

CIS 223 Computer Science II

Assignment Two

Due: April 05, 2022

There are two credit levels for this assignment. Completing the “Step 1” level correctly is worth 12/20 points, completing the “Step 2” level is worth 20/20. You are encouraged to get the code for the “Step 1” level working first and then complete the “Step 2” level. You should include a comment at the top of your source code that indicates what level you are attempting.

- *You may work either alone or in a pair. Place both of your and your partner's names in the header*

StringSet Class

The String class is provided in the C++ library and it supports the overloaded operators. Define your own string class called StringSet that will be used to store a set of STL strings. Provide your own implementation for the member functions and overload the following operators.

Use an array (dynamic array), or a vector to store the string.

Create a constructor that takes as an input parameter an array of strings for the initial values in the set. Define a copy constructor and destructor.

Then write member functions to add a string remove a string from the set, clear the entire set to empty, check the string is in the set or not, search for a string, return the size of strings in the set, and output all strings in the set.

Also define a function to read data from an input file.

Implement the following overloaded operators:

- Overload the + operator so that it returns the unions of two StringSet objects.
- Overload the *operator so that it returns the intersection of two StringSet objects.
- Overload the << and >> operator so that it output and input StringSet objects.

- Overload the == operator so that it returns true if two StringSet objects are the same.

Write a program to test all member functions and overloaded operators in your class.

Step 2: StringSet Similarity

Add a member function that computes the similarity between the current StringSet and an input parameter of type StringSet.

The field of information retrieval is concerned with finding relevant electronic documents based upon a query. For example, given a group of keywords (query), a search engine retrieves Web pages (documents) and displays them sorted by relevance to the query.

A way to make this comparison is to compute the binary cosine coefficient. The coefficient is a value between 0 and 1, where 1 indicates that the query is very similar to the document and 0 indicates that the query has no keywords in common with the document. This approach treats each document as a set of words.

For example, given the following sample document:

“Write a program to test all member functions and overloaded operators in your class.”

This document would be pared into keywords where case is ignored and punctuation discarded and turned into the set containing the words

{write, a, program, to, test, all member, functions, and, overloaded, operators, in, your, class}

An identical process is performed on the query to turn it into a set of strings. Once we have a query Q represented as a set of words and a document D represented as a set of words, the similarity between Q and D is computed by:

$$\text{Sim} = \frac{|Q \cap D|}{\sqrt{|Q|} \sqrt{|D|}}$$

Create two text files named doc1.txt and doc2.txt. Write some text content of your choice in each file, but make sure that each file contains different content.

Write a program that allows the user to input from the keyboard a set of strings that represents a query. The program should then compare the query to both text files and the output the similarity to each one using the binary cosine coefficient. Test your program with different queries to see if the similarity metric is working correctly.

Submission

You are going to submit electronically:

- All files constituting the solution of this assignment.
- Write a **report** file which describes the programs and input/output **screenshot**
- Place your name and student number in your report, and all source codes as comments.
- Include function prototypes for all of the required functions and any that you may create.
- Don't forget to get rid of all compiler warnings