

Program #7 - Due Tuesday, Oct. 27

You are to write a hash table to help perform a word repetition count of a data file. Part of this assignment is to experiment with different hash codes and different compression functions and to keep track of collisions that occur. You will be looking for the codes and compression functions that results in the fewest number of collisions.

Your program should

1. implement a polynomial hash code and a cyclic shift hash code (separately). Experiment with at least three different primes for the polynomial hash code and at least three different cyclic shifts.
2. implement a MAD compression function. Experiment with at least three different primes. Choose your primes large enough to have a sufficiently small load factor.
3. handle collisions using separate chaining. There are two types of “collisions.” The first isn’t really a collision at all. It is entirely possible that a word in your data file will be repeated, and therefore, it will have the same key as a previous instance of the same word. This is what you will want to count - the number of times each word is repeated. The second type of collision is the real “collision” - when two different words have the same key. This is what you will need to handle using separate chaining.
4. calculate the number of true collisions that occur for each experiment.
5. generate a list of the 50 most used words in the file with a count for the number of times each word is used.

Go to www.gutenberg.org to download three different plain text novels to use as your data files to test your program. (Of course, tell me which novels you are using.)