CS250 Intro to CS II

Spring 2013

Review

- Review Reading: pp. 265-284; 377-424

- Topics
  - Files (Reading & Writing)
  - Arrays (1D & 2D)
 - Character Processing

Files

- What is a stream?

- How do we create a stream for reading from a file?

Bioinformatics

- bioinformatics

  - is the study of methods for storing, retrieving, and analyzing biological data

  - generates new knowledge useful in fields such as drug design and drug discovery

  - develops tools to help study how normal cellular activities are altered in different disease states

  - Perl is a language of choice for solving many bioinformatics problems but C++ can also be used quite effectively

Problem #1
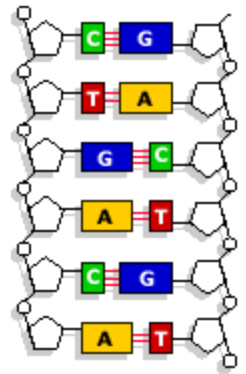
- A DNA string represents the order of nucleobases along one strand of a double-stranded DNA molecule. The other strand is the reverse complement of the string.

- DNA strings are constructed from the alphabet {A, C, G, T} representing the bases adenine, cytosine, guanine, and thymine

- The DNA string AAGATGCCGT has length 10 nucleobases (or just bases).

- The file dna.txt contains a dna string. Write a function dnaLength that returns the length of a DNA string
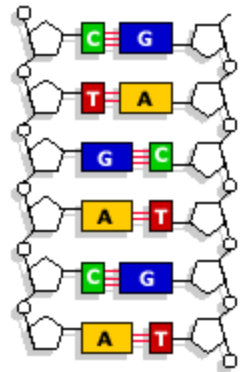
One-dimensional arrays

- Consider
  ```
  int intArry[] = {1, 2, 3, 4, 5};
  double doubleArry[10];
  ```

1. What are the index values for each array?

2. How many elements does each array have?

3. Arrays consist of homogeneous data. What does this mean?

Problem #2

- In DNA strings, the symbols A and T are complements of each other as are C and G.

- The reverse complement of a DNA string s is formed by reversing the contents of s and then taking the complement of each symbol.

- The DNA string AAAACCCGGT has the reverse complement ACCGGGTTTT

- The data file dnastrings.txt contains an unknown number of DNA strings where each string has length at most 1000 bases. Output each DNA string and the reverse complement of each DNA string.

- Let's write well-defined function prototypes for solving the above problem and then write each function definition as well as the main function for solving the stated problem

Two-dimensional arrays

- Consider
  ```
  int intArry[][2] = {{1, 2}, {3, 4}};
  double doubleArry[10][5];
  ```

1. What are the index values for each array?

2. How many elements does each array have?

3. Show how `doubleArry` is passed to a function and what a function prototype might look like.

Problem #3

- A data file lottery.txt contains an unknown number (but at most 1000) of lottery ticket data where each ticket contains 6 lottery numbers between 1 and 52.

- Write a C++ program that loads a 2D array with the lottery numbers and then has the user enter the actual winning lottery numbers from the keyboard into a 1D array.

- Print the number of winning (matched all numbers) lottery tickets. The ticket numbers and the winning numbers must match exactly (i.e. order is important)
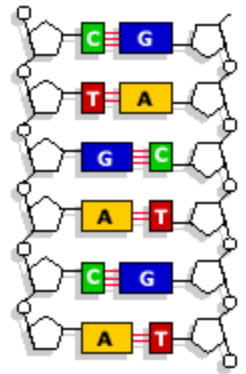
Problem #3

- Here are some other functions you might wish to consider:

  1. Print the number of tickets that matched 0, 1, … 6 numbers

  2. Print the number of winning tickets where order is not important

  3. Print the number of tickets that matched 0, 1, … 6 numbers where order is not important

Problem #4
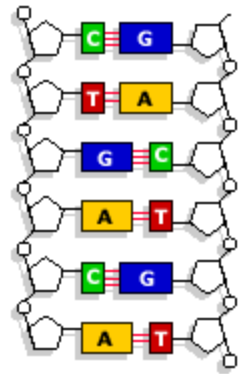
- Consider two DNA strings $s1$ and $s2$ of equal length. The p-distance, denoted $p(s1, s2)$, is the proportion of corresponding symbols that differ between $s1$ and $s2$.

- A collection of at most 10 equal length (at most 1000 bases) DNA strings exist in a file dnastrings.txt. Compute and return a matrix containing the p-distance on the given strings.
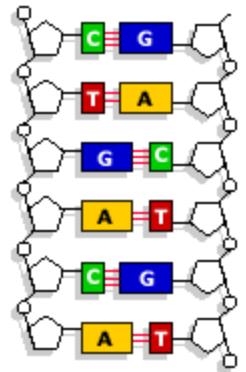
Problem #4

- Given the following strings:
  TTTCCATTTA
  GATTCATTTC
  TTTCCATTTT
  GTTCCATTTA

- The p-distance matrix would be:
  ```
  0.0 0.4 0.1 0.1
  0.4 0.0 0.4 0.3
  0.1 0.4 0.0 0.2
  0.1 0.3 0.2 0.0
  ```

Problem #4

- How might we break this problem up into well-defined functions. Let's write the function prototypes and the main program before writing any function definitions.