# MC39 Homework 3: Which Web Page is Most Popular?

Max Hailperin

Due: February 27, 1996

In this homework, you will work on a program for finding which textual line appears most frequently in an input file; for example, this can be used to process the list of web pages Gustavus's server served in a given month to find the most popular page. The key data structure needed is a table for keeping track of the occurrence counts for the lines. In this lab you will

1. read more C++ code
2. work on a complicated implementation (using hashing) of a conceptually simple class (the table), by adding the necessary code to "grow" the table (replace it with a larger one) automatically when it gets full
3. implement from scratch another implementation of the *same* table interface, this one simpler, using binary search trees as we sketched in class Monday
4. compare the performance of the two alternative implementations of the table class

You should familiarize yourself with the `.C` and `.h` files in the directory `~max/www-docs/MC39/hw3`, which are the starting point for your work. Printouts of these files are attached. One other file in that directory, `hw3.data`, can be used as a large input data file to show the intended application of finding the most popular web page.

The program is complete except that there is no complete implementation of the `StringToIntTable` class; the file `HashStringToIntTable` contains a nearly complete implementation of this class, however. The first two tasks below needn't be done in the order listed; note that when you link your program together you will need to include either `HashStringToIntTable.o` or `BstStringToIntTable.o` but not both.

1. Finish the hash table implementation by writing the missing member function, and test your implementation.
2. Write an alternative file, `BstStringToIntTable.C` that also defines the `StringToIntTable` class, but uses a binary search tree structure instead. Test this implementation.
3. Compare the performance of your two versions on the `hw3.data` file.