Burt Rosenberg

Problem Set 7

Rabin–Karp String Searching

Implement the Rabin–Karp string searching algorithm, where you program, `my-grep`, accepts two command line arguments, a string to search, and a file-name to search in.

```
my-grep find-this-string in-this-file
```

The output is a line number and the line for each matching string found. For extra-credit write the program so that if given just one argument, the program accepts input from stdin (filter mode).

The Rabin–Karp algorithm was described in class. Here it is again: hash the input string and compare the hash value to the hash values of equal lengthed substrings from the file. Methodically test the hash value against all equal lengthed substrings from the input file by incrementally updating the hash value as you walk character by character through the file. When the hash values match, test the input string and file substring directly for a complete and exact match.

Use this hash function:

$$x = a[i]d^{M-1} + a[i+1]d^{M-2} + \ldots + a[i+M-1] \pmod{Q},$$

where $d = 32$ and $Q = 33554393$.