

Programming Assignment #3 Chained Hashing

Overview

You are to write a program that reads in a body of text, and inserts the text into a Chained Hash Table Class that you write from scratch. You should keep an instance count for each word in the text. Each time a duplicate word is added you should be able to update the instance count for that word. At the end of the text insertion you will show statistics for the run.

Details

The purpose of this exercise is to give you familiarity with Chained Hash Tables. You will create a chained hash table with a length of 127 buckets. You will read in a lengthy text file; Bram Stokers Dracula as text from guttenberg.org <http://www.gutenberg.org/ebooks/345.txt.utf-8>. If you have problems search for 345 on Gutenberg.org, and select the **Plain Text UTF-8** link. As you read the text you will map all words to lower case, and remove all punctuation characters (Only numbers and letters should remain). Once the words are pre-processed you will insert them into a Chained Hash Table. You are to use a hash function that adds the ASCII code for each letter together, and mods the result by 127. You will insert each word to the head of the chain located at the bucket when the bucket is determined.

Do not modify the text file you download, although there are headers and footers from Gutenberg, please leave them intact.

When the hash table is completely loaded you will perform queries to exhibit the following things:

- How many times do the following words appear in the text?
 - transylvania
 - harker
 - renfield
 - vampire
 - expostulate
 - fangoriously
- What is the length of the longest chain?
- What is the length of the shortest chain?
- How many distinct words are there in the text?
- What word is at the head of the chain in buckets 0, 27, 43, and 126?
- What is the total word count inclusive of duplicates? *add up all inside counts*
- Which word occurs most frequently?

No User Interface is required for this project.

Suggestions (Use at your discretion)

- Download the text file to your local machine.
- Write a program to open the text file and display it verbatim before you start adding it to the table.
- The individual chains can contain information to help with the queries (length of chain, etc).

Grading

To receive full credit for this program the following must be submitted:

- | | | |
|----|---|-------|
| 1) | Pseudo code for the algorithms utilized. | (25%) |
| 2) | Well commented Syntactically Correct Java code. | (60%) |
| 3) | A run as instructed above. | (15%) |

Late programs will not be accepted. Waiting until the last minute (to start this) is probably not a good idea. The program is due at the start of class on May 3rd.