**Introduction**

The task is to ask the user which implementation they would like to use bst or hash table

Then read a file with a word and embedding per line

Store them in their respective data structure chose then given a pair of two words calculate the similarities using the formula

sim(w0, w1) = (e0 · e1 )/(|e0||e1|)

Proposed solution

First ask the user which one they would like to use like so

option=input("Binary tree or hash table")

depending on the input read the document into the correct data structure.

To compare them take the words and emending and apply the formula.

**Conclusion**

Hash tables are great for searching and finding items in constant time. Where as a binary tree may take log n time to compare and find elements.

**Academic Certification statement**

I certify that the attached written or electronically transmitted material is my own work. I further certify that (a) I have formally cited or otherwise fully acknowledged the quotations, ideas, and wording used here from other sources, whether published or unpublished, in written or electronic form, (b) I have engaged in no falsification or misrepresentation of data or experience in this submission, (c) I have disclosed any collaboration not specifically authorized by the instructor.

I understand that the relative autonomy of the legal profession carries with it special obligations of self-regulation as outlined in the Columbia Law School Bulletin and that any violation of this certification will subject me to discipline, including possible suspension or expulsion by the Law School, declination to certify for admission to the Bar, and sharing of information about discipline with Bar admissions committees. I further understand that the Law School is required to notify Bar admissions committees if a student has been the subject of disciplinary proceedings, regardless of the outcome of those proceedings.

**Apendix**

#Jusitn Ruiloba

#lab 5

#CS2302 Data structures

#4/5/19

class HashTableC(object):

# Builds a hash table of size 'size'

# Item is a list of (initially empty) lists

# Constructor

def \_\_init\_\_(self,size):

self.item = []

self.num\_items= size

for i in range(size):

self.item.append([])

def InsertC(H,k,l):

# Inserts k in appropriate bucket (list)

# Does nothing if k is already in the table

b = h(k,len(H.item))

H.item[b].append([k,l])

def FindC(H,k):

# Returns bucket (b) and index (i)

# If k is not in table, i == -1

b = h(k,len(H.item))

for i in range(len(H.item[b])):

if H.item[b][i][0] == k:

return b, i, H.item[b][i][1]

return b, -1, -1

def h(s,n):

r = 0

for c in s:

r = (r\*255 + ord(c))% n

return r

#######################################

########################

#binary tree

# Code to implement a binary search tree

# Programmed by Olac Fuentes

# Last modified February 27, 2019

class BST(object):

# Constructor

def \_\_init\_\_(self, item, left=None, right=None):

self.item = item

self.left = left

self.right = right

def Insert(T,newItem):

if T == None:

T = BST(newItem)

elif T.item > newItem:

T.left = Insert(T.left,newItem)

else:

T.right = Insert(T.right,newItem)

return T

def Delete(T,del\_item):

if T is not None:

if del\_item < T.item:

T.left = Delete(T.left,del\_item)

elif del\_item > T.item:

T.right = Delete(T.right,del\_item)

else: # del\_item == T.item

if T.left is None and T.right is None: # T is a leaf, just remove it

T = None

elif T.left is None: # T has one child, replace it by existing child

T = T.right

elif T.right is None:

T = T.left

else: # T has two chldren. Replace T by its successor, delete successor

m = Smallest(T.right)

T.item = m.item

T.right = Delete(T.right,m.item)

return T

def InOrder(T):

# Prints items in BST in ascending order

if T is not None:

InOrder(T.left)

print(T.item,end = ' ')

InOrder(T.right)

def InOrderD(T,space):

# Prints items and structure of BST

if T is not None:

InOrderD(T.right,space+' ')

print(space,T.item)

InOrderD(T.left,space+' ')

def SmallestL(T):

# Returns smallest item in BST. Returns None if T is None

if T is None:

return None

while T.left is not None:

T = T.left

return T

def Smallest(T):

# Returns smallest item in BST. Error if T is None

if T.left is None:

return T

else:

return Smallest(T.left)

def Largest(T):

if T.right is None:

return T

else:

return Largest(T.right)

def Find(T,k):

# Returns the address of k in BST, or None if k is not in the tree

if T is None or T.item == k:

return T

if T.item<k:

return Find(T.right,k)

return Find(T.left,k)

def FindAndPrint(T,k):

f = Find(T,k)

if f is not None:

print(f.item,'found')

else:

print(k,'not found')

# Code to test the functions above

T = None

A = [70, 50, 90, 130, 150, 40, 10, 30, 100, 180, 45, 60, 140, 42]

for a in A:

T = Insert(T,a)

InOrder(T)

print()

################

################################

#hash

# Implementation of hash tables with chaining using strings

option=input("Binary tree or hash table")

f = open("text.txt", "r")

print(f.readline())

print(f.readline())

if option == "hashtable":

H = HashTableC(11)

A = ['data','structures','computer','science','university','of','texas','at','el','paso']

for a in A:

InsertC(H,a,alen(a))

print(H.item)

for a in A: # Prints bucket, position in bucket, and word length

print(a,FindC(H,a))