CSE 250 Recitation

April 24 - 28: Hash Tables

PA4 Testing Tips

In PA4 you will be de-anonymizing data based on a person's voter record and health record. Each record contains a birthday and a zip code field, which will be used to determine unique matches.

How can we tell if two records are "unique"?

PA4 Testing Tips

In PA4 you will be de-anonymizing data based on a person's voter record and health record. Each record contains a birthday and a zip code field, which will be used to determine unique matches.

How can we tell if two records are "unique"?

How do we deal with null values? (they act as wildcards)

Hashing

Take the items A-E and their corresponding hash values:

- hash(A) = 636
- hash(B) = 712
- hash(C) = 459
- hash(D) = 12
- hash(E) = 154
1. Start with a 5-bucket hash table (with chaining) and insert the above items
2. Rehash the table, doubling its size to 10

Open Addressing

Take the items A-E and their corresponding hash values:

- hash(A) = 636
- hash(B) = 712
- hash(C) = 459
- hash(D) = 12
- hash(E) = 154
1. Start with a 5-bucket hash table (with open addressing) and insert the above items
2. Run through the process of looking up records A-E and F (hash(F) = 232)
3. Remove item B
4. Rehash, doubling the array size to 10 and repeat steps 2 and 3

Cuckoo Hashing

Take the items A-E and their corresponding hash values:

- $hash_1(A) = 312$ $hash_2(A) = 636$
- $hash_1(B) = 242$ $hash_2(B) = 712$
- $hash_1(C) = 684$ $hash_2(C) = 459$
- $hash_1(D) = 871$ $hash_2(D) = 12$
- $hash_1(E) = 154$ $hash_2(E) = 939$

1. Start with a 5-bucket hash table (with cuckoo hashing) and insert the above items (rehash as needed)