

CSE13S Winter Quarter 2022

Assignment 7: Author Identification

Description of the program:

- This assignment attempts to identify the author of a document by looking at their diction and comparing their word usage to other authors

Files to be included in the Asgn7 directory

- bf.h
- bf.c
- bv.h
- bv.c
- ht.h
- ht.c
- identify.c
- metric.h
- node.h
- node.c
- parser.h
- parser.c
- pq.h
- pq.c
- salts.h
- speck.h
- speck.c
- text.h
- text.c
- Makefile
- DESIGN.pdf
- WRITEUP.pdf

Pseudocode:

- HashTable *ht_create(uint32_t size)
 - malloc hashtable
 - initialize size as size
 - allocate memory for every word or entry
- void ht_delete(HashTable **ht)
 - for s in range(size)
 - free memory allocated to words

- free hashtable
- uint32_t ht_size(HashTable *ht)
 - return size of hashtable
- Node *ht_lookup(HashTable *ht, char *word)
 - increment occurrences for node
 - look in hashtable for word
- Node *ht_insert(HashTable *ht, char *word)
 - increment occurrences for node
 - set index of hashtable is word
- HashTableIterator *hti_create(HashTable *ht)
 - iterate through hashtable size
- void hti_delete(HashTableIterator **hti)
 - set pointer as 0
- Node *ht_iter(HashTableIterator *hti)
 - return *hti
- Node create functions can be taken from the previous assignment
- BloomFilter *bf_create(uint32_t size)
 - allocate memory
 - run through all three hash iterations
- void bf_delete(BloomFilter **bf)
 - free memory
- uint32_t bf_size(BloomFilter *bf)
 - return size
- void bf_insert(BloomFilter *bf, char *word)
 - set bitvectors for all three iterations
- bool bf_prove(BloomFilter *bf, char *word)
 - if all the indices of bf_insert are not 0 return true
- BitVector *bv_create(uint32_t length)
 - allocate memory
- void bv_delete(BitVector **bv)
 - free memory
- uint32_t bv_length(BitVector *bv)
 - return bit vector length
- bool bv_set_bit(BitVector *bv, uint32_t i)
 - bit twiddle to set bitvector
 - return true
 - if i is greater than bv_length return false
- bool bv_get_bit(BitVector *bv, uint32_t i)
 - same as bv_set_bit but get bit using temp variable and bit twiddle
- Recycle priority queue