Programming Assignment #6

Plagiarism catcher support

CS 3358.501, Summer I 2012
Instructor: Jill Seaman

Due: Monday, 7/2/2012 (upload electronic copy by 4:30pm)

Problem:

You will be writing **part** of a tool that can be used to catch plagiarists. The goal of the tool is to determine similarities between documents in a large set to find out if plagiarism is going on within the group.

For this assignment you will write a program that will be able to compare two documents and indicate how many n-word sequences they have in common.

Your program should take as input the name of two files in the working directory, and the value for n (the number of words in a sequence). These can be either input on the command line, or your program can prompt the user to enter them. Then your program should output the number of n-word sequences that the files have in common (ignoring duplicates in a given file).

You should use a hash table to compute the results for your program. Implement a hash table for containing strings using the class declaration given in hashtable_3358.h. Implementing the hash table is a requirement for this programming assignment.

You should also use the program you wrote for assignment 5 (or the solution on TRACS) as a starter for reading in the two files and collecting the n-word sequences.

NOTES:

- I will be putting a zip file of text files on the class website soon. You can use these files to test your tool (see the file called "catchmeifyoucan.txt").

- See the end of this file for code to implement isPrime and nextPrime (declared in HashTable_3358)

- I recommend making use of these member functions from the string class:
  bool empty(): returns true if the string is empty
  void clear(): sets string content to empty string (size == 0)

- Write your own hashtable_test.cpp file to test your implementation before you start writing the main program. This will save you from headaches later.

Style:

See the Style Guidelines document on the course website.

Logistics:

Please submit the following files in a single zip file. You can call it assign6_xxxxxx.zip.

`hashtable_3358.h hashtable_3358.cpp stop_plagiarism.cpp`

The xxxxx is your TX State NetID (your txstate.edu email id).

Submit: an electronic copy only, using the Assignments tool on the TRACS website for this class.

```cpp
bool HashTable_3358::isPrime(int number)
{   // Given:   num an integer > 1
    // Returns: true if num is prime
    //               false otherwise.

    int i;
    for (i=2; i<number/2; i++)
    {
         if (number % i == 0)
         {
              return false;
         }
    }
    return true;
}

int HashTable_3358::nextPrime (int number) {
    int x = number+1;
    while (!isPrime(x))
        x++;
    return x;
}
```