

## COSC 160 Data Structures Spellchecker Based on Hashing

Implement a spelling checker by using a hash table. Assume the dictionary comes from an external file (en\_US.dic<sup>1</sup>). Your spell checker is required to do the following:

- read in two files from the command line ( a dictionary file and a textfile to spell check)
- output to the screen and to a log file the following:
  - misspelled word and line number containing the misspelled word
  - total words checked
  - total number of misspelled words
  - number of collisions in creating hash table
  - total number of unsuccessful probes (word not found in hash table)
  - total number of successful probes (word found in dictionary, note count each access/probe)
  - load factor (lamda) of the table
  - size of hash table
- You can use any of the following hashing methods
  - separate chaining
  - linear probing
  - quadratic probing
  - double hashing

\*\*\*\*\* FOR Extra Credit (2.5 pts) implement a second hashing function using Universal Hashing and compare access times \*\*\*\*\*

- Use Horner's rule or a similar hashing routine to generate a key for the strings.

good luck

Note:

- You should remove punctuations ( . , ! ? ) ( - " ' ) before checking a word.
- You can ignore capitalization
- The dictionary file is a text file, where each line contains an acceptable spelling for a word. There may be additional extraneous information after the accepted spelling. So you should remove any extraneous information before storing the word in the hash table.

Example

WORD IN DICTIONARY FILE

Abilene/M

WORD TO STORE IN HASH TABLE

Abline

---

<sup>1</sup> taken from <http://wordlist.aspell.net>