- Once biologists have sequences of DNA or RNA from difference species, they want to be able to compare them.
- This comparison can't be a simple equality comparison. The desired comparison is one of homology, could one have evolved from the other or could they have shared a common ancestor.
- The details of how these comparison are done is a bit beyond the scope of this class. However, we can speculate on it a bit to see what actually matters.

Longest Common Subsequence

- A standard problem in computer science that can give us some insight into how sequence comparisons work is the longest common subsequence problem.
- You are given two strings and are asked what is the longest string that appears in proper order in both of the strings.
- Let's look at some examples and consider different ways that we might go about trying to solve this problem.

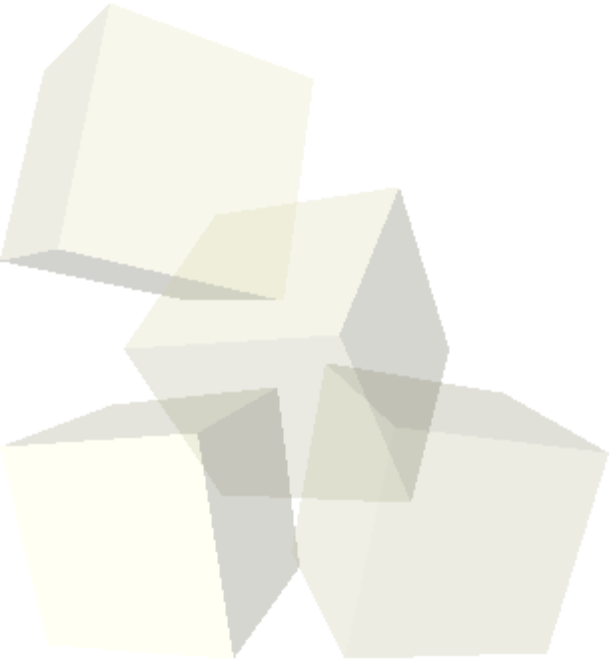
Recursive Solution

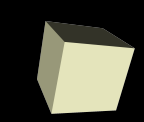
- The most intuitive solution to the longest common subsequence problem is a recursive one. First we can build up a recurance relationship in mathematical terms.
- Once we have it in this form we can convert it into code fairly easily.

$$lcs(m,n)=\begin{cases} 0 \; if \; m=0 \vee n=0 \\ lcs(m-1,n-1)+1 \; if \; s1(m)=s2(n) \\ max(lcs(m-1,n), lcs(m,n-1) \; otherwise) \end{cases}$$

Speed Issues

- What is the order of the code we just created? Will we be able to apply this to DNA sequences with many thousands of base pairs?





Reminders

- Quiz #5 will be next class.