Dictionaries and hash tables

Comp310 Object oriented data structures in Java

# TopicS

* Dictionaries
* Hash Tables

# Readings

Carrano: Chapters 19 & 20 (*read as much as you will need for the assignment)*

# Objectives

Understand dictionary and hash table

* Concepts
* Methods of implementation

# Instructions

## Implementation

### Overview

**Write a Java program that will**

* Read in a text file
* Count the number of occurrences of each word.

### Details

* Store the words and their counts in a hash table with the word as the key.
* User determines the hash table size.
* The method used for converting a key to its hash code (index into the hash table) is up to you, but a simple option is to sum the Unicode value of each character. For example:

Key **“The”**

Hash code = (int)’T’ + (int)’h’+ (int)’e’

= 84 + 104 + 101

= 289 *(then compress to table size)*

* Use separate chaining for storing words
* Once the end-of-file is reached, print the hash table.

## B. Testing

Run your program numerous times

* varying
  + Input files (i.e., number of different words and the number of occurrence’s of each word)
  + Hash tables sizes.
* Keeping a record of your results for each test

# Analysis / Summary

Answer the following in your Journal, either in paragraph form or as answers to the numbered items

## Analysis

1. What are your observations about the relationship between file size and hash table size?
2. Do you think there is a universally ideal hash table size for this problem? If not, what do you think the hash table size should depend on?
3. What should be the correlation between word frequency and chain length? How can we influence chain length in this respect?
4. What is your observation about how well disbursed the words are in the hash table *(i.e., are most of the words hashed to just a few locations, or do most of the buckets have words)*? Did varying the hash table size have an effect on disbursement of words? If so, in what way?
5. What might be a better hash code, or what made your hash code work?

## Summary

* 1. If you worked in pairs:
     1. **How did you “divide up” the work so that each student still met the objectives for the assignment (i.e., learned, understood and applied the concepts).**
     2. **What was your contribution?**
     3. **How did you coordinate code changes/testing?**
     4. **Other observations about working with a partner?**
  2. Where did you have trouble with this assignment? How did you move forward? What topics still confuse you?
  3. What did you learn from this assignment? *(Please be specific)*